Transcriptome-level assessments and computational characterization of disease models based on induced pluripotent stem cells (iPSCs)

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Wasco Wruck

aus Lübeck

Düsseldorf, 04.10.2021

aus dem Institut für Stammzellforschung und Regenerative Medizin des Universitätsklinikum Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. James Adjaye

2. Prof. Dr. Gunnar Klau

Tag der mündlichen Prüfung:
25.03.2022

Die vorliegende Arbeit ist eine kumulative Dissertation.

HEINRICH-HEINE-UNIVERSITÄT DÜSSELDORF

Abstract

Mathematisch-Naturwissenschaftliche Fakultät

Medical Faculty

Institute for Stem Cell Research and Regenerative Medicine

Doctor rerum naturalium

Transcriptome-level assessments and computational characterization of disease models based on induced pluripotent stem cells (iPSCs)

by Wasco Wruck

Since the groundbreaking discovery of induced pluripotent stem cells (iPSCs) by Shinya Yamanaka and his team much research effort has been spend in differentiation of iPSCs to adult somatic cells of various lineages. The rational behind that is to get personalized cells of different tissue types which have multiple applications in research and therapy. While in research of today iPSC-based disease models can already be employed, therapeutic approaches such as the development of whole organs from stem cells are a long-term goal on the way to which however progress has already been made by generation of organoids for many organs. For disease models, often maturity is an important issue because cells differentiated from pluripotent stem cells by nature resemble the fetal phenotype when the differentiation process is considered to recapitulate the development in the embryo. On the other hand, aging is an essential factor in many or even most diseases, e.g. in Alzheimer's disease and cancer. This could result in obstacles when cells generated with a fetal phenotype shall be used to model a phenotype only appearing in aged individuals. Assessment of iPSC-based disease models in comparison to adult cell types is indispensable. These adult cell types are originated in biopsies or in transformed cell lines which have disadvantages of limited availability sometimes only post-mortem and association with a cancerous phenotype. The iPSC-based models would overcome most of these drawbacks but would need improvement in reaching an adult or even aged phenotype.

In this dissertation I assessed transcriptome data from stem-cell-based disease models in comparison to biopsy-derived transcriptome data. Main focus were Non-alcoholic fatty liver disease (NAFLD), Alzheimer's disease and brain aging and in consequence liver biopsies from NAFLD patients, iPSC-derived hepatocyte-like cells (HLCs), post-mortem brain biopsies from AD patients and aged individuals and iPSC-derived neuronal cells. Additionally, also urine-derived renal progenitor cells (UdRPCs) were investigated which

represent a mature phenotype because they were not reprogrammed to the pluripotent state. Biopsy-derived data was investigated in meta-analyses of microarray and next-generation sequencing gene expression data retrieved from public repositories. For the meta-analysis of aging in the human prefrontal cortex an approach was developed to infer relationships between genes from gene expression time series based on the concept of Granger causality. Applying this approach predicted Granger causality between *CAMK4* and *GFAP*. The comparison of NAFLD in an iPSC-based model and biopsy-derived transcriptome data showed congruency in the results - most outstanding the involvement of the gene *PLIN2* coding a protein surrounding lipid droplets and thus a crucial biomarker of the disease.

I want to conclude that this dissertation showed that the stem-cell-based models are already capable of capturing many relevant facets of a disease but need further research to improve them in their potential to fully reflect the mature or aged phenotype which is a precondition for the development of the investigated diseases.

Zusammenfassung

Seit der bahnbrechenden Entdeckung der induzierten pluripotenten Stammzellen (iPSCs) durch Shinya Yamanaka und sein Team wurde viel Forschungsarbeit in die Differenzierung von iPSCs in adulte somatische Zellen verschiedener Linien investiert. Die Überlegung dahinter ist, personalisierte Zellen von verschiedenen Gewebetypen zu erhalten, die vielfache Anwendungen in Forschung und Therapie haben. Während in der heutigen Forschung iPSC-basierte Krankheitsmodelle bereits eingesetzt werden können, sind therapeutische Ansätze wie die Entwicklung ganzer Organe aus Stammzellen ein langfristiges Ziel, auf dem Weg zu dem allerdings Fortschritte durch die Erzeugung von Organoiden für viele Organe gemacht wurden. Für Krankheitsmodelle ist oft der Reifegrad ein wichtiges Kriterium, weil aus Stammzellen differenzierte Zellen naturgemäß dem fötalen Phenotyp ähneln, wenn der Differenzierungsprozess als eine Rekapitulation der embryonalen Entwicklung aufgefasst wird. Andererseits ist die Alterung ein essenzieller Faktor in vielen oder sogar den meisten Krankheiten, z.B. in der Alzheimerkrankheit oder in Krebs. Das kann in Hindernissen resultieren, wenn mit einem fötalen Phenotyp erzeugte Zellen benutzt werden sollen, um eine nur in älteren Individuen auftretende Krankheit zu modellieren. Insofern ist die Auswertung von iPSC-basierten Krankheitsmodellen im Vergleich zu adulten Zelltypen unverzichtbar. Diese adulten Zelltypen kommen aus Biopsien oder transformierten Zelllinien, die die Nachteile der limitierten Verfügbarkeit - teilweise erst post-mortem - und die Assoziierung mit dem Krebs-Phenotyp aufweisen. Die iPSC-basierten Modelle überwinden die meisten dieser Nachteile, benötigen aber Verbesserungen im Erreichen des adulten oder sogar gealterten Phenotyps.

In dieser Dissertation habe ich Transkriptomdaten von stammzellbasierten Krankheitsmodellen im Vergleich zu aus Biopsien gewonnenen Transkriptomdaten ausgewertet. Der wesentliche Fokus waren dabei die Nicht-alkoholische Fettlebererkrankung (NAFLD), die Alzheimererkrankung im Zusammenhang mit dem Alterungsprozess des Gehirns und in Konsequenz Leberbiopsien von NAFLD-Patienten, iPSC-abgeleitete Hepatozyt-ähnliche Zellen (HLCs), Post-mortem Gehirn-Biopsien von AD-Patienten und älteren Individuen und iPSC-abgeleitete neuronale Zellen. Zusätzlich wurden auch aus Urin gewonnene Nierenvorläuferzellen (UdRPCs) untersucht, die einen reifen Phenotyp repräsentieren, weil sie nicht in den pluripotenten Zustand reprogrammiert wurden. Die Biopsie-gestützten Daten wurden in Meta-Analysen von Microarray- und NGS- (Next Generation Sequencing) -Genexpressionsdaten aus öffentlichen Repositorien untersucht. Für die Meta-Analyse der Alterung im humanen präfrontalen Kortex wurde ein Ansatz entwickelt, um auf dem Konzept der Granger-Kausalität basierte Beziehungen zwischen Genen aus Zeitreihen abzuleiten. Die Anwendung dieses Verfahrens erbrachte die Vorhersage von Granger-Kausalität zwischen den Genen *CAMK4* und *GFAP*. Der Vergleich von NAFLD in einem iPSC-basierten Modell und Biopsie-gestützten Transkriptomdaten zeigte Kongruenz in den Ergebnissen. Dabei trat die Involvierung des Gens *PLIN2* in den Vordergrund, das ein Fetttropfen umgebendes Protein kodiert und dadurch ein entscheidender Biomarker der Krankheit ist.

In Zusammenfassung zeigte diese Dissertation, dass die stammzellbasierten Modelle bereits viele wichtige Facetten einer Krankheit abbilden können, aber weitere Forschung notwendig ist, um sie zu verbessern in ihrem Potenzial, den reifen oder gealterten Phäno-

typ voll zu reflektieren, der eine Voraussetzung für die Entstehung der untersuchten Krankheiten ist.

Contents

List of Figures

Abbreviations

AD Alzheimer's disease
ALB Albumin
BP Biological Processes
BLAT BLAST-like alignment tool
BLAST Basic Local Alignment Search Tool
BWA Burrows-Wheeler Aligner
CC Cellular Components
cDNA complementary DNA
ChIP chromatin immunoprecipitation
ChIP-seq ChIP sequencing
CSA compressed suffix array
CRISPR Clustered Regularly Interspaced Short Palindromic Repeats
DAVID database for annotation, visualization and integrated discovery
DNA Deoxyribonucleic acid
FDR false discovery rate
FGF fibroblast growth factor
GABA gamma-Aminobutyric acid
GEO Gene Expression Omnibus
GFAP glial fibrillary acidic protein
GO gene ontology
GO-CAMs GO causal activity models
GSEA Gene Set Enrichment Analysis
GSVA Gene Set Variation Analysis
GTEx Genotype-Tissue Expression project
HLCs hepatocyte-like cells
ICG indocyanine green
ICM inner cell mass
iPSC induced pluripotent stem cell
KEGG Kyoto Encyclopedia of Genes and Genomes
KRT18 Keratin 18
LOAD Late-Onset-Alzheimer's Disease
MF Molecular Functions
MSigDB Molecular Signatures Database
NAFLD non-alcoholic fatty liver disease
NANOG Nanog homeobox
NASH non-alcoholic steatohepatitis
NCBI National Center for Biotechnology Information
NGS Next Generation Sequencing
OCT4 Octamer-binding protein 4/POU class 5 homeobox 1
ORA Over-Representation Analysis
PCA principal component analysis
PCR polymerase chain reaction
PLAGE Pathway Level Analysis of Gene Expression
PLIN2 Perilipin 2
RMA Robust Multi-array Average
RNA-Seq RNA sequencing

SOX2 Sex-determining region Y (SRY)-box 2
SBL sequencing by ligation
SBS sequencing by synthesis
SVD singular value decomposition
TE trophectoderm

Chapter 1

Introduction

The discovery of a way to reprogram adult cells to pluripotent cells - so called induced pluripotent stem cells (iPSCs) - by Yamanaka and his colleagues in mice (Takahashi and Yamanaka 2006) and men (Takahashi et al. 2007) was a big advancement of biological research implicating a plethora of possible applications in stem cell research and regenerative medicine. Pluripotent stem cells can be differentiated into any other cell type and that can be used to generate specific tissue types paving the avenues for research on organoids targeting the long-term goal of generating fully functional organs.

1.1 Pluripotent stem cells

Pluripotent stem cells are cells which possess the ability to differentiate into each cell type of an organism. In contrast to totipotent stem cells which can give rise to a whole organism pluripotent stem cells lack the potential to differentiate into extra-embryonic tissues. On the other hand, they are distinguished from adult stem cells which are multipotent or even unipotent meaning that they can be differentiated into cells of a limited range or only one tissue type (De Los Angeles et al. 2015). The state of pluripotency can be further subdivided into *naïve* and *primed* state (Devika et al. 2019) describing early and late phases of epiblast ontogeny (De Los Angeles et al. 2015). *Naïve* cells can develop into germline cells and all other embryonic lineages and have the flexibility to generate chimeras while at the later *primed* stage cells are primed to the three germ layers ectoderm, mesoderm and endoderm and can no longer generate chimeras.

Figure 1 illustrates that pluripotent stem cells can be differentiated into the three germ layers ectoderm, endoderm and mesoderm. Ectoderm gives rise to cell types such as neurons, endoderm to hepatocytes and mesoderm to cardiomyocytes. In the upper right

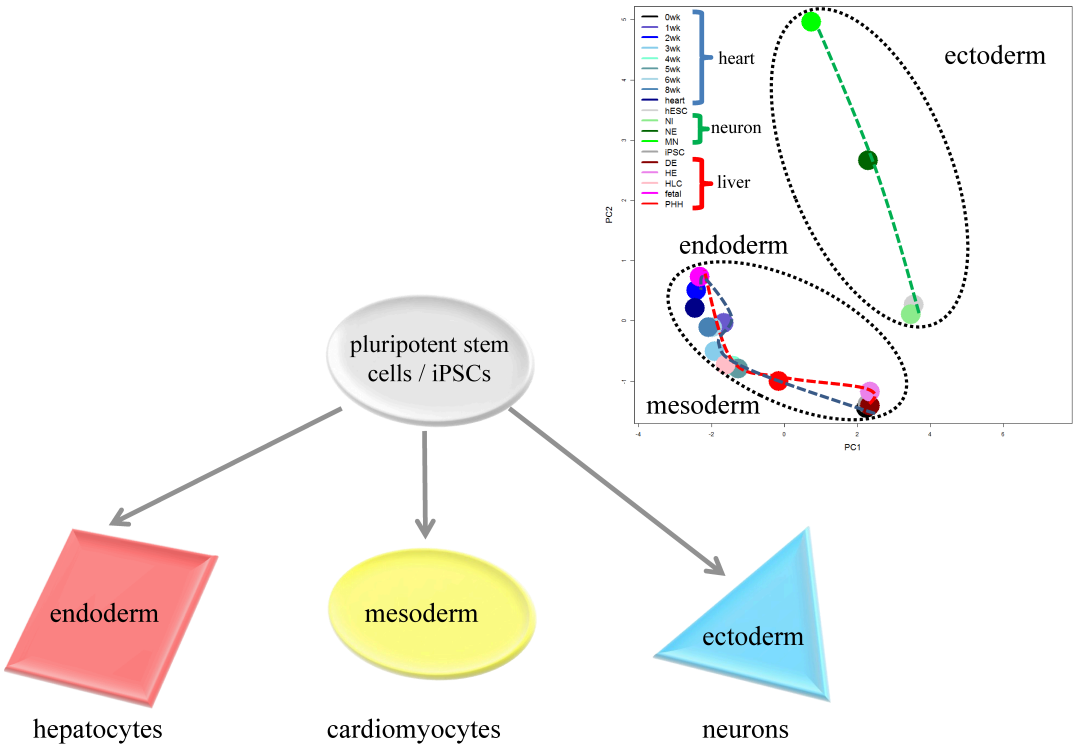


Figure 1: Differentiation of pluripotent stem cells into the three germ layers.

corner a principal component analysis (PCA) of gene expression data of POU family genes was calculated for different developmental stages of heart, neurons and liver. These datasets clusters by the three germ layers although showing close similarity between endoderm and mesoderm for this pluripotency-associated gene family. The trajectories indicated by the dotted lines correlate with developmental stages.

1.1.1 Embryonic stem cells

Embryonic stem cells are pluripotent stem cells derived from the inner cell mass of a blastocyst (Thomson et al. 1998), (Adjaye et al. 2005). Evans et al. first reported establishment of a mouse embryonic cell line derived from a blastocyst in 1981 (Evans and Kaufman 1981) while Thomson et al. reported this for human embryonic cell lines in 1998 (Thomson et al. 1998). Blastocysts develop after the eight-cell morula stage in the human embryo from about day 5 to day 8 after fertilization. The morula gives rise to the trophectoderm (TE), from which the cytotrophoblast and syncytiotrophoblast

of the placenta develop, and the inner cell mass (ICM) from which the embryo and extraembryonic tissues develop (Adjaye et al. 2005).

The great potential of human embryonic stem cells lies in the ability to give rise to any cell of the adult human body and thus they would open up perspectives for a plethora of therapies in the context of regenerative medicine. Future research will target the intermediate steps of the derivation and improvement of organ-specific cell types and the development of organoids - small functional parts of whole organs - with the ultimate goal of the generation of functional organs from pluripotent stem cells. However, one serious challenge in transplanting cells or organs derived from embryonic stem cells represents the immunological rejection of biological material from a donor different from the acceptor. As furthermore embryonic stem cells are associated with ethical and legal issues varying between countries, in the future most developments will be based on induced pluripotent stem cells (iPSCs) which are described in the next paragraph. Nevertheless, embryonic stem cells are often used as a "gold standard" in particular when induced pluripotent stem cells shall be characterized.

1.1.2 Induced pluripotent stem cells

The finding of specific expression of *OCT4* (*POU5F1*) and other genes such as *SOX2* and *NANOG* in embryonic stem cells and the findings that they are key regulators of pluripotency-guarding transcriptional networks (Babaie et al. 2007), (Pesce, Gross, and Schöler 1998),(Nichols et al. 1998) led to the discovery by Yamanaka and colleagues that inducing expression of pluripotency-related genes in adult cells can reprogram them to a pluripotent state in mice (Takahashi and Yamanaka 2006) and men (Takahashi et al. 2007). This Nobel-Laureated discovery with one strike overcame the drawbacks of embryonic stem cells: immunological problems, ethical and legal issues. In the meantime reprogramming of adult cells of a plethora of cell types to so-called induced pluripotent stem cells (iPSCs) has become routine for stem-cell-committed laboratories (Martins et al. 2019), (Schröter et al. 2016),(Bohndorf et al. 2017), (M. Zhang et al. 2014). In the era of personalized medicine iPSCs are a valuable means to generate differentiated cell types of patients with isogenic gene profiles. Additionally, iPSCs allow access to cell types which are otherwise complicated or even impossible to gain from healthy individuals such as cells from brain, kidney or liver.

1.2 Stem cell based disease models

Stem cell based disease models take advantage of the differentiation potential of pluripotent stem cells. They proceed a path - appearing complicated on the first glimpse - of reprogramming adult cells to iPSCs and then again differentiating them to cells of dedicated tissue types, e.g. instead of directly taking liver biopsies skin biopsies are taken, the fibroblast from the skin biopsies are reprogrammed to iPSCs and the iPSCs are differentiated into liver cells. However, this putatively complicated procedure circumvents the problems of complications during the liver biopsy and the limited culturability of liver cells. In other cases such as Alzheimer's disease (AD) research stem cell based disease models enable studies on neuronal cell cultures or brain organoids with patient-specific genotypes which obviously would not be possible by direct access to the target cell type.

1.2.1 Differentiation of pluripotent stem cells into hepatocyte-like cells

Many protocols for deriving hepatocyte-like cells (HLCs) from pluripotent stem cells have been published (Matz et al. 2017),(Wruck and Adjaye 2018), (Graffmann et al. 2016),(Jozefczuk et al. 2010), (Hay et al. 2008). The in vitro differentiation of pluripotent cells progresses over the intermediate states of definitive endoderm and hepatic endoderm to HLCs. Definitive endoderm is usually induced by Activin A (or Nodal) thus recapitulating processes taking place in gastrulation in vivo (Józefczuk 2009).

In vivo studies provided insight how the endoderm is organized to give rise to specific organs at specific time points - in particular to the liver which develops from the ventral foregut endoderm. Hepatogenesis is regulated by exchange of fibroblast growth factor (FGF) signals between cardiac mesoderm and ventral foregut endoderm in a way that different FGFs initiate distinct phases, e.g. FGF1 or FGF2 induce hepatic gene expression in mice while FGF8 contributes to morphogenetic outgrowth of the hepatic tissue after specification. Importantly, FGFs also block development of the pancreatic lineage from endoderm which emerges in the absence of these factors (Zaret 2002). The further development of liver is controlled by transcription factors operating on the intracellular level including FOXA1 (HNF3A) and FOXA2 (HNF3B), HNF4A, HNF1A/B, HNF6 (ONECUT1) and CCAAT enhancer binding protein (CEBPA) and cytokines operating on the extracellular level including Activin A (NODAL), FGFs, BMPs , OSM and HGFs (Józefczuk 2009). Of these HNF4A has a central role in hepatocyte function and differentiation particularly in converting the hepatic parenchyma to epithelial cells and maintaining the organization of the sinusoids (Zhao and Duncan 2005). The knowledge gained about the role of these factors in liver development has been transferred to the development of protocols for in vitro differentiation of HLCs targeting at cells which most closely resemble mature liver cells. The quality of the HLCs can be measured by expression of several marker genes - among which Albumin (ALB) is most important -

and several functional assays such as glycogen storage, indocyanine green (ICG) uptake and release, urea and bile acid production, cytochrome activity (Matz et al. 2017).

1.2.2 Differentiation of pluripotent stem cells into neurons

Neuronal cells can be derived from pluripotent stem cells according to several published protocols, e.g. the protocol by Zhang et al. (S. C. Zhang et al. 2001) for the induction of neural precursors or the protocol by Liu et al. for the induction of Forebrain GABA interneurons (Liu et al. 2013). Zhang et al. used FGF2 to differentiate human embryonic stem cells to neural tubular structures (S. C. Zhang et al. 2001).

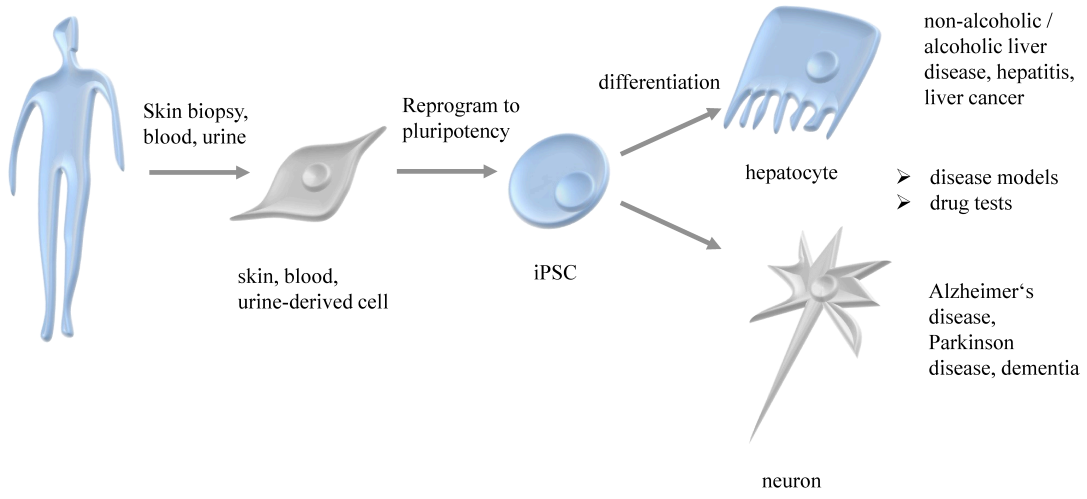


Figure 2: iPSC-based disease models.

Figure 2 depicts the process of generating iPSC-based disease models for the examples of liver diseases and neurodegenerative diseases. Biological material which can easily

be accessed from the specific patient such as urine, blood or skin is used as basis to reprogram iPSCs. The iPSCs in turn are differentiated into the hepatocyte or neuron lineage. If the material comes from patients of the disease of interest the differentiated cells can be used directly as disease model. Otherwise, the disease can be mimicked in manifold ways such as adding oleic acid to hepatocytes or introducing disease-causing mutations via CRISPR-Cas9 (Doudna and Charpentier 2014). In case of Alzheimer's disease also the addition of an amyloid-β peptide can be employed to simulate facettes of the disease (Martins et al. 2020).

1.3 Bioinformatic analyses of pluripotent stem cells and derived disease models

1.3.1 Gene expression analysis

1.3.1.1 Microarrays

Microarrays have been first employed in 1983 by Tse Wen Chang in an antibody microarray to measure expression of multiple cell surface antigens in parallel (Chang 1983). Later a publication from the Pat Brown lab introduced the microarray technology for monitoring the expression of multiple genes in parallel via an array of roboter-printed complementary DNA (cDNA) probes (Schena et al. 1995). The technology was advanced to increase the densities of spots printed with roboters on to nylon membranes or glass slides eventually reaching the goal of representing the whole transcriptome of an organism such as the human on a single microarray. For the technology at that stage sophisticated image analysis and visualization software was indispensable as the resulting scanned images of fluorescent or radioactive labeled hybridizations came with spots arranged in considerably distorted grid arrays (Steinfath et al. 2001), (Wruck et al. 2002). Nevertheless, it was possible to assess the whole genomes, e.g. the genes from the ENSEMBL database (Hubbard et al. 2002) on such a microarray (Adjaye 2005). Following these advancements companies such as Affymetrix and Illumina launched microarrays manufactured in industrial processes providing high precision and accuracy what was confirmed in a large-scale microarray quality assessment project (MAQC Consortium et al. 2006), (Mane et al. 2009) . Also other companies offered microarrays but here I will focus on the most widespread companies Illumina and Affymetrix. The basic principle of microarrays is to sample genes via short cDNA sequences so-called "oligonucleotides". In the case of Illumina the oligonucleotides consist of 50 bases, Affymetrix used 25 bases in the beginning but later also increased this. Illumina uses a single probe in high redundancy (e.g. 30-fold) to measure a gene while Affymetrix uses a so-called "probesets" consisting of multiple probes, e.g. 10 probes - matching the RNA sequence of the transcript at multiple specific locations. Affymetrix microarrays called "GeneChips" are built using a photolithographic process. Photolithographic masks for bases A,C,G and T are applied to the chip in multiple iterations to synthesize oligonucleotides in situ.

Illumina uses a technology based on beads to which the 50-mer probes are attached (Gunderson et al. 2004). These beads are taken from pools and are randomly distributed over the array with a ~30-fold redundancy for each probe. The random distribution necessitates a decoding of the location of each probe which is performed during the manufacturing process using hybridizations of additional adapters to the beads. Thus, the location of the probes is unique for each chip and is provided to the user.

1.3.1.2 Analysis of gene expression microarray data

Most microarray platforms include image processing modules and deliver signal intensities of probes or probesets (Affymetrix) referring to genes or dedicated transcript variants as results. Although some proprietary software modules also provide data analysis facilities superior performance has been shown for the plethora of software modules developed by the Bioinformatics community in the open source R/Bioconductor statistical environment (Gentleman et al. 2004), (Gautier et al. 2004), (Du, Kibbe, and Lin 2008). Gene expression microarray analysis includes multiple steps including background correction, normalization and statistical tests. Background correction refers to removal of background noise appearing in the image. Pixels without biological information nevertheless have intensities which often have dedicated spatial distributions and are also dependent on global parameters such as brightness and contrast. Affymetrix extended this notion to also include portions of the signal due to unspecific binding. Simple ways to correct the background are to add a socket value to intensities subtracted by the minimum intensity or to subtract an quantile-derived offset value derived from a predefined quantile of the instensities (Du, Kibbe, and Lin 2008). Some proprietary methods which result in negative values should not be used because they impede logarithmic transformations often applied during follow-up processing. Before normalization data is usually transformed to a logarithmic scale which for high values is equivalent to the arcus sinus hyperbolicus function suggested by Huber et al. for stabilization of the variance (Huber et al. 2002).

Normalization is indispensable to make different datasets comparable meaning that the true biological but not technical differences, e.g. different scanner brightness adjustments can be assessed. Methods for normalization range from median-normalization (division by the median) over Lowess-normalization discarding bias between the two color channels of microarrays (Cleveland 1979) to quantile-normalization assigning the median of all sorted datasets to the sorted datasets themselves and thus bringing them all to the same quantile distribution (Gautier et al. 2004). For Affymetrix the quantile normalization has been extended to Robust Multiarray average (RMA) normalization which additionally includes a median polish procedure condensing the values of the probes to one value for the probeset. Special variants of RMA have been developed such as GCRMA taking into account GC-content for normalization and fRMA (McCall, Bolstad, and Irizarry 2010) enabling normalization on single array base which allows for later addition of array without the need for re-normalization.

Experimental design in the biomedical area usually includes comparison to control experiments what allows for a better dissection of factors involved in processes induced by a treatment or a disease. A good choice of the control which should be very close to the tested condition, e.g. same gender and about the same age, is therefore paramount in order to pinpoint the crucial determinants of the investigated condition. Differential gene expression tests are employed to detect genes which are involved in the etiology of a disease or induced by a specific treatment. The classical parametric test which is employed for differential gene expression is Student's t-test often used in the Welch-test version which corrects unequal variances. These tests are based on the assumption that data are approximately normally distributed what usually is the case. However, this precondition can be omitted when the non-parametrical Wilcoxon rank-sum test is applied which provides similar yet more conservative results but needs larger sample sizes (Herwig et al. 2001). Superior results are often achieved - particularly when sample sizes are small - by taking advantage of the multitude of measurements incorporated on the microarrays via the application of linear models on the data and basing the statistical test on these models as is done for the statistical test from the R package limma (Smyth 2004). P-values from the statistical tests should be adjusted for the multiple testing problem arising from the multitude of tests executed in parallel. One method to achieve this is the q-value method (Storey 2002) providing a false discovery rate (FDR). Differentially expressed genes are then often determined using a logical combination of thresholds for the p/q-value for differential expression, the ratio and the detection-p-value., e.g. (q<0.05) AND ((ratio<0.5) OR (ratio>2)) AND ((detection_p_control<0.05) OR (detection_p_treatment<0.05)). The detection-p-value is a measure often provided by the proprietary software judging how good a probe intensity separates from the background what is usually interpreted as a measure for gene expression. This tells - absolutely - if a gene is expressed in contrary to the differential expression which tells if a gene is expressed differently in comparison to the control. Thus, the detection-p-value can be used as additional criterion for the determination of differentially expressed genes in order to filter out genes which have expression in the noise level. Using also a ratio threshold is important to account for the effect size and to sort out too small changes which nevertheless might have low p-values.

1.3.1.3 Next-generation sequencing (NGS)

The Human Genome project was mainly performed employing Sanger sequencing and resulted in the identification of the whole human DNA sequence in a world-wide effort of many leading research institutes lasting several years. After it had been declared complete in 2003 sequencing methods and devices have been continuously improved up to the degree that it was enabled to sequence the whole genome of a human in less then one day. A further achievement important for upcoming applications in diagnostics and personalized medicine was reaching the milestone of the 1000-Dollar genome - meaning that a human genome could be fully sequenced for less than 1000 Dollar. That shifted the technology into an area where it was possible to use it for routine diagnostics and large-scale projects in the context of personalized medicine. An important project in this context was the 1000 genomes project which delivered the DNA sequence of the genomes of more than 1000 individuals from all over the whole world and thus providing a valuable source of variations in a publicly accessible database (Clarke et al. 2012). The short-

read sequencing (35-700bp) technologies usually applied in Next-generation-Sequencing can be categorized as sequencing by ligation (SBL) and sequencing by synthesis (SBS) (Goodwin, McPherson, and McCombie 2016): in SBL a fluorescent probe hybridizes to a DNA fragment using DNA ligase to identify the nucleotide. In SBS fluorescently labeled reversible terminally blocked nucleotides which are complementary to the measured base are sequentially imaged and afterwards cleaved and washed for the new cycle.

1.3.1.4 Analysis of NGS data

NGS data analysis can target distinct goals such as determining gene expression (RNA-Seq), finding variants in the DNA or finding interactions between proteins and DNA (ChIP-seq). In any case, the NGS raw data - the reads - have to be aligned against a reference genome. Alignment software has been continuously improved to face several computational challenges such as to cope with the huge amount of memory and computational power needed or algorithmic challenges such as to cope with mismatches within the reads to be aligned. Soon after the launch of the first next-generation sequencing devices software such as BWA (H. Li and Durbin 2010) was developed which outperformed existing programs like BLAT or BLAST. Later the so-called "Tuxedo-suite" software from the Trapnell laboratory provided one of the best performing alignment algorithms named BOWTIE (Langmead et al. 2009) and additionally whole pipelines also incorporating transcript assembly (software TOPHAT) and differential expression (software CUFFLINKS) assessment (Trapnell et al. 2012). However, many other software solutions for alignment and transcript assembly exist which differ in hardware requirements, performance and also in the number and accuracy of mapped reads. Shortly after the Tuxedo-suite the STAR software was developed (Dobin et al. 2013) and was state of the art for a period of time but was soon after caught up by a new evolution of the Tuxedo suite including the HISAT aligner (Kim, Langmead, and Salzberg 2015) and the StringTie transcript assembly (Pertea et al. 2015). Later, improved transcript assembly and abundance estimation was reported in a publication about the CIDANE software (Canzar et al. 2016) . Modern fast alignment algorithms are usually based on Burrows-Wheeler transforms and suffix trees or suffix arrays - combined to the compressed suffix array (CSA) introduced by Lippert et al. (Lippert 2005) - which are employed to index the genome and make it accessible efficiently (Döring et al. 2008), (Delcher et al. 2002).

1.3.2 Functional analysis of gene sets

Most biological processes involve not only single genes but complex networks of genes interacting with each other via various regulatory mechanisms. With respect to this systemic perspective the gene expression analysis has to shift from the Mendelian approach associating functionality with a single gene to an approach associating functionality with a gene set. Gene sets usually result from gene expression analysis after filtering for significance criteria including correction for the multiple testing problem. The next level

of analysis then aims at associating functionality with these gene sets. This is usually achieved by so-called over-representation tests - or sometimes also called enrichment or functional annotation tests, e.g. in the context of the DAVID tool (Huang, Sherman, and Lempicki 2009).

1.3.2.1 Statistical methods

Over-representation can be tested with the hypergeometric test or Fisher's exact test which for the case of 2x2 contingency tables is equivalent to the hypergeometric test. Therefore, here we focus on the hypergeometric test. The hypergeometric test is based on the hypergeometric distribution which relates to the probability of drawing k objects of a specific type in n trials, without replacement, from a population of size N containing K objects of the type of interest, e.g. the probability to draw k red marbles in n trials from an urn containing K red marbles among N marbles. The probability of drawing exactly k marbles is then:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \tag{1.1}$$

However, usually one is not interested in the probability of drawing an exact number of marbles but in the probability of drawing more marbles than the drawn ones. This would refer to the sum of the probabilities of drawing all possible numbers of marbles greater or equal to the drawn number of marbles. This can be calculated via the cumulative distribution function of the hypergeometric distribution:

$$P(X \leq k) = \sum_{y=0}^{k} \frac{\binom{K}{y}\binom{N-K}{n-k}}{\binom{N}{n}} \tag{1.2}$$

The probability of drawing k or more marbles is:

$$P(X \geq k) = P(X > k-1) = 1 - P(X \leq (k-1)) \tag{1.3}$$

This probability relates to retrieving more than k significant genes from a pathway/GO with n genes when there are K significant genes among N genes in total (i.e. genes mappable to pathways/GOs).

Another way of analyzing gene sets is the so-called gene set enrichment analysis (GSEA) which was introduced by Subramanian et al. (Subramanian et al. 2005). This method ranks the genes in the gene set in comparison to a so-called *background* gene set of e.g. all genes on a microarray. GSEA is based on the Kolmogorov-Smirnov-test and takes into account the expression values of the involved genes while the over-representation test solely works on numbers of gene symbols. Thus, GSEA has the advantage to incorporate more information into the statistical assessment but on the other hand cannot be applied in the frequent cases where only gene symbols but no expression values are available. A plethora of further approaches has tried to improve the performance of gene set analysis, e.g. the GSVA method which can analyze gene sets using microarray or NGS gene expression data in a sample-wise manner and also in other experimental designs deviating from the traditional case-control comparison (Hänzelmann, Castelo, and Guinney 2013). Tarca et al. compared several approaches including GSVA, GSEA and the traditional overrepresentation analysis (ORA) implemented in many tools such as the DAVID web tool (Huang, Sherman, and Lempicki 2009) and the R package GOStats (Falcon and Gentleman 2007). Interestingly, although ORA does not directly integrate gene expression values it can compete with methods doing so such as GSEA. Tarca et al. found the software tools PLAGE (Tomfohr, Lu, and Kepler 2005) and GLOBALTEST (Goeman et al. 2004) were performing best. Pathway Level Analysis of Gene Expression (PLAGE) uses singular value decomposition (SVD) to calculate a meta-gene representing the expression of a gene set. GLOBALTEST employs a logistic regression model to test the null hypothesis that all regression coefficients of all genes in a gene set are zero meaning that the genes from the gene set cannot predict the phenotype.

Many comparisons between methods of gene set analysis including those by Tarca et al. (Tarca, Bhatti, and Romero 2013) and Hänzelmann et al. (Hänzelmann, Castelo, and Guinney 2013) have favored differing approaches. Each publication introducing a new method usually tries to prove the superiority of the method proposed by the authors. Therefore, there is no clear favorite method superior to all others. Instead other issues such as the possibility to use the approach without gene expression values or the availability of a comprehensive database of gene sets such as in the DAVID tool (Huang, Sherman, and Lempicki 2009) may have more relevance. Another important issue are the annotations of datasets which often differ between sources hence demonstrating that there usually is not full clarity about which genes belong to a pathway and what roles they play in the pathway. Thus, further refinement of the topology of genes within a pathway or possibly the assignment of higher weights to more relevant genes would be other points which would have the potential to improve the quality of pathway analyses.

1.3.2.2 Gene ontologies

Gene ontologies (GOs) are a hierarchical classification system of genes which are grouped into functional entities describing either Biological Processes (BP), Cellular Components (CC) or Molecular Functions (MF) which are the top level GOs. These three top-level GOs are stepwise refined in a hierarchical structure, e.g. the GO BP is continuously detailed via the GOs *developmental process*, *anatomical structure development*, *multicellular organism development* to the GO *embryo development*. GOs have been introduced in 2000 and are maintained by the gene ontology consortium (Ashburner et al. 2000). Recently, models for causal relationships between GOs have been proposed - the so-called GO-CAMs (GO causal activity models) which can help to further elucidate how genes interact to provide biological functionality (Thomas et al. 2019). They substitute previous GO terms about positive or negative regulation by new causal relationships which may have positive or negative effect. This adds more detail to the presentation of genetic knowledge and may have the potential to enable more sophisticated approaches - possibly even artificial intelligence approaches - to get more mechanistic insight into genetic networks from gene expression data. Analysis of GOs has been traditionally made via the R/Bioconductor package GOStats - an ORA approach which can also account for the hierarchy of GOs (Falcon and Gentleman 2007) - but can also be achieved by most other tools used for gene set analysis such as DAVID (Huang, Sherman, and Lempicki 2009), ConsensuPathDB (Kamburov et al. 2011) or GSVA (Hänzelmann, Castelo, and Guinney 2013) or by specific GO analysis tools such as GOminer (Zeeberg et al. 2003), the Cytoscape plugin BINGO (Maere, Heymans, and Kuiper 2005) and REVIGO (Supek et al. 2011).

1.3.2.3 Pathways

Pathways are collections of genes that interact to achieve specific biological functions. The authors of the Reactome pathway database define a pathway as a grouping of reactions into causal chains (Joshi-Tope et al. 2005). Pathways can be further refined into metabolic e.g. Glycolysis, regulatory and signal transduction pathways, e.g. Insulin-signaling pathway. Several resources for pathway data exist including KEGG (Kanehisa et al. 2010), Reactome (Joshi-Tope et al. 2005) and WikiPathways (Slenter et al. 2018). Furthermore, there have been several approaches to integrate or collect data from multiple pathway databases such as the ConsensusPathDB (Kamburov et al. 2011), the DAVID functional annotation tool (Huang, Sherman, and Lempicki 2009) and the Molecular Signatures Database (MSigDB) at the Broad Institute which is often used in combination with the GSEA method (Subramanian et al. 2005). Pathway data from distinct sources may have considerable differences which in consequence can lead to deviating analysis results. These may emerge when the effects of the investigated data are small - however large effects will usually result in consistent results independent of the pathway resource. The classical pathway analysis approaches such as ORA and GSEA regard

all genes as equivalent and do not give genes playing central roles, e.g. MTOR in the MTOR-signaling pathway, higher weights than genes playing marginal roles. Development of approaches accounting for such weights may have the potential to improve the results. However, another solution to detect genes with high relevance in the topology of a pathway is to mark their expression or differential expression directly in the pathway map with tools such as PathView (Luo and Brouwer 2013).

1.3.2.4 Tissues

Particularly in the area of stem cell research it is relevant to be able to identify the tissue type of a cell. Pluripotent and also multipotent stem cells are often differentiated into an adult cell type for various purposes including disease models and therapeutic approaches. For instance, a disease model for NAFLD could be constructed by differentiating an iPSC into hepatocyte-like cells which subsequently would be challenged by oleic acid and possibly other disease-inducing factors. In the course of these experiments, it is of eminent importance to judge if the cell type one has generated is in fact congruent or at least sufficiently similar with the target cell type, in this case hepatocytes. Cell types have distinct gene expression profiles. Thus, one can use the genes expressed in cells of interest to determine their tissue type. Public databases such as PaGenBase (Pan et al. 2013) and the Genotype-Tissue Expression (GTEx) project (GTEx Consortium 2013) store associations of tissues with the genes expressed in them . Maturity of cell types often is an issue with cells derived from pluripotent stem cells because analogously to human development the differentiated cells are at a fetal stage and would years or even decades to arrive at the adult phenotype. If one studies Late-Onset-Alzheimer's Disease (LOAD) the neuronal cells of interest are usually over 65 years old what imposes a major challenge to the disease model. With respect to this point the employment of developmental datasets as provided by the Allen Brain Atlas are a feasible strategy (Jones, Overly, and Sunkin 2009). In principle, tissues can be detected the same way as GOs or pathways by ORA based on tissue-specific genesets but beyond that some tools have emerged for classification of tissue types including KeyGenes (Roost et al. 2015) and CellNet (Cahan et al. 2014). These also account for tissue-specific gene expression data.

1.3.2.5 Protein interaction networks

Proteins coded by genes are the central functional interactors in cells. Several databases have been built up of a plethora of protein interactions derived from experimental as well as computational sources. As examples here only the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database (Szklarczyk et al. 2015) and the BioGrid (Biological General Repository for Interaction Datasets) database (Oughtred et al.

2019) shall be mentioned. The STRING database gets protein interactions from primary databases, pathway databases, text-mining on Medline abstracts and full-text articles, prediction algorithms based on genomic information and coexpression analysis, orthologic transfer of interactions from other organisms (Szklarczyk et al. 2015) . STRING provides an intuitive user interface allowing to analyse interactions of single genes or sets containing multiple genes and delivers high-quality graphical presentations of the resulting networks. Networks can be expanded iteratively by a numer of interaction proteins in a way that initially not connected gene sets can be connected indirectly via interacting proteins.

BioGRID started as an open-source repository and in September 2018 has collected about 1.6 Mio protein, genetic and chemical interactions by text-mining. Interactions are well annotated facilitating retrieval of the data sources (Oughtred et al. 2019) . Recently, data from genome-wide CRISPR/Cas9 screens has been added to BioGRID.

Integration of gene expression data with protein interaction networks raises the problem to find optimal functional modules within the large protein interaction networks. Particularly in weighted graphs where vertices can have also negative weights, e.g. Pearson correlation coefficients from a gene expresssion experiment, identification of an optimal connected sub-network can be a not-trivial NP-hard mathematical problem. Ideker et al. first described this problem in the context of their cytoscape system and proposed heuristic approaches as solution (Ideker et al. 2002) . Dittrich et al. improved that solution by transforming the weighted graphs to Price-collecting Steiner trees (PCST) introducing a cost function for the edges (Dittrich et al. 2008). The cost function for the edges is initialized with the minimum of the weights in the graph. An optimal PCST can the be calculated with a time efficient algorithm. Furthermore, the approach also can deliver suboptimal sub-networks for visual inspection.

However, in many use cases protein interaction networks shall be generated from gene sets without associated weights. Besides STRING the java tool SubNet (Q. Zhang and Zhang 2013) offers this feature but different from STRING which uses its own database SubNet can work on arbitrary protein interaction databases provided by the user as tab-delimitted files. SubNet offers multiple extraction methods including extraction by shell, i.e. using interacting vertices connected with one edge (first shell), two edges (second shell) and further, (ii) using the shortest path between two vertices or (iii) a method similar to Google's PageRank working with the node centrality metric.

In publications associated with this work we used one approach combining elements from the extraction by shell and PageRank methods by using the first-level interactors with the n=30 most interactions to the gene set (Spitzhorn et al. 2019) and another one implementing a Shortest Path method (Wruck and Adjaye 2020) .

1.3.3 Transcription factor analysis

Transcription factors play an essential role in regulation of gene expression. Binding at transcription factor binding sites in the promoter or enhancer regions of genes they can initiate, enhance or repress transcription of the DNA sequence of a gene into mRNA. Promoter regions initiating transcription are close to the transcription start site while enhancer regions enhancing or repressing transcription may be located up to a few kB (kilo bases) upstream or downstream of it. Several transcription factors can be involved in the transcription of a gene and vice versa, one transcription factor can regulate a multitude of genes. Databases such as Transfaq (Heinemeyer et al. 1999) and JASPAR (Sandelin et al. 2004) store these often highly conserved relationships between transcription factors and genes for humans and other species. The classical *in vitro* experimental method to identify transcription factor binding is chromatin immunoprecipitation (ChIP) which precipitates transcription factors together with their DNA binding sites. The DNA sequences can be detected via microarrays or NGS (ChIP-seq) approaches. *In silico* transcription factor analysis can be performed by ORA or similar approaches using datasets of genes which are associated with the same transcription factors. The oPOSSUM-3 tool (Kwon et al. 2012) can be employed to perform such a transcription factor analysis using the Fisher-exact-test. Furthermore, the tool can account for the frequency of nucleotides in binding sites in the promoter regions of the investigated gene set and compare this to a background distribution of binding site nucleotides in the whole genome via a normal-approximation Z-score of the binomial distribution.

1.3.4 Time series analysis

Time series analysis is a discipline predominantly elaborated in other fields than bioinformatics such as economics and meteorology, e.g. for analysis of stock quotations or carbon dioxide measurements at Mauna Loa. However, also in bioinformatics several approaches for gene expression time series analysis have been proposed (Bar-Joseph 2004). In a review Bar-Joseph identifies several challenges including tackling missing data, sampling rates, synchronization, inference of regulatory networks and describes some proposed solutions such as interpolation between data points to cope with missing or noisy data and dynamic Bayesian networks to reconstruct regulatory networks (Bar-Joseph 2004). In general, time series analysis has to cope with cyclic effects such as the obvious seasonal local maxima and minima overlaying the major continuously rising carbon dioxide curve at the Mauna Loa or 24-hour cycles in genes associated with circadian rhythm. One solution for this problem is to use stationary curves resulting from derivations of the original curve. A test to check the stationarity of a curve is the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test which uses the null hypothesis that the curve is stationary. An important capability of time series analysis is the Granger test for causality (Granger 1969). The goal of this test is to determine causal relationships between time series, e.g. to solve the problem if the chicken or the egg came first by comparing time series of

chicken and egg production as was a bit ironically proposed in an essay about the Granger test (Thurman and Fisher 1988). The Granger test works by checking if a model using past lagged values of the time series itself and another putative causative time series is better predicting values of the time series than a model with only past values of the time series itself. In fact, that does not actually prove causality but shows that both features are related and one feature is preceding the other. To exclude that a third is responsible for the relation the Granger test in the opposite direction must not be significant.

1.4 Outline

In this cumulative thesis publications about meta-analyses and stem-cell-based disease models are brought together. Table 1 lists all included publications subdivided into two sections containing publications as first author (or equal contribution) and publications as co-author. The abstracts of 44 additional PubMed-listed publications are provided in the appendix. The first author publications predominantly cover meta-analyses or specialized analyses such as the multi-omics analysis of NAFLD and the phylogenetic analysis of SARS-CoV-2. Particular emphasis is laid on the two SARS-CoV-2-related first author publications which are highlighted in the table in bold letters. These represent an application of the methods developped so far to a rapidly emerging problem manifested in the Covid-19 pandemic. While the meta-analyses exploring NAFLD, AD and brain aging have counterparts in co-authored publications about iPSC-based disease models the SARS-CoV-2-related publications have built a foundation for future projects about stem-cell-based models of SARS-CoV-2 infection of multiple organs.

Table 1: Publications in this thesis

publications as first author or equal contribution	**journal/year**	**pubmed/DOI**
Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes	Sci Data. 2015	26646939
Meta-Analysis of Transcriptome Data Related to Hippocampus Biopsies and iPSC-Derived Neuronal Cells from Alzheimer's Disease Patients Reveals an Association with FOXA1 and FOXA2 Gene Regulatory Networks	J Alzheimers Dis. 2016	26890743
Concise Review: Current Status and Future Directions on Research Related to Nonalcoholic Fatty Liver Disease	Stem Cells. 2017	27374784
Meta-analysis reveals up-regulation of cholesterol processes in nonalcoholic and down-regulation in alcoholic fatty liver disease	World J Hepatol. 2017	28357032
Nijmegen Breakage Syndrome fibroblasts and iPSCs: cellular models for uncovering disease-associated signaling pathways and establishing a screening platform for anti-oxidants	Sci Rep. 2017	28790359
Human pluripotent stem cell derived HLC transcriptome data enables molecular dissection of hepatogenesis	Sci Data. 2018	29533390
The FGF, TGF and WNT axis Modulate Self-renewal of Human SIX2+ Urine Derived Renal Progenitor Cells	Sci Rep. 2020	31959818
Meta-analysis of human prefrontal cortex reveals activation of GFAP and decline of synaptic transmission in the aging brain	Acta Neuropathol Commun. 2020	32138778
SARS-CoV-2 receptor ACE2 is co-expressed with genes related to transmembrane serine proteases, viral entry, immunity and cellular stress	Sci Rep. 2020	33293627
Transmission of SARS-COV-2 from China to Europe and West Africa: a detailed phylogenetic analysis	bioRxiv 2020	DOI: 10.1101/ 2020.10.02.323519
publications as contributing author in this thesis		
Induced pluripotent stem cell-derived neuronal cells from a sporadic Alzheimer's disease donor as a model for investigating AD-associated gene regulatory networks	BMC Genomics. 2015	25765079
Footprint-free human fetal foreskin derived iPSCs: A tool for modeling hepatogenesis associated gene regulatory networks	Sci Rep. 2017	28740077
Modeling Nonalcoholic Fatty Liver Disease with Human Pluripotent Stem Cell-Derived Immature Hepatocyte-Like Cells Reveals Activation of PLIN2 and Confirms Regulatory Functions of Peroxisome Proliferator-Activated Receptor Alpha	Stem Cells Dev. 2016	27308945
IPSC-Derived Neuronal Cultures Carrying the Alzheimer's Disease Associated TREM2 R47H Variant Enables the Construction of an Aβ-Induced Gene Regulatory Network	Int J Mol Sci. 2020	32630447
44 other PubMed-listed publications as first or contributing author listed in the appendix of this thesis		

Chapter 2

Results

2.1 Publications as first author or equal contribution

2.1.1 Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes.

Sci Data. 2015 Dec 8;2:150068. doi: 10.1038/sdata.2015.68.

Wruck W(1), Kashofer K(2), Rehman S(3), Daskalaki A(4), Berg D(5), Gralka E(6), Jozefczuk J(4), Drews K(4), Pandey V(4), Regenbrecht C(7), Wierling C(4), Turano P(6), Korf U(5), Zatloukal K(2), Lehrach H(4), Westerhoff HV(3)(8)(9), Adjaye J(1)(4).

Author information: (1)Medical Faculty, Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, 40225 Düsseldorf, Germany. (2)Institute of Pathology, Medical University of Graz, Graz 8036, Austria. (3)The Manchester Centre for Integrative Systems Biology, Manchester Institute of Biotechnology, University of Manchester, Manchester M1 7DN, UK. (4)Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestrasse 63, Berlin 14195, Germany. (5)Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. (6)Magnetic Resonance Center (CERM), University of Florence, 50019 Florence, Italy. (7)Institute for Pathology &Comprehensive Cancer Center, Cancer Stem Cell Group, Charité-Universitätsmedizin, 10117 Berlin, Germany. (8)Netherlands Institute for Systems Biology, VU University Amsterdam, HV NL-1081 Amsterdam, The Netherlands. (9)Synthetic Systems Biology, Swammerdam Institute for Life Sciences, University of Amsterdam, 1018 WS Amsterdam, The Netherlands.

Non-alcoholic fatty liver disease (NAFLD) is a consequence of sedentary life style and high fat diets with an estimated prevalence of about 30% in western countries. It is associated with insulin resistance, obesity, glucose intolerance and drug toxicity. Additionally, polymorphisms within, e.g., APOC3, PNPLA3, NCAN, TM6SF2 and PPP1R3B, correlate with NAFLD. Several studies have already investigated later stages of the disease. This study explores the early steatosis stage of NAFLD with the aim of identifying molecular mechanisms underlying the etiology of NAFLD. We analyzed liver biopsies and serum samples from patients with high- and low-grade steatosis (also pre-disease states) employing transcriptomics, ELISA-based serum protein analyses and metabolomics. Here, we provide a detailed description of the various related datasets produced in the course of this study. These datasets may help other researchers find new clues for the etiology of NAFLD and the mechanisms underlying its progression to more severe disease states.

DOI: 10.1038/sdata.2015.68
PMCID: PMC4672680
PMID: 26646939 [Indexed for MEDLINE]
URL: https://www.nature.com/articles/sdata201568

Contribution: Wasco Wruck performed transcriptomics analyses, was responsible for data management and wrote the manuscript.

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Non-alcoholic fatty liver disease
» Metabolomics
» Gene expression analysis
» Systems biology

Received: 16 March 2015
Accepted: 20 October 2015
Published: 8 December 2015

Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes

Wasco Wruck[1], Karl Kashofer[2], Samrina Rehman[3], Andriani Daskalaki[4], Daniela Berg[5], Ewa Gralka[6], Justyna Jozefczuk[4], Katharina Drews[4], Vikash Pandey[4], Christian Regenbrecht[7], Christoph Wierling[4], Paola Turano[6], Ulrike Korf[5], Kurt Zatloukal[2], Hans Lehrach[4], Hans V. Westerhoff[3,8,9] & James Adjaye[1,4]

Non-alcoholic fatty liver disease (NAFLD) is a consequence of sedentary life style and high fat diets with an estimated prevalence of about 30% in western countries. It is associated with insulin resistance, obesity, glucose intolerance and drug toxicity. Additionally, polymorphisms within, e.g., *APOC3, PNPLA3, NCAN, TM6SF2* and *PPP1R3B*, correlate with NAFLD. Several studies have already investigated later stages of the disease. This study explores the early steatosis stage of NAFLD with the aim of identifying molecular mechanisms underlying the etiology of NAFLD. We analyzed liver biopsies and serum samples from patients with high- and low-grade steatosis (also pre-disease states) employing transcriptomics, ELISA-based serum protein analyses and metabolomics. Here, we provide a detailed description of the various related datasets produced in the course of this study. These datasets may help other researchers find new clues for the etiology of NAFLD and the mechanisms underlying its progression to more severe disease states.

Design Type(s)	parallel group design • transcription profiling by array design • protein expression profiling
Measurement Type(s)	transcription profiling assay • protein expression profiling • metabolite profiling
Technology Type(s)	microarray platform • ELISA • nuclear magnetic resonance assay
Factor Type(s)	disease stage • obesity
Sample Characteristic(s)	Homo sapiens • serum • liver

[1]Medical Faculty, Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, 40225 Düsseldorf, Germany. [2]Institute of Pathology, Medical University of Graz, Graz 8036, Austria. [3]The Manchester Centre for Integrative Systems Biology, Manchester Institute of Biotechnology, University of Manchester, Manchester M1 7DN, UK. [4]Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestrasse 63, Berlin 14195, Germany. [5]Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [6]Magnetic Resonance Center (CERM), University of Florence, 50019 Florence, Italy. [7]Institute for Pathology & Comprehensive Cancer Center, Cancer Stem Cell Group, Charité—Universitätsmedizin, 10117 Berlin, Germany. [8]Netherlands Institute for Systems Biology, VU University Amsterdam, HV NL-1081 Amsterdam, The Netherlands. [9]Synthetic Systems Biology, Swammerdam Institute for Life Sciences, University of Amsterdam, 1018 WS Amsterdam, The Netherlands. Correspondence and requests for materials should be addressed to J.A. (email: James.Adjaye@med.uni-duesseldorf.de).

Background & Summary

With an estimated prevalence of about 30% in western countries, NAFLD is a major public health issue[1]. Sedentary life-style and excessive food consumption correlate with rate at which NAFLD cases appear. Epidemiologic studies showing a prevalence of the disease that differs between countries as well as between groups in the same country, appear to reflect an interplay of environmental and genetic factors in its etiology[1]. Additionally, polymorphisms in, e.g., *APOC3, PNPLA3, NCAN, TM6SF2 and PPP1R3B*, correlate with NAFLD[2,3]. Over-feeding directly induces insulin resistance[4]. Causality between steatosis and the metabolic syndrome of insulin resistance, obesity, and glucose intolerance, is still unresolved[5]. While the correlation between steatosis and insulin resistance is established there is debate about the relationship between steatosis and hepatic insulin resistance[6]. Samuel *et al.* showed that activated *PKC-ϵ* and *JNK* can induce insulin resistance via impaired *IRS1* and *IRS2* tyrosine phosphorylation in rats fed with high fat diet[7]. An investigation on the insulin-like growth factor (IGF) axis in the Nurses' Health Study[8] and another population study of 3863 people[9] addressed connections between the IGF axis, insulin resistance, diabetes risk and NAFLD. *IGFBP3* is associated with various cancers and up-regulation of *IGF1* receptor (*IGF1R*) is considered an early event in hepatocarcinogenesis[10]. Thus, the IGF axis might play an important role in a direct development of carcinoma from steatosis without the formerly assumed intermediary phase of cirrhosis[11].

The progression of NAFLD from mild steatosis up to severe steatohepatitis and even liver cirrhosis and hepatocellular carcinoma, varies widely between individual patients. Insulin resistance, dysregulation of cytokines as a basis for inflammation, and oxidative stress appear to foster progression to steatohepatitis[12]. A two-step progression from simple steatosis to steatohepatitis and fibrosis has been proposed[13], and suggests that after fat accumulation in the liver due to insulin resistance, lipids are peroxidized with cytokines and Fas ligand induced by excessive ROS. However, this two-step progression has been questioned[5]. We found that in fibroblasts derived from steatosis patients *AKT/mTOR* signaling was reduced and that the insulin-resistant phenotype is exhibited not only by insulin-metabolizing central organs, e.g., the liver, but also by skin fibroblasts[14]. Transcriptome data identified a regulatory network orchestrated by the transcription factor *SREBF1* and linked to a metabolic network of glycerolipid and fatty acid biosynthesis. The downstream transcriptional targets of *SREBF1* which include the phosphatidic acid phosphatase *LPIN1* and *LDLR*, were also involved.

Moreover, there is the possible involvement of ROS in disease progression. Houstis *et al.*[15] demonstrated that oxidative stress can induce insulin resistance and that anti-oxidants may ameliorate insulin resistance. Depletion of glutathione can improve insulin sensitivity in mice[16]. Glutathione is known as the body's master antioxidant, protecting cells against damage caused by numerous reactive intermediates[17]. Detoxification of these reactive metabolites results in the consumption of glutathione either via oxidation or conjugation. Maintenance of the intracellular glutathione level is thereby a critical liver function, which could be impaired following insult/injury or in steatosis and steatohepatitis.

Several other studies exploring various aspects of NAFLD have been published. A recent publication by Moylan *et al.* showed that it is possible to discriminate mild versus severe fibrosis stages of NAFLD patients via their gene expression profiles[18]. Another study from Speliotes *et al.* investigated NAFLD via a genome-wide association study (GWAS) approach[3]. Besides the most prominent association of *PNPLA3* this study reported several other associations including one at locus 19p13.11 which is in strong linkage disequilibrium with a recently found steatosis-linked polymorphism in *TM6SF2*, transmembrane6 superfamily member 2 (refs 19,20) . A knockdown of TM6SF2 in human hepatoma cell lines and in mice led to an increase in lipid droplet area while overexpression led to a decrease[19].

Interestingly, the above mentioned genes associated with NAFLD in GWAS were not detected in a large-scale GWAS about obesity and insulin biology although the metabolic syndrome connects NAFLD and obesity[21]. Feldstein *et al.* found CK-18 as a non-invasive biomarker for NASH by comparison of plasma samples from patients with biopsy proven NAFLD[22]. Du Plessis *et al.* used analysis results from subcutaneous and visceral fat and liver biopsies to construct a model which predicts NAFLD liver histology[23]. This model involves the genes *CCL2, DMRT2, GADD45B, IL1RN*, and *IL8*. In contrast to the studies of Moylan *et al.* and Feldstein *et al.* our study highlights potential means of classifying distinct grades of Steatosis in NAFLD—the very early stage of the disease. Although it is evident that a complex interplay of genetic and environmental factors contribute to the development of steatosis, to date there has not been a systemic study of the disease employing a multi-omic approach- transcriptome, ELISA-based proteome and metabolome. Therefore, the intention of this study is to provide a more comprehensive view of steatosis based on transcriptomic, metabolomic and protein biomarker profiles. Additionally, this should lay down the foundation for follow-up systems biology-based studies.

In the current study we analyzed patient liver biopsies and associated serum samples, from patients with the insulin resistance phenotype confirmed by the HOMA-IR model[24]. Here, we describe these valuable data sets deposited in public repositories, which might support other researchers in identifying new clues for the etiology of NAFLD and the mechanisms underlying its progression to more severe disease states.

ID	gender	Age	BMI	% steatosis	grouping by pathologist	steatosis grouping	medical centre	liver illumina array rep.1 (GSE46300)	liver illumina array rep.2 (GSE46300)	serum NMR data	serum ELISA data
H0004	f	54	47	10%		obese, low steatosis	Graz (Austria)	GSM1128362	GSM1128363	MTBLS174	10.6084/m9.figshare.1333564
H0007	f	33	51	40%		obese, high steatosis	Graz (Austria)	GSM1128364	GSM1128365	MTBLS174	10.6084/m9.figshare.1333564
H0008	m	61	46	40%	obese, high steatosis	obese, high steatosis	Graz (Austria)	GSM1128366	GSM1128367	MTBLS174	10.6084/m9.figshare.1333564
H0009	f	48	49	5–10%	obese, low steatosis	obese, low steatosis	Graz (Austria)	GSM1128368	GSM1128369	MTBLS174	10.6084/m9.figshare.1333564
H0011	f	58	45	70%	obese, high steatosis	obese, high steatosis	Graz (Austria)	GSM1128370	GSM1128371	MTBLS174	10.6084/m9.figshare.1333564
H0012	f	50	35	0	obese, low steatosis	obese, low steatosis	Graz (Austria)	GSM1128372	GSM1128373	no	no
H0018	f	35	41	30–40%	obese, high steatosis	obese, high steatosis	Graz (Austria)	GSM1128374	GSM1128375	MTBLS174	10.6084/m9.figshare.1333564
H0021	m	49	41	0%		no steatosis	Graz (Austria)	GSM1128376	GSM1128377	MTBLS174	10.6084/m9.figshare.1333564
H0022	m	45	49	40%		obese, high steatosis	Graz (Austria)	GSM1128378	GSM1128379	MTBLS174	10.6084/m9.figshare.1333564
H0024	m	29	44	50%		obese, high steatosis	Graz (Austria)	no	no	MTBLS174	10.6084/m9.figshare.1333564
H0025	f	53	46	15–20%		obese, low steatosis	St Gallen (Switzerland)	no	no	MTBLS174	10.6084/m9.figshare.1333564
H0026	f	46	39	0%		no steatosis	St Gallen (Switzerland)	no	no	MTBLS174	10.6084/m9.figshare.1333564
H0027	m	44	42	50%		obese, high steatosis	St Gallen (Switzerland)	no	no	MTBLS174	10.6084/m9.figshare.1333564
H0028	f	28	43	20%		obese, low steatosis	St Gallen (Switzerland)	no	no	MTBLS174	10.6084/m9.figshare.1333564
H0029	f	40	39	< 5%		no steatosis	St Gallen (Switzerland)	no	no	MTBLS174	10.6084/m9.figshare.1333564
H0030	m	22	45	30%		obese, low steatosis	St Gallen (Switzerland)	no	no	MTBLS174	10.6084/m9.figshare.1333564
H0031	m	22	41	0%		no steatosis	St Gallen (Switzerland)	no	no	MTBLS174	10.6084/m9.figshare.1333564
H0033	f	44	43	40%		obese, high steatosis	St Gallen (Switzerland)	no	no	MTBLS174	10.6084/m9.figshare.1333564
H0034	m	50	42	10%		obese, low steatosis	St Gallen (Switzerland)	no	no	MTBLS174	10.6084/m9.figshare.1333564

Table 1. Samples related to data sets in repositories (Data Citations 1–Data Citation 3).

Methods

Patient recruitment, sample collection and clinical measurements

All patients participating in this study were recruited in the Multidisciplinary Obesity Research (MORE) project at the Medical University of Graz, Austria or at the Interdisciplinary Adipositas Center at the Kantonsspital St Gallen, Switzerland. Patients with morbid obesity who admitted into hospital for treatment by bariatric surgery (gastric banding, gastric bypass, sleeve gastrectomy) were invited to participate in the study and to sign the informed consent. The study was approved by the institutional review board of the Medical University of Graz (reg. IRB00002556 at the Office for Human Research Protections of the US Departments of Health and Human Services) under license 20–143 ex 08/09. All experiments were performed in accordance with approved guidelines. Written informed consent was obtained from all participants. In the course of the bariatric surgery, samples of blood, skin and a liver biopsy were taken. Out of 18 patients (Table 1), 9 liver biopsies were of high quality enabling their use in the transcriptome analyses. Serum plasma was available from all the patients. The overall experimental design of this study is illustrated in Fig. 1. A pathological diagnosis of the liver phenotype, including liver steatosis grading based on H&E morphology, was performed by an experienced, board certified pathologist (CL). We simplified Kleiner's scoring scheme by condensing Steatosis grades 2 (34–66%) and 3 (> 66%) to our 'high-grade' while adopting grades 0 ('none') and 1 ('low')[25]. This simplification was made because the inter-patient-variability in this complex heterogeneous disease did not allow a more detailed grading on the omics levels. Two examples of liver biopsies are shown in Fig. 2a.

Illumina bead chip hybridization and data analysis

Microarray experiments were carried out on the Illumina BeadStation 500 platform (Illumina, San Diego, CA, USA). Briefly, 500 ng DNase-treated total RNA were used as input for amplification and biotin labeling reactions (Illumina TotalPrep RNA Amplification Kit, Ambion) prior to hybridization of the resulting cRNAs onto Illumina HumanHT-12_v4_BeadChips, washing, Cy3-streptavidin staining and scanning according to the manufacturer's instructions.

Transciptomics data analysis

Illumina data was processed via R/Bioconductor[26] and packages lumi[27], limma[28] and biomaRt. Background-corrected log2-transformed data was normalized via quantile normalization from the lumi package.

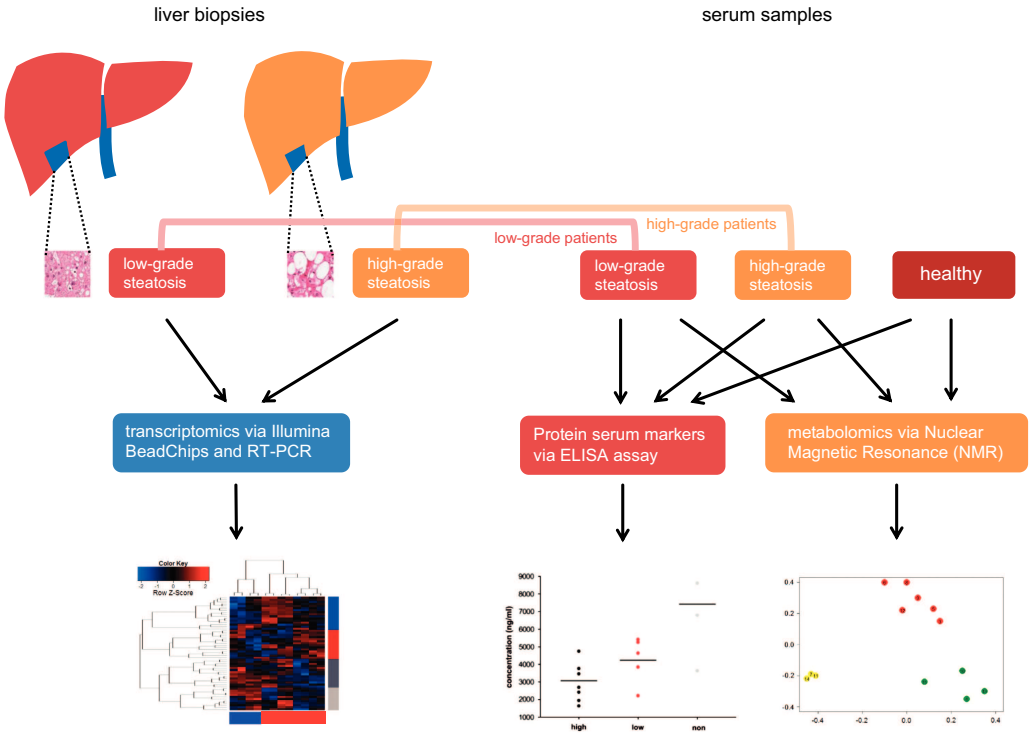


Figure 1. Scheme of experiments for multi-omics comparison of steatosis grades. The scheme shows how the distinct severities of non-alcoholic fatty liver disease (NAFLD) are compared in terms of transcriptomics, metabolomics and potentially relevant parts of the proteome. Liver biopsies were taken from NAFLD patients and classified by pathologists as low-grade (5–33% steatosis area) and high-grade (>33% steatosis area). The transcriptome of liver biopsies were assessed on Illumina HumanHT-12 v4 BeadChips and on RT-PCR. Serum samples of these NAFLD patients and from healthy persons were taken and investigated at the protein level employing ELISA assays and at the metabolome level via Nuclear Magnetic Resonance (NMR).

qRT-PCR

Quantitative real-time polymerase chain reaction (qRT-PCR) was carried out to confirm the microarray-derived data. Reactions were carried out on the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems). Data analysis was carried out using the ABI PRISM SDS 2.2.1 software (Applied Biosystems) and Microsoft Excel (Microsoft Corporation, Redmond, WA, USA). GAPDH-normalized, relative mRNA levels of each gene (high steatosis versus low steatosis) were calculated based on the 2-ΔΔCT Method. Primer sequences for QRT-PCR validation are described in Table 2.

ELISA-based assay for biomarkers

ELISA measurements from plasma samples were carried out using the Ciraplex platform (Aushon Biosystems, Billerica, MA, US). Commercial assays were purchased and measurements were carried out according to instructions provided by the manufacturer. The following 29 targets were analyzed either as single-plex assays or as multiplex assay: hFGFb; hGROa; hLIF;hIFNg; hIL1b; IL4; IL5; hIL6; hIL10; hIL12p70; hIL13; hTNFa; hI309; hIL8; hIP10; hMCP4; hMIP1a; hMIP1b; hCRP; hLeptin; hPAIactive; hResistin; hIGFBP1; hIGFBP3; hIGFBP2; hMIF; hApoA1; hCRP; hAcrp30.

NMR sample preparation

Frozen plasma samples were thawed at room temperature and shaken before use. According to standard methodologies a total of 300 µl of buffer (70 mM Na2HPO4; 20% (v/v) D2O; 6.15 mM NaN3; 6.64 mM TMSP; pH 7.4) was added to 300 µl of each serum sample. A total of 450 µl of this mixture was transferred into a 4.25 mm NMR tubes (Bruker BioSpin) for analysis.

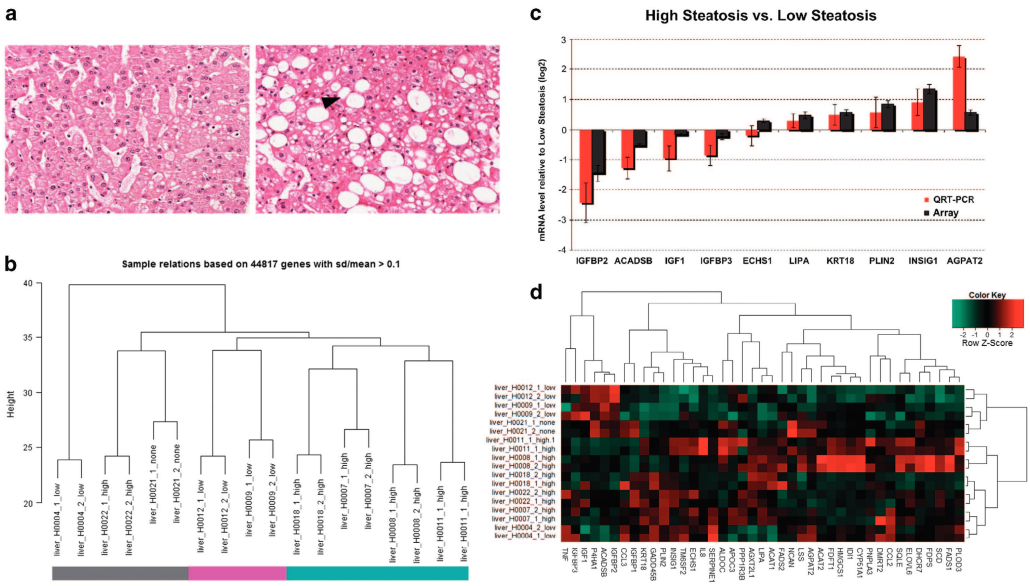


Figure 2. Histopathological and transcriptome characterization of liver tissue. (**a**) Liver tissue with only marginal pathological changes (H9, low-grade steatosis group). The hepatocytes are arranged in one cell thick plates, separated by sinusoids. They contain only few small isolated fat valuoles (H&E stained section). Hepatocytes of the intermediate and central lobular areas contain macrovesicular fat (image to the right, H8, steatosis group, hepatocytes with fatty change are indicated by arrow heads; H&E stained section). (**b**) Hierachical clustering of the transcriptomes of patient liver samples. We identified three clusters: high (>33%) steatosis (cyan), low (5–33%) steatosis (magenta) and heterogeneous clusters of high, low and no steatosis (grey). (**c**) Quantitative QRT-PCR confirmation of genes differentially expressed in high versus low steatotic livers. The columns represent the mean of four biological replicates (high steatosis) versus two biological replicates (low steatosis). Error bars indicate standard errors of the mean. Array-derived and RT-PCR-derived columns are depicted in dark grey and red respectively. (**d**) Heatmap of genes differentially expressed in high versus low steatotic livers and genes found in literature and in genome-wide association studies.

NMR spectra acquisition and processing

NMR spectra from 18 plasma samples from morbidly obese patients that underwent different type of bariatric surgery and additionally have developed steatosis were collected (Table 1). 1H-NMR spectra were acquired using a Bruker spectrometer (Bruker Biospin). Unsupervised and supervised methods were used in order to identify a disease-related metabolomic profile that might contain a signature of steatosis.

Data Records

Data record 1

The microarray experiments discussed in this publication were carried out on the Illumina BeadStation 500 platform (Illumina, San Diego, CA, USA). The data have been deposited in NCBI's GEO and are accessible through GEO Series accession number GSE46300 (Data Citation 1).

Data record 2

Metabolomic raw data from Nuclear magnetic resonance (NMR) measurements have been deposited at the MetaboLights database (http://www.ebi.ac.uk/metabolights) of the European Bioinformatics Institute (EBI) under MTBLS174 (Data Citation 2).

Gene	Fwd	Rev	Product size
ACADSB	CACCATTGCAAAGCATATCG	GCAAGGCACTTACTCCCAAC	117
AGPAT2	GGGGCGTCTTCTTCATCA	TTGAGGTTCTCCCTGACCAT	91
ECHS1	AACCTTTGCCACTGATGACC	CAAGCAGAGGTGTGAAGCAG	112
IGF1	TGCAGGAGGGACTCTGAAAC	AGCTGCGTGATATTTGAAAGG	111
IGFBP2	CTCCCTGCCAACAGGAACTG	TCTTGCACTGTTTGAGGTTGTACAG	147
IGFBP3	CAACTGTGGCCATGACTGAG	CCTGACTTTGCCAGACCTTC	92
INSIG1	CAACACCTGGCATCATCG	CTCGGGGAAGAGAGTGACAT	118
KRT18	GAGGTTGGAGCTGCTGAGAC	CAAGCTGGCCTTCAGATTTC	99
LIPA	CATCTGTGTGAAGCCAAAGC	AATCCCTGAGCTGAGTTTGC	112
PLIN2	GCTGAGCACATTGAGTCACG	TGGTACACCTTGGATGTTGG	102

Table 2. Primer sequences for QRT-PCR validation of genes differentially expressed between high-grade and low-grade steatosis.

sample	liver_H0007_1_high	liver_H0007_2_high	liver_H0008_1_high	liver_H0008_2_high	liver_H0011_1_high	liver_H0011_1_high	liver_H0018_1_high	liver_H0018_2_high	liver_H0022_1_high	liver_H0022_2_high	liver_H0004_1_low	liver_H0004_2_low	liver_H0009_1_low	liver_H0009_2_low	liver_H0012_1_low	liver_H0012_2_low	liver_H0021_1_none	liver_H0021_2_none
liver_H0007_1_high	1.0000	0.9944	0.9861	0.9859	0.9892	0.9891	0.9910	0.9905	0.9878	0.9871	0.9851	0.9858	0.9895	0.9914	0.9882	0.9878	0.9889	0.9889
liver_H0007_2_high	0.9944	1.0000	0.9893	0.9886	0.9925	0.9924	0.9921	0.9917	0.9908	0.9910	0.9866	0.9865	0.9901	0.9923	0.9918	0.9910	0.9922	0.9930
liver_H0008_1_high	0.9861	0.9893	1.0000	0.9965	0.9911	0.9912	0.9881	0.9878	0.9924	0.9926	0.9855	0.9847	0.9831	0.9864	0.9911	0.9911	0.9897	0.9908
liver_H0008_2_high	0.9859	0.9886	0.9965	1.0000	0.9905	0.9904	0.9875	0.9868	0.9915	0.9922	0.9842	0.9840	0.9832	0.9857	0.9910	0.9910	0.9891	0.9902
liver_H0011_1_high	0.9892	0.9925	0.9911	0.9905	1.0000	0.9979	0.9911	0.9910	0.9922	0.9920	0.9862	0.9860	0.9863	0.9891	0.9928	0.9915	0.9918	0.9927
liver_H0011_1_high	0.9891	0.9924	0.9912	0.9904	0.9979	1.0000	0.9910	0.9908	0.9922	0.9922	0.9860	0.9856	0.9859	0.9886	0.9931	0.9919	0.9917	0.9927
liver_H0018_1_high	0.9910	0.9921	0.9881	0.9875	0.9911	0.9910	1.0000	0.9960	0.9916	0.9909	0.9896	0.9895	0.9919	0.9927	0.9910	0.9905	0.9915	0.9917
liver_H0018_2_high	0.9905	0.9917	0.9878	0.9868	0.9910	0.9908	0.9960	1.0000	0.9912	0.9910	0.9892	0.9891	0.9900	0.9916	0.9905	0.9905	0.9902	0.9907
liver_H0022_1_high	0.9878	0.9908	0.9924	0.9915	0.9922	0.9922	0.9916	0.9912	1.0000	0.9979	0.9871	0.9866	0.9861	0.9887	0.9921	0.9920	0.9928	0.9939
liver_H0022_2_high	0.9871	0.9910	0.9926	0.9922	0.9920	0.9922	0.9909	0.9910	0.9979	1.0000	0.9869	0.9861	0.9855	0.9887	0.9924	0.9925	0.9928	0.9939
liver_H0004_1_low	0.9851	0.9866	0.9855	0.9842	0.9862	0.9860	0.9896	0.9892	0.9871	0.9869	1.0000	0.9974	0.9889	0.9908	0.9834	0.9830	0.9867	0.9867
liver_H0004_2_low	0.9858	0.9865	0.9847	0.9840	0.9860	0.9856	0.9895	0.9891	0.9866	0.9861	0.9974	1.0000	0.9890	0.9910	0.9830	0.9831	0.9864	0.9858
liver_H0009_1_low	0.9895	0.9901	0.9831	0.9832	0.9863	0.9859	0.9919	0.9900	0.9861	0.9855	0.9889	0.9890	1.0000	0.9961	0.9854	0.9854	0.9877	0.9876
liver_H0009_2_low	0.9914	0.9923	0.9864	0.9857	0.9891	0.9886	0.9927	0.9916	0.9887	0.9887	0.9908	0.9910	0.9961	1.0000	0.9885	0.9885	0.9904	0.9905
liver_H0012_1_low	0.9882	0.9918	0.9911	0.9910	0.9928	0.9931	0.9910	0.9905	0.9921	0.9924	0.9834	0.9830	0.9854	0.9885	1.0000	0.9971	0.9920	0.9931
liver_H0012_2_low	0.9878	0.9910	0.9911	0.9910	0.9915	0.9919	0.9905	0.9905	0.9920	0.9925	0.9830	0.9831	0.9854	0.9885	0.9971	1.0000	0.9915	0.9924
liver_H0021_1_none	0.9889	0.9922	0.9897	0.9891	0.9918	0.9917	0.9915	0.9902	0.9928	0.9928	0.9867	0.9864	0.9877	0.9904	0.9920	0.9915	1.0000	0.9968
liver_H0021_2_none	0.9889	0.9930	0.9908	0.9902	0.9927	0.9927	0.9917	0.9907	0.9939	0.9939	0.9867	0.9858	0.9876	0.9905	0.9931	0.9924	0.9968	1.0000

Table 3. Pearson correlation coefficients of transcriptome data of all samples versus each other.

Data record 3

ELISA measurements have been deposited at figshare (http://www.figshare.com) (Data Citation 3).

Technical Validation

Transcriptomic data

Microarray data passed the proprietary Illumina quality controls. All samples were investigated in duplicates. Fig. 2b shows that—as would be expected—the duplicates cluster together demonstrating the validity of experiments in terms of whole-genome gene expression. The Pearson correlation coefficients of all samples versus each other were calculated with the intention to detect outliers. However, all correlation coefficients were greater than 0.98 and all correlation coefficients of duplicates even greater than 0.99 so that all samples passed this quality check (Table 3). Genes with significant differential gene expression were selected for validation via RT-PCR experiments (Fig. 2c). Genes were termed differentially expressed if the multiple-testing-corrected limma[28] *P*-value was less than 0.05, the ratio was less than 0.75 or greater than 1.33 and the gene was expressed (detection *P*-value less than 0.05) in at least one of both cases. Furthermore, we analysed clusters of genes differentially expressed in high versus low steatotic livers together with genes found in literature[19,23] and genome-wide association studies[3] (Fig. 2d). This analysis confirms high similarity between duplicates and clustering—to some extent but not fully—according to steatosis grade. Fig. 3a shows a plot of the first two components of the Principal Component Analysis (PCA) of the microarray data.

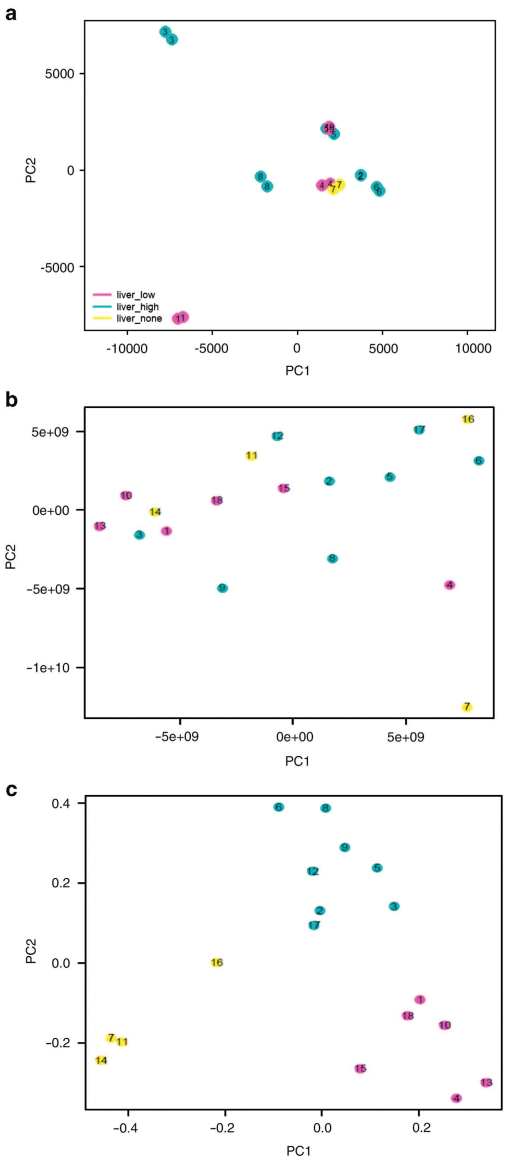


Figure 3. Transcriptomic and metabolomic profiles. (**a**–**c**) Transcriptomics and metabolomics PCA plots. Distinct colours are used to aid visualizing patients with distinct levels of steatosis: yellow, patients with < 5% no steatosis; magenta, patients with 5–33%, low level steatosis; cyan, patients with high steatosis >33%, high steatosis. (**a**) Unsupervised PCA plot for 18 liver biopsies, Illumina microarray data. (**b**) Unsupervised PCA plot for 18 plasma samples, metabolomics data. (**c**) Supervised discrimination analysis (pls/ca: partial least squares/canonical analysis) of metabolites in patient plasma samples. The correspondence between numbers in the plot and sample names in Table 1 is: 1 = H0004, 2 = H0007, 3 = H0008, 4 = H0009, 5 = H0011, 6 = H0018, 7 = H0021, 8 = H0022, 9 = H0024, 10 = H0025, 11 = H0026, 12 = H0027, 13 = H0028, 14 = H0029, 15 = H0030, 16 = H0031, 17 = H0033, 18 = H0034, 19 = H0012.

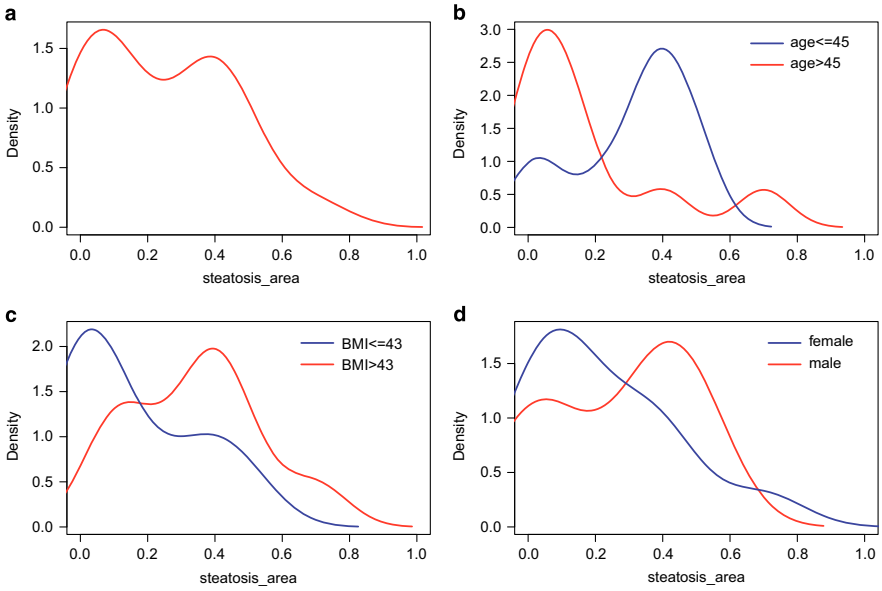


Figure 4. Distribution plots of percentage parenchymal involvement in steatosis. (**a**) all patients. (**b**) Kernel density plot of patients above/below median age (median = 45). (**c**) Kernel density plot of patients above/below median BMI (median = 43). (**d**) Kernel density plot of male/female patients.

ELISA-based assay for biomarkers

Samples below and above quantification limit as well as samples with coefficient of variation (cv) greater than 20% were marked in the measurements table (Data Citation 3). Independent validation of the ELISA-based measurements was checked by visual inspection of plots comparing disease states.

Metabolomics

Assignment of all metabolites were done manually, signals were assigned on template one-dimensional NMR profiles by using matching routines of AMIX 7.3.2 (Bruker BioSpin) in combination with the BBIOREFCODE Version 2-0-0 reference database and published literature when available. Additional confirmation was done using data provided in our lab -database containing spectra of standard pure compounds. To assess which metabolites (i.e., NMR peaks) were significantly different between different sets a univariate paired Wilcoxon test was used. A P-value ≤ 0.05 was considered statistically significant (P-value not corrected for multiple testing).

Robust validation of statistical analysis results was done using a cross-validation technique.

The accuracy of the classification was assessed by means of a single cross-validation scheme. The original data set was split into a training set (80% of the samples) and a test set (20% of the samples) prior to any step of statistical analysis. The number of PLS components was chosen on the basis of a 5-fold cross validation performed on the training set only, and the best model was used to predict the samples in the test set. The whole procedure was repeated 200 times with a Monte Carlo cross validation scheme, and the results averaged.

Figure 3b shows a plot of unsupervised discrimination analysis and Fig. 3c shows separation of steatosis grades in a plot of supervised discrimination analysis (pls/ca: partial least squares/canonical analysis) of metabolites in patient plasma samples. The clustering of Fig. 3c results from a supervised PLS/CA based only on the metabolomic NMR profiles. The algorithm takes into account the supervised information relative to the 3 steatosis groups.

Distribution plots

Figure 4a shows the distribution of percentage parenchymal involvement in steatotic patients derived from Table 1. The percentage is converted to a scale from zero to one and plotted with the kernel density function from the R statistical package. Fig. 4b–d display distributions separated into groups of age above/below median (median = 45), body mass index (BMI) above/below median (median = 43) and gender. Fig. 4d would suggest a slight tendency for more severe steatosis in males. A similar trend has

been reported in a NAFLD study on Australian adolescents where 3.1% of males and only 2.2% of females had moderate to severe steatosis while 7.0% of males and 14.1% of females had mild steatosis[29].

Usage Notes

All patients in this study underwent bariatric surgery. This should be taken into account when generalizing results although these are typical cases of morbid obesity which is connected to the metabolic syndrome including NAFLD. The sample size of these datasets—in particular the transcriptomics dataset—poses certain limits onto its usage. Due to its small size it will not enable rigorous analysis of gender effects. Therefore it would likely need to be combined with other data sources, such as data from Moylan *et al.*[18] and Du Plessis *et al.*[23].

References

1. Lazo, M. & Clark, J. M. The epidemiology of nonalcoholic fatty liver disease: a global perspective. *Semin. Liver Dis.* **28,** 339–350 (2008).
2. Petersen, K. F. *et al.* Apolipoprotein C3 Gene Variants in Nonalcoholic Fatty Liver Disease. *N. Engl. J. Med.* **362,** 1082–1089 (2010).
3. Speliotes, E. K. *et al.* Genome-Wide Association Analysis Identifies Variants Associated with Nonalcoholic Fatty Liver Disease That Have Distinct Effects on Metabolic Traits. *PLoS Genet* **7,** e1001324 (2011).
4. Samocha-Bonet, D. *et al.* Overfeeding Reduces Insulin Sensitivity and Increases Oxidative Stress, without Altering Markers of Mitochondrial Content and Function in Humans. *PLoS ONE* **7,** e36320 (2012).
5. Ratziu, V., Bellentani, S., Cortez-Pinto, H., Day, C. & Marchesini, G. A position statement on NAFLD/NASH based on the EASL 2009 special conference. *J. Hepatol.* **53,** 372–384 (2010).
6. Farese, R. V. Jr., Zechner, R., Newgard, C. B. & Walther, T. C. The Problem of Establishing Relationships between Hepatic Steatosis and Hepatic Insulin Resistance. *Cell Metab.* **15,** 570–573 (2012).
7. Samuel, V. T. *et al.* Mechanism of Hepatic Insulin Resistance in Non-alcoholic Fatty Liver Disease. *J. Biol. Chem.* **279,** 32345–32353 (2004).
8. Rajpathak, S. N. *et al.* Insulin-like growth factor axis and risk of type 2 diabetes in women. *Diabetes* **61,** 2248–2254 (2012).
9. Völzke, H. *et al.* Association between hepatic steatosis and serum IGF1 and IGFBP-3 levels in a population-based sample. *Eur. J. Endocrinol.* **161,** 705–713 (2009).
10. Aleem, E., Nehrbass, D., Klimek, F., Mayer, D. & Bannasch, P. Upregulation of the insulin receptor and type I insulin-like growth factor receptor are early events in hepatocarcinogenesis. *Toxicol. Pathol.* **39,** 524–543 (2011).
11. Baffy, G., Brunt, E. M. & Caldwell, S. H. Hepatocellular carcinoma in non-alcoholic fatty liver disease: An emerging menace. *J. Hepatol.* **56,** 1384–1391 (2012).
12. Marra, F., Gastaldelli, A., Svegliati Baroni, G., Tell, G. & Tiribelli, C. Molecular basis and mechanisms of progression of non-alcoholic steatohepatitis. *Trends Mol. Med.* **14,** 72–81 (2008).
13. Angulo, P. Nonalcoholic fatty liver disease. *N. Engl. J. Med.* **346,** 1221–1231 (2002).
14. Jozefczuk, J. *et al.* A Systems Biology Approach to Deciphering the Etiology of Steatosis Employing Patient-Derived Dermal Fibroblasts and iPS Cells. *Front. Physiol* **3,** 339 (2012).
15. Houstis, N., Rosen, E. D. & Lander, E. S. Reactive oxygen species have a causal role in multiple forms of insulin resistance. *Nature* **440,** 944–948 (2006).
16. Findeisen, H. M. *et al.* Glutathione Depletion Prevents Diet-Induced Obesity and Enhances Insulin Sensitivity. *Obesity* **19,** 2429–2432 (2011).
17. Hayes, J. D., Flanagan, J. U. & Jowsey, I. R. Glutathione Transferases. *Annu. Rev. Pharmacol. Toxicol.* **45,** 51–88 (2005).
18. Moylan, C. A. *et al.* Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatol. Baltim. Md* **59,** 471–482 (2014).
19. Kahali, B. *et al.* TM6SF2: catch-22 in the fight against nonalcoholic fatty liver disease and cardiovascular disease? *Gastroenterology* **148,** 679–684 (2015).
20. Kozlitina, J. *et al.* Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* **46,** 352–356 (2014).
21. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518,** 187–196 (2015).
22. Feldstein, A. E. *et al.* Cytokeratin-18 fragment levels as noninvasive biomarkers for nonalcoholic steatohepatitis: A multicenter validation study. *Hepatology* **50,** 1072–1078 (2009).
23. Du Plessis, J. *et al.* Association of Adipose Tissue Inflammation With Histologic Severity of Nonalcoholic Fatty Liver Disease. *Gastroenterology* **149,** 635–648, e14 (2015).
24. Matthews, D. R. *et al.* Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28,** 412–419 (1985).
25. Kleiner, D. E. *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41,** 1313–1321 (2005).
26. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5,** R80 (2004).
27. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24,** 1547–1548 (2008).
28. Smyth, G. K. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat. Appl. Genet. Mol. Biol.* **3** doi:10.2202/1544-6115.1027 (2004).
29. Ayonrinde, O. T. *et al.* Gender-specific differences in adipose distribution and adipocytokines influence adolescent nonalcoholic fatty liver disease. *Hepatol. Baltim. Md* **53,** 800–809 (2011).

Data Citations

1. Wruck, W. *et al. NCBI GEO* GSE46300 (2014).
2. Wruck, W. *et al. MetaboLights* MTBLS174 (2015).
3. Wruck, W. *et al. Figshare* http://dx.doi.org/10.6084/m9.figshare.1333564 (2015).

Acknowledgements

The authors acknowledge support from the German Federal Ministry of Education and Research (BMBF GRANT 0315717A), which is a partner of the ERASysBio+ initiative supported under the EU ERA-NET Plus scheme in FP7. Hans V. Westerhoff, & Samrina Rehman thank the EPSRC and BBSRC (and

BBSRC-BRIC) for support of the Manchester Centre for Integrative Systems Biology, as well as various other funding agencies (EU-FP7 [SYNPOL #311815], H2020 [Epipredict, #642691 and Corbel #654248], NWO, the SysMO and ERASysBio funder communities), in particular the ERASysBio+ support by BBSRC. Additionally, James Adjaye and Wasco Wruck acknowledge support from the Medical faculty of the Heinrich Heine University Düsseldorf.

Author Contributions

W.W. performed transcriptomics analyses, UK and D.B. performed RPPA analyses, E.G. and P.T. NMR analyses, S.R. and H.W. flux analyses (data not included here). K.K. and K.Z. were responsible for patient recruitment and RNA isolation from patient-derived liver biopsies. J.J. and K.W. performed the RT-PCR experiments. A.D. and C.W. performed pathway analysis and built systems biological models (data not included here) and V.P. performed flux coupling analysis (data not included here). W.W. was responsible for data management. W.W., A.D., S.R., H.W., C.W. and J.A. wrote the manuscript. H.L. and J.A. conceived the concept. J.A. coordinated the work.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Wruck, W. *et al.* Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes. *Sci. Data* 2:150068 doi: 10.1038/sdata.2015.68 (2015).

2.1.2 Meta-Analysis of Transcriptome Data Related to Hippocampus Biopsies and iPSC-Derived Neuronal Cells from Alzheimer's Disease Patients Reveals an Association with FOXA1 and FOXA2 Gene Regulatory Networks.

J Alzheimers Dis. 2016;50(4):1065-82. doi: 10.3233/JAD-150733.

Wruck W, Schröter F, Adjaye J.

Although the incidence of Alzheimer's disease (AD) is continuously increasing in the aging population worldwide, effective therapies are not available. The interplay between causative genetic and environmental factors is partially understood. Meta-analyses have been performed on aspects such as polymorphisms, cytokines, and cognitive training. Here, we propose a meta-analysis approach based on hierarchical clustering analysis of a reliable training set of hippocampus biopsies, which is condensed to a gene expression signature. This gene expression signature was applied to various test sets of brain biopsies and iPSC-derived neuronal cell models to demonstrate its ability to distinguish AD samples from control. Thus, our identified AD-gene signature may form the basis for determination of biomarkers that are urgently needed to overcome current diagnostic shortfalls. Intriguingly, the well-described AD-related genes APP and APOE are not within the signature because their gene expression profiles show a lower correlation to the disease phenotype than genes from the signature. This is in line with the differing characteristics of the disease as early-/late-onset or with/without genetic predisposition. To investigate the gene signature's systemic role(s), signaling pathways, gene ontologies, and transcription factors were analyzed which revealed over-representation of response to stress, regulation of cellular metabolic processes, and reactive oxygen species. Additionally, our results clearly point to an important role of FOXA1 and FOXA2 gene regulatory networks in the etiology of AD. This finding is in corroboration with the recently reported major role of the dopaminergic system in the development of AD and its regulation by FOXA1 and FOXA2.

DOI: 10.3233/JAD-150733 PMID: 26890743 [Indexed for MEDLINE]

URL: https://content.iospress.com/articles/journal-of-alzheimers-disease/jad150733

Contribution: Wasco Wruck performed the meta-analysis of transcriptome data and wrote the manuscript.

Journal of Alzheimer's Disease 50 (2016) 1065–1082
DOI 10.3233/JAD-150733
IOS Press

Meta-Analysis of Transcriptome Data Related to Hippocampus Biopsies and iPSC-Derived Neuronal Cells from Alzheimer's Disease Patients Reveals an Association with FOXA1 and FOXA2 Gene Regulatory Networks

Wasco Wruck, Friederike Schröter and James Adjaye*
Institute for Stem Cell Research and Regenerative Medicine, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

Handling Associate Editor: George Acquaah-Mensah

Accepted 13 November 2015

Abstract. Although the incidence of Alzheimer's disease (AD) is continuously increasing in the aging population worldwide, effective therapies are not available. The interplay between causative genetic and environmental factors is partially understood. Meta-analyses have been performed on aspects such as polymorphisms, cytokines, and cognitive training. Here, we propose a meta-analysis approach based on hierarchical clustering analysis of a reliable training set of hippocampus biopsies, which is condensed to a gene expression signature. This gene expression signature was applied to various test sets of brain biopsies and iPSC-derived neuronal cell models to demonstrate its ability to distinguish AD samples from control. Thus, our identified AD-gene signature may form the basis for determination of biomarkers that are urgently needed to overcome current diagnostic shortfalls. Intriguingly, the well-described AD-related genes APP and APOE are not within the signature because their gene expression profiles show a lower correlation to the disease phenotype than genes from the signature. This is in line with the differing characteristics of the disease as early-/late-onset or with/without genetic predisposition. To investigate the gene signature's systemic role(s), signaling pathways, gene ontologies, and transcription factors were analyzed which revealed over-representation of response to stress, regulation of cellular metabolic processes, and reactive oxygen species. Additionally, our results clearly point to an important role of FOXA1 and FOXA2 gene regulatory networks in the etiology of AD. This finding is in corroboration with the recently reported major role of the dopaminergic system in the development of AD and its regulation by FOXA1 and FOXA2.

Keywords: Alzheimer's disease, energy metabolism, forkhead box proteins, gene expression, induced pluripotent stem cells, meta-analysis, microarray analysis, transcription factors

INTRODUCTION

Alzheimer's disease (AD) is estimated to affect about 35 million of the world's population and due to continuous increase in our aging population, it is obvious that its prevalence will increase

*Correspondence to: James Adjaye, Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-Universität Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. Tel.: +49 0 211 81 08191; Fax: +49 0 211 81 17858; E-mail: James.Adjaye@med.uni-duesseldorf.de.

ISSN 1387-2877/16/$35.00

[1]. Epidemiology shows only slight variation in the prevalence with a band from 5% to 7% in most regions of the world—however with some noticeable outliers of about 2% in West Sub-Saharan Africa [2]. A dichotomy of rare genetic variants leading to early-onset AD (EOAD) and more frequent but less malignant variants leading to late-onset (LOAD) has been proposed [3]. The most prominent discovery for LOAD has been the association of a higher risk with the APOE 4 allele [4]. However, the effect is mediated by its action on the amyloid-β product of a variant APP gene. Nonetheless, age and gender remain the major risk factors for sporadic forms of AD.

The interplay of genetic and environmental influences leading to the disease is only partially understood. There is now increasing awareness that age-associated metabolic impairment might also have a contributory effect on LOAD. In our earlier studies on aging in mouse brain, heart, and kidney, we postulated that metabolic stability of gene regulatory networks modulating glutathione metabolism, oxidative phosphorylation, insulin signaling, and inflammation which are biological processes intimately associated with aging are under constraint and should be tightly controlled to attain healthy aging [5, 6]. The metabolic stability theory which is the ability of cells to maintain stable concentrations of reactive oxygen species (ROS) and other critical metabolites is in fact further supported by emerging evidence that associates type 2 diabetes mellitus (T2DM), brain insulin resistance, oxidative stress, and cognitive impairment with AD [7]. More recent data also imply an interplay between metabolic homeostasis and neurogeneration [8].

Pistollato et al. hypothesize that progress in AD research has been hampered by overuse of animal models inadequate for the study of LOAD which accounts for about 95% of AD cases [9]. Demetrius and Driver consider the sporadic late-onset forms of AD a metabolic disease [10]: the amyloid cascade hypothesis postulating the imbalance of amyloid production and clearance due to missense mutations can explain the familial EOAD forms but cannot stringently explain the etiology of sporadic forms as it ignores the impact the brain energy metabolism has on neuronal dynamics. The simplistic view of the brain energy metabolism consisting of glucose as main energy substrate has been refined by the findings that neurons acquire their energy predominantly from oxidative phosphorylation and astrocytes from glycolysis via anaerobically produced lactate [10, 11]. Moreover, a dichotomy of AD pathological changes between the youngest-old (individuals at about 75 years) and the oldest-old (individuals at about 95 years) has been reported in view of the fact that for the youngest-old the severity of dementia correlates very well with neural plaques and neurofibrillary tangles while that correlation is reduced for the oldest-old [12, 13]. However, correlation between cerebral atrophy and dementia consistently is strong for the youngest-old and oldest-old [13]. These observations for the oldest-old are hard to explain with the amyloid-cascade hypothesis but in particular the correlation to brain atrophy is very well in line with the metabolic model. Inter-relations between metabolism, brain atrophy, and age have already been reported in several publications, e.g., by Yoshii et al. [14], and cerebral glucose metabolism measures have even been suggested as AD markers [15]. Besides the age-dependence the metabolic model also referred to as *Inverse Warburg Hypothesis* can elucidate two further hallmarks of sporadic AD which are in conflict with the amyloid cascade model: the selective vulnerability of neurons in different brain regions and the inverse cancer comorbidity [16].

As a consequence of our fragmentary understanding of the molecular basis of AD, fully functional therapies are not available. Furthermore, unsatisfying accuracy of diagnoses has been reported, e.g., in a large study 39% of clinically diagnosed non-AD dementia patients showed postmortem histopathology consistent with AD [17].

Meta-analyses have been performed related to polymorphisms [18], secreted cytokines [19], and cognitive training [20]. Here, we performed meta-analysis based on hierarchical clustering analysis of transcriptome data of a reliable training set of hippocampus biopsies which led to a gene expression signature distinguishing AD from healthy control samples. This gene expression signature was also applied to various test sets of brain biopsies and induced pluripotent stem cells (iPSC)-derived neuronal cell transcriptome models to demonstrate its ability to distinguish AD from control in various experimental setups. Besides the EOAD PSEN1 mutation brain biopsy dataset from Antonell et al. [21] and the iPSC-derived sporadic AD model from Hossini et al. [22], the biopsy datasets we analyzed are representative of the LOAD form, while the iPSC-based AD models are based on EOAD-related APP- and PSEN-mutations.

Recently, links between the monoaminergic/dopaminergic system and AD were described [23]. The predominant locations of this system (locus

coeruleus, raphe nuclei, and tuberomammillary nucleus) are degenerated and thus impair functionality of hippocampus and cortical neurons. Furthermore, genome-wide association studies of AD patients detected polymorphisms in genes from the dopaminergic system, e.g., polymorphisms in Catechol-O-methyltransferase (COMT) are associated with availability of dopamine and, on the other hand, with loss of behavioral control and psychosis in AD patients [24]. Moreover, Ferri et al. [25] and Stott et al. [26] reported that Foxa1 and Foxa2 regulate development and maintenance of dopaminergic neurons in the mouse midbrain. This finding is in corroboration with the observation that FOXA1 and FOXA2 regulate dopaminergic neurons and that the dopaminergic system plays a major role in the development of AD.

MATERIALS AND METHODS

Determining an AD-correlated gene signature from a training set of transcriptomes from hippocampus-derived biopsies

Figure 1 illustrates the scheme of this meta-analysis. A sound dataset of hippocampus biopsies (GSE29378 downloaded from the public repository NCBI GEO) with a relatively high sample size of 72 produced on a state-of-the-art microarray platform (Illumina Beadchip) is used as training set [27]. The correlation of each gene to the AD phenotype is calculated on a binary scale (AD = 1, healthy control = 0) using Spearman correlation. Genes with the '*n*' most significant correlations are determined based on the test for association between paired samples using the R function cor.test(). This implies that highly correlated and anti-correlated (down-regulated in AD) genes are integrated into the gene signature.

Recalibration of the gene signature on an alternative platform

In order to place the gene signature on a more robust basis which is independent of the technical platform, it was recalibrated with the dataset GSE36980 downloaded from NCBI GEO [28] on an alternative technical platform - the Affymetrix Human Gene 1.0 ST Array. In this dataset, again the correlation with the AD phenotype was determined employing the approach used for the first dataset. The probesets found to be correlated with the AD phenotype in this probeset ($p < 0.05$) were intersected with the gene signature from the first step. The result is a reduced gene signature, which, however, is more robust as it works at least for the Affymetrix and the Illumina microarray platforms.

Cluster analysis of brain biopsies

The re-confirmed gene signature was employed for cluster analysis in several test sets of brain biopsy-derived microarray data. Test datasets for human brain biopsies were downloaded from NCBI GEO for Affymetrix and Illumina microarray platforms. These comprise the datasets: GSE4757 [29], GSE26927 [30], GSE39420 [21]. In detail, the raw data were processed via R/Bioconductor [31] packages affy [32], lumi [33] and oligo [34] and normalized using a platform-specific method (rma for Affymetrix, quantile normalization for Illumina). The probes or probesets matching genes from the gene signature are extracted from the normalized test dataset and subjected to cluster analysis via the R function heatmap.2.

Cluster analysis of iPSC-based AD models

As a complement to the brain biopsies, cluster analysis was also performed for several test sets of iPSCs-based AD models microarray data for the re-confirmed gene signature. Test datasets for iPSCs-based AD models were downloaded from NCBI GEO for Affymetrix and Illumina microarray platforms. These comprise the datasets: GSE28379 [35], GSE43326 [36], GSE34879 [37], and GSE42492 [22]. These test datasets were processed in the same way as described above for the brain biopsy test datasets.

Transcription factor analysis

Transcription factors were analyzed using the oPOSSUM-3 tool [38]. Single Site Analysis was performed on a locally installed oPOSSUM database (downloaded in June 2014) in order to detect over-represented conserved transcription factor binding sites within the AD gene signature. Parameters were set to the following values: species "human", 2000 base pairs upstream and downstream each, use only JASPAR Transcription Factor Binding Site (TFBS) profile matrices which belong to the tax group "vertebrates", a minimum relative TFBS position weight matrix (PWM) score of 0.85 and a minimum infor-

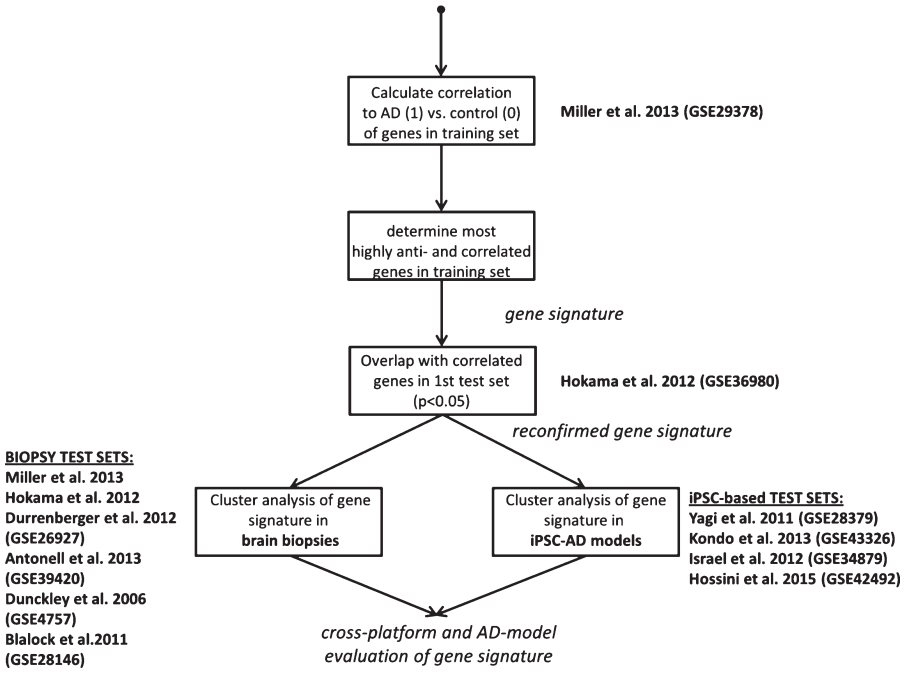


Fig. 1. Scheme of the meta-analysis method. Genes with the highest correlation to the AD phenotype are extracted from the training set of brain biopsies analyzed on the Illumina platform. In order to achieve platform independency the resulting gene signature is recalibrated with another training set on the Affymetrix platform. This gene signature is extracted from test sets of brain biopsies on the one hand and from test sets of iPSC-based AD models on the other hand and subjected to cluster analysis.

mation content (specificity) of JASPAR TFBS profile matrices of 8.

For the network plot, we used the R package network [39] to connect transcription factors with an oPOSSUM Z-score > 3 with their target genes. The oPOSSUM Z-score accounts for an enrichment of TFBS in the investigated up- and downstream regions of the set of genes compared to the background. Genes and transcription factors were distinguished by red (white in b/w) circles for genes and green (grey in b/w) circles with sizes corresponding to the Z-scores for the transcription factors.

Pathway and gene ontology over-representation analysis

KEGG pathways were analyzed via the hypergeometric test from the R package [40]. The qvalue package was employed to adjust for multiple testing [41]. Over-represented gene ontologies were determined employing the GOstats package [42] from the R/Bioconductor environment [31].

Gene expression analysis

Quantitative real-time polymerase chain reaction (qRT-PCR) was carried out as described [22]. Briefly the data presented fold induction based on the $\Delta\Delta$Ct calculation over fetal brain RNA for the adult and AD brain and undifferentiated H9 or AD5 cells for the iPSC model. Only one neuronal differentiation of a single well of embryonic stem cell line H9 and sAD iPSC line AD5 was used for this analysis.

RESULTS

The meta-analysis of the datasets listed in Table 1 was performed according to the flow chart illustrated in Fig. 1. We started with the determination of genes most significantly correlated or reversely correlated with the AD phenotype in the dataset from Miller et al. Figure 2 shows that the gene signature consisting of the 40 most significantly correlated or anti-correlated Illumina probes (mapping to 34

Table 1
Datasets employed for the meta-analysis: (A) brain biopsy datasets, (B) iPSC-derived AD model datasets

A					
Literature	Dataset	Mutation	location of biopsy	NR of Samples	Figure
Miller et al., [27]	GSE29378	–	hippocampus CA1 and CA3	35x AD 37x control	2
Hokama et al., [28]	GSE36980	–	hippocampus	7x AD, 10x control	3
Durrenberger et al., 2012 [30]	GSE26927	–	entorhinal cortex	11x AD, 7x control	S6A
Antonell et al., [21]	GSE39420	PSEN1	hippocampus	7x EOAD, 7x PSEN1-mut, 7x control	S5A
Dunckley et al., [29]	GSE4757	–	entorhinal cortex	10x NFT, 10x control	S6B
Blalock et al., [44]	GSE28146	–	hippocampus	22x AD, 8x control	S5B
B					
Literature	Dataset	Mutation	iPSC Protocol to derive neurons	NR of Samples	Figure
Kondo et al., [36]	GSE43326	fAD: APP	cortical neurons	1x APPE693, 1x APPwild	4
Yagi et al., [35]	GSE28379	EOAD, fAD: PSEN2	adapted to their murine protocol	2x AD, 2x control	S7A
Israel et al., [37]	GSE34879	fAD: APP	FACS purification of neural progenitor cells	1x APP, 2x NDC	S7B
Hossini et al., [22]	GE42492	LOAD, sAD	4-weeks of TGFb-receptor and MEK1/2 inhibiton	2x sAD, 1x ES control	S7C

AD, Alzheimer's disease; NFT, neurofibrillary tangles; EOAD, early-onset Alzheimer's disease; fAD, familial Alzheimer's disease; LOAD, late-onset Alzheimer's disease; sAD, sporadic Alzheimer's disease.

distinct gene symbols, Table 2) enables a good separation into two main clusters. One cluster (healthy control cluster) contains only healthy control brain biopsies while the other cluster (AD cluster) contains 35 (87.5%) AD brain biopsies and 5 outliers in 3 groups (one group contains two biopsies from the same patient). Table 2 lists these 34 genes most significantly correlated with the AD phenotype. 32 genes correlate positively (most significant: *GEM, S100A6, YAP1*, and *PFKFB3*) while only two genes (*FXYD5* and *WDFC1*) are anti-correlated.

The good separation of AD and healthy controls by the gene signature is additionally demonstrated by the principal component analysis (PCA) plotted in Supplementary Fig. 1. While we used a supervised clustering approach (filtering genes by their correlation to the AD phenotype), unsupervised clustering analyses revealed no clear separation of AD and control clusters. For comparison, Supplementary Fig. 2 shows the result of a clustering analysis of genes from the KEGG AD pathway, and Supplementary Fig. 3 shows the result of a clustering analysis of genes from the secretome on the training dataset GSE29378. The results of the reconfirmation of the gene signature via the GEO dataset GSE36980 from Hokoma et al. [28] is presented in Fig. 3. The adjusted gene signature then is reduced to 29 genes.

Intriguingly, neither the well-described AD-related genes *APP* and *APOE* nor the later discovered presenilin1/2 (*PSEN1*, *PSEN2*) were within the signature because their gene expression profiles showed a lower correlation to the disease phenotype than genes from the signature. This is in line with the distinct characteristics of the disease as EOAD or LOAD or with or without genetic predisposition and the low frequency of the AD-causing mutations in *APP, APOE, PSEN1*, and *PSEN2* [43]. To investigate the gene signature and its relevance to the AD phenotype in more detail, associated pathways, gene ontologies, and transcription factors were analyzed. Finally, we demonstrated that our gene signature could better distinguish AD from healthy controls than a collection of genes encoding secretory factors. Thus, it may be a good base for determination of biomarkers that are urgently needed to overcome the current shortfall in accurate diagnostics.

Cluster analysis of hippocampus biopsy datasets

The good correlation of the clustering of dataset GSE36980 from Hokama et al. [28] with the AD phenotype is illustrated by Fig. 3. Also with other hippocampus datasets, the gene signature performed very well. Using the gene signature for cluster analysis of the hippocampus biopsy dataset GSE39420 with EOAD and PSEN1 mutations from Antonell et al. [21] facilitated a good separation of one pure AD cluster and one healthy control clusters containing two AD cases (Supplementary Fig. 4). Additionally,

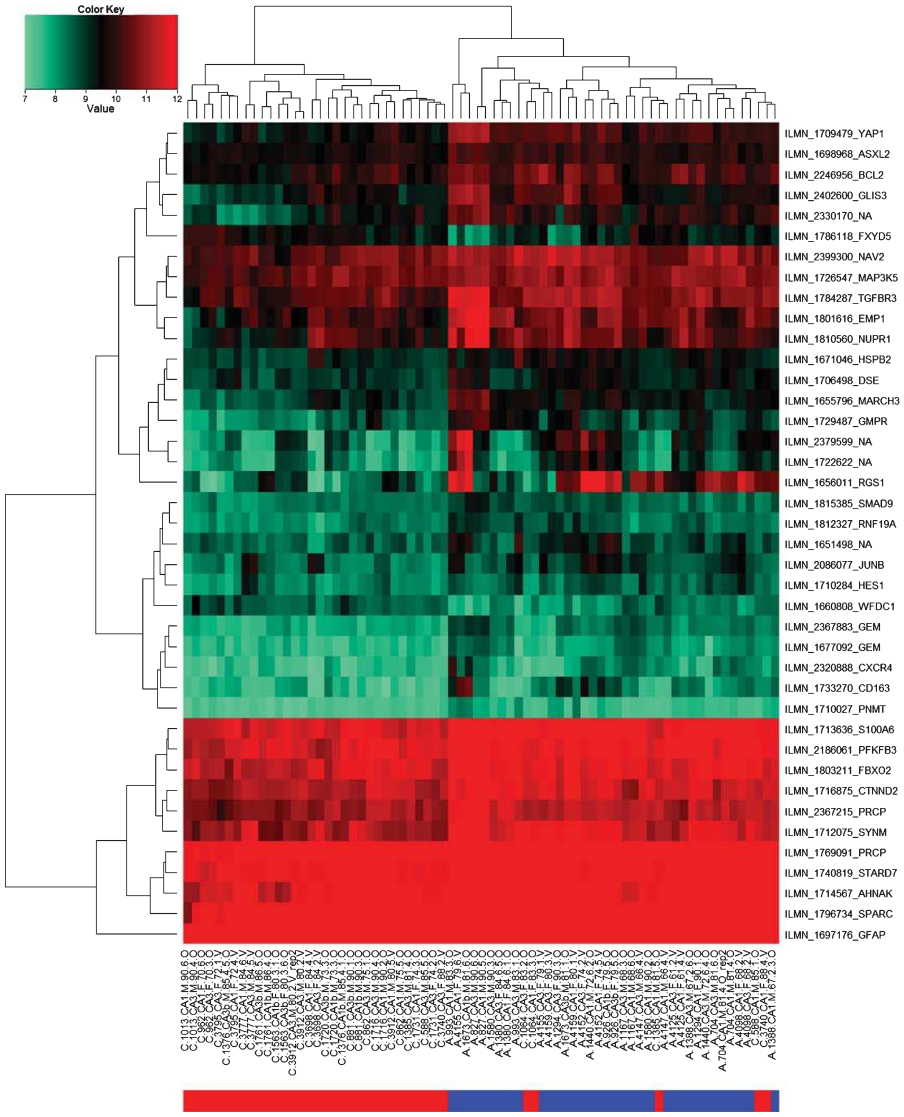


Fig. 2. Cluster analysis of the training set separates one healthy control and one AD cluster. The gene signature of the 40 most significantly correlated or anti-correlated Illumina probes was employed for cluster analysis of the training set of hippocampus biopsies. Red bars at the bottom indicate healthy control while blue bars indicate AD. Two main clusters are clearly separated one pure healthy control cluster and one AD cluster with 3 small outlier groups consisting of 5 healthy control samples.

the dataset GSE28146 from Blalock et al. [44] derived from formalin-fixed, paraffin-embedded (FFPE) hippocampus biopsies was subjected to cluster analysis and enabled separation into one AD cluster and one mixed cluster with healthy controls and not-severe AD cases (Supplementary Fig. 5). The results also showed that in most datasets, there was a high level of variation between the various transcript variants and not all correlated with the AD phenotype. Using the variants with the highest correlation to the AD phenotype unveiled the results shown in (Supplementary Fig. 5).

Table 2
Genes from the AD gene signature and their correlation with the AD phenotype

Probeid	Correlation	p	Symbol (alias)	PubMedID	Reference
ILMN_2367883	0.7104	6.85E-12	GEM		
ILMN_1713636	0.7029	2.35E-12	S100A6	15590066	[50] (t)
ILMN_1709479	0.6494	2.07E-10	YAP1	21178287	[64] (a)
ILMN_2186061	0.6417	3.23E-10	PFKFB3	23427097	[65] (p8)
ILMN_1656011	0.6356	1.78E-09	RGS1		
ILMN_1784287	0.6138	2.66E-09	TGFBR3	17080199	[48] (t)
ILMN_1769091	0.6065	4.59E-09	PRCP		
ILMN_1712075	0.5995	1.25E-08	SYNM		
ILMN_1740819	0.5888	3.10E-08	STARD7		
ILMN_1801616	0.5861	2.07E-08	EMP1	23978990	[66] (a)
ILMN_1716875	0.5819	5.26E-08	CTNND2	22984439	[67] (t)
ILMN_1655796	0.5786	1.04E-07	MARCH3		
ILMN_1698968	0.5734	8.41E-08	ASXL2		
ILMN_1803211	0.5718	1.28E-07	FBXO2	24469452	[68] (t)
ILMN_1796734	0.5709	2.60E-08	SPARC	20435134	[69] (Table 3)
ILMN_1810560	0.5681	1.85E-07	NUPR1		
ILMN_1812327	0.5666	1.45E-07	RNF19A		
ILMN_1714567	0.5654	2.00E-07	AHNAK	11312263	[70] (t)
ILMN_2246956	0.5600	2.65E-07	BCL2	8922409	[49] (t)
ILMN_1706498	0.5592	3.62E-07	DSE		
ILMN_1815385	0.5562	2.17E-07	SMAD9	22815752	[71] (p9,11)
ILMN_1726547	0.5556	1.70E-07	MAP3K5(ASK1)	15592360	[59] (t)
ILMN_1710027	0.5553	8.07E-08	PNMT	11378842	[72] (t)
ILMN_1729487	0.5548	1.08E-07	GMPR	25448601	[73] (Table 3)
ILMN_2402600	0.5536	2.44E-07	GLIS3	23562540	[74] (a)
ILMN_2320888	0.5526	4.07E-07	CXCR4	17764962	[75] (t)
ILMN_2399300	0.5502	2.54E-07	NAV2	25859259	[76] (a)
ILMN_1733270	0.5481	3.49E-07	CD163	24528486	[77] (t)
ILMN_1671046	0.5317	8.78E-08	HSPB2	16599941	[78] (a)
ILMN_1710284	0.5243	3.49E-07	HES1	22849569	[79] (t)
ILMN_2086077	0.5191	4.05E-08	JUNB	8313943	[80] (t)
ILMN_1697176	0.5159	2.64E-07	GFAP	25991443	[51] (p364)
ILMN_1786118	–0.5254	4.07E-07	FXYD5		
ILMN_1660808	–0.5682	2.84E-08	WFDC1	23705665	[27] (Table 3)

The 34 genes most significantly correlated with the AD phenotype. 32 genes are positively correlated (most significant: *GEM, S100A6, YAP1,* and *PFKFB3*) while only two genes (*FXYD5* and *WDFC1*) are anti-correlated. Beside the reference, the text location is indicated (t, title; a, abstract).

Cluster analysis of entorhinal cortex biopsy datasets

Cluster analysis of entorhinal cortex biopsies (Supplementary Figs. 6 and 7) did not yield as good results as for hippocampus biopsies (Supplementary Figs. 4 and 5). Cluster analysis with the gene signature on the test set of entorhinal cortex biopsies AD dataset GSE26927 from Durrenberger et al. [30] resulted in three clusters one pure AD cluster containing only two samples and two incoherent clusters containing AD as well as control samples (Supplementary Fig. 6). Red bars at the bottom indicate healthy control while blue bars indicate AD. The dataset GSE4757 from Dunckley et al. [29] consists of healthy controls and entorhinal cortex biopsies with neurofibrillary tangles which are considered an early event in AD pathology (Supplementary Fig. 7). Cluster analysis with the gene signature on this test set resulted in three clusters that were more or less mixed up and only had a few coherent sub-clusters. Red bars at the bottom indicate healthy control while blue bars indicate neurofibrillary tangles. The inferior performance of the gene signature on entorhinal cortex biopsies might be due to its specificity to hippocampus. AD-related gene expression and regulatory networks exhibit distinct behavior between hippocampus and entorhinal cortex.

Cluster analysis of iPCS-derived AD model datasets

In comparison to brain biopsy-derived datasets, a major disadvantage of iPSC-derived AD models is

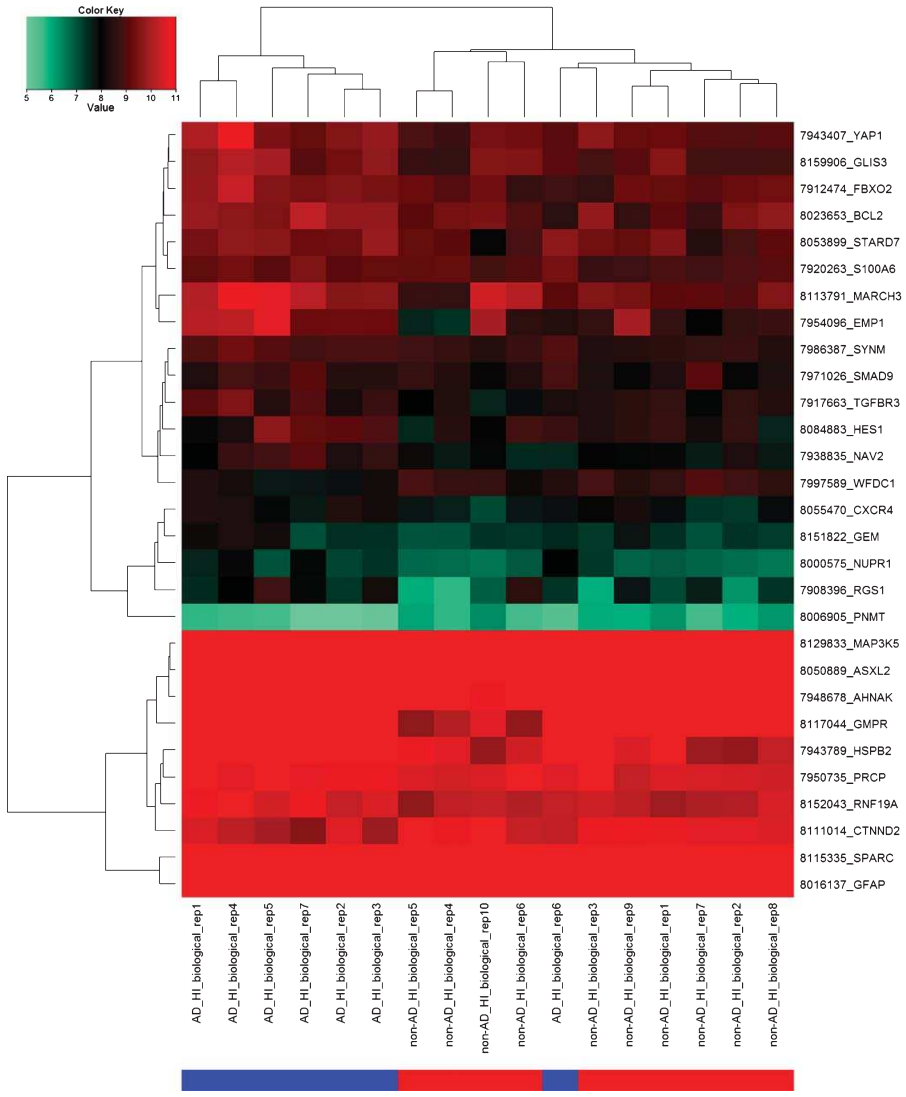


Fig. 3. Cluster analysis of the alternative-platform test set separates AD from healthy control. After recalibration of the gene signature with data from the GSE36980 Affymetrix hippocampus biopsy dataset 29 AD-correlated genes remain. This signature is considered the platform-independent AD gene signature. It enables clustering of the GSE36980 dataset into one pure AD cluster and one healthy control cluster containing one outlier AD sample.

the smaller sample size of the datasets due to the complex process of patient-specific iPSC-derivation and differentiation into neuronal cells. On the other hand, the advantages of the iPSC-based models are obvious: patient-specific *in-vitro* models of AD can be perturbed in order to test new hypotheses. A summary of the analyzed iPSC-derived AD models with their mutation and the neural protocols is provided in Table 1B.

The cluster analysis with the gene signature for the dataset GSE43326 from Kondo et al. [36] with relatively high sample size for iPSC-AD-models showed good results: it separated one familial AD (*APP* mutations) from the healthy con-

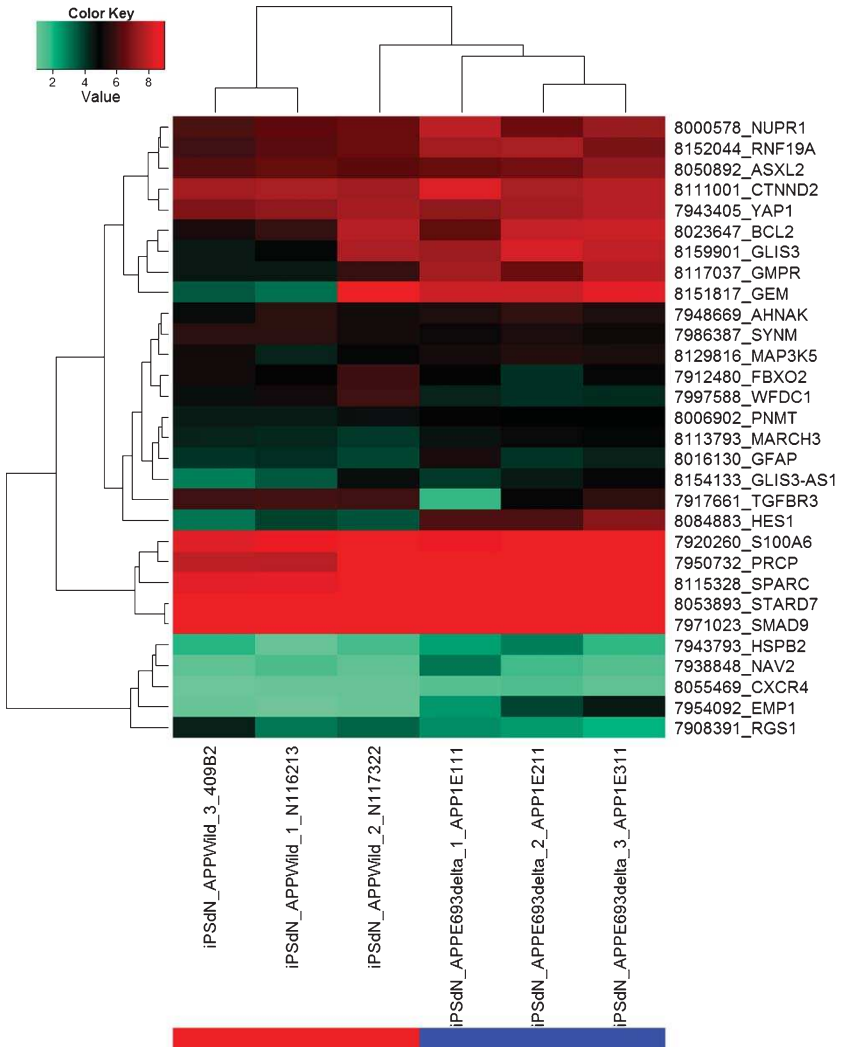


Fig. 4. Gene signature demonstrates functionality of the iPSC-based AD model in reproducing the AD phenotype. Cluster analysis using the gene signature with an iPSC-based AD model from Kondo et al. (2013, GEO accession no. GSE43326) enables separation of an AD-associated cluster. The dataset consists of iPSC-derived neurons from patients with a familial amyloid precursor protein (APP)-E693D mutation and from control subjects with a wildtype APP. This cluster analysis also showed the relevance of transcript variants as the optimal cluster separation was only possible with specific transcripts.

trol samples (*APP* wild type) as demonstrated in Fig. 4.

As mentioned above, the small sample size has major impact on the results of the cluster analysis: in most datasets it is so low that a reasonable clustering is barely possible. Dataset GSE28379 from Yagi et al. [35] contains two iPSC-derived AD model cell lines carrying *PSEN2* mutations and two control iPSC-derived neuron samples—one of them from a Parkinson's disease (PD) patient (Supplementary Fig. 8). Here, one *PSEN2* mutation sample clusters with the PD sample while the other *PSEN2* mutation sample clusters separately from the other samples. In dataset GSE34879 from Israel et al. [37], there is only

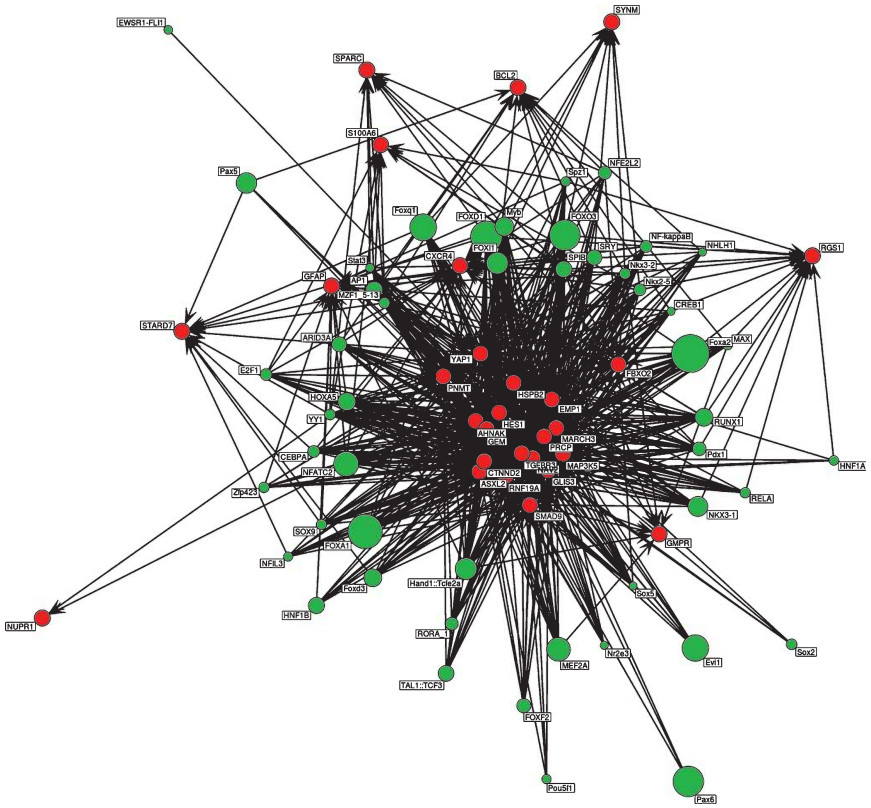


Fig. 5. Transcription factor analysis of the gene signature reveals FOX-family factors as main regulators. Transcription factor analysis with the oPOSSUM-3 tool was performed with genes from the AD gene signature. Red circles denote genes while green circles denote transcription factors with a size corresponding to their significance (z-score). Factors from the FOX family (FOXA1, FOXA2, FOXD1, FOXO3, FOXQ1, FOXI1) have the highest significance.

one AD patient-specific iPSC sample and one iPSC-derived neuron sample. Nevertheless, these samples showed some differences with the healthy control samples in the cluster analysis of the gene signature (Supplementary Fig. 9). The dataset from Hossini et al. [22] comprised two iPSC-derived neuron cell cultures derived from a single sporadic AD (sAD) and one healthy control derived from embryonic stem cell line H9 (Supplementary Fig. 10). Applying the gene signature to this test set clustered the two AD samples separate from the healthy control revealing some prominent differences between the two AD samples and the control, e.g., the S100 calcium binding protein A6 (S100A6). In summary, our AD gene signature could separate AD samples from healthy controls in the iPSC-based AD models, with the exception of the Yagi et al. (GSE28379, [35]) dataset which was probably masked by the similarity of the PD control to the AD phenotype.

Characterization of the gene signature

The gene signature was further investigated with the aim to identify associated gene regulatory networks employing gene set and transcription factor analyses. Table 3 shows the 20 most significant terms from gene ontology over-representation analysis of the AD gene signature. These gene ontology terms point to a major role of regulation of glial cells and development of nervous system. Table 4 shows the results of a KEGG pathway over-representation analysis. Protein processing in endoplasmatic reticulum, amyotrophic lateral sclerosis and Neurotrophin signaling pathway are found as significant with a

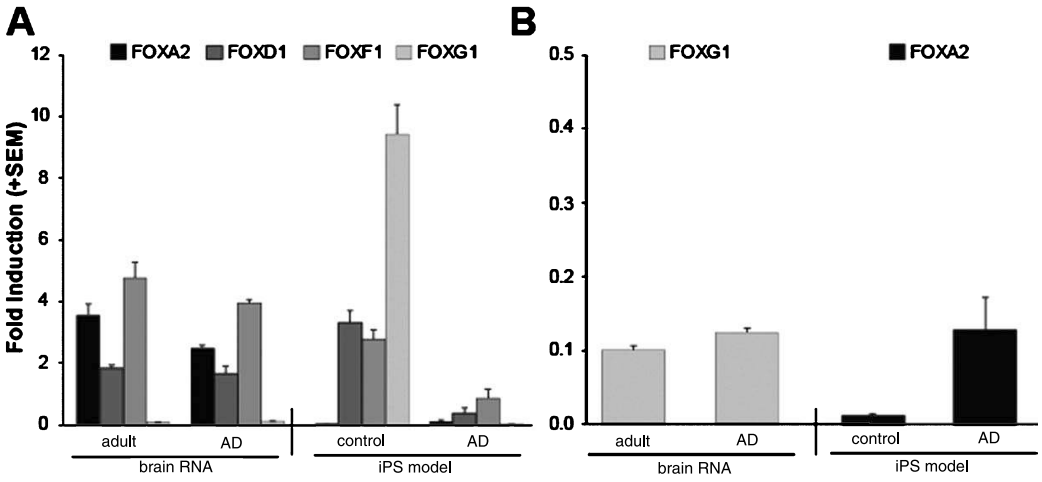


Fig. 6. FOX gene expression analysis from commercially available brain RNA and iPSC model. Quantitative real time PCR (qRT-PCR) analysis of healthy adult and AD brain as well as healthy (H9)- and diseased (AD5)-derived neuronal network of [22] are shown. Both analyses are presented relative to the "housekeeping" gene β-Actin and normalized to fetal brain RNA (for adult and AD brain) and to undifferentiated H9 or AD5 cells (for stem cell-derived neurons).

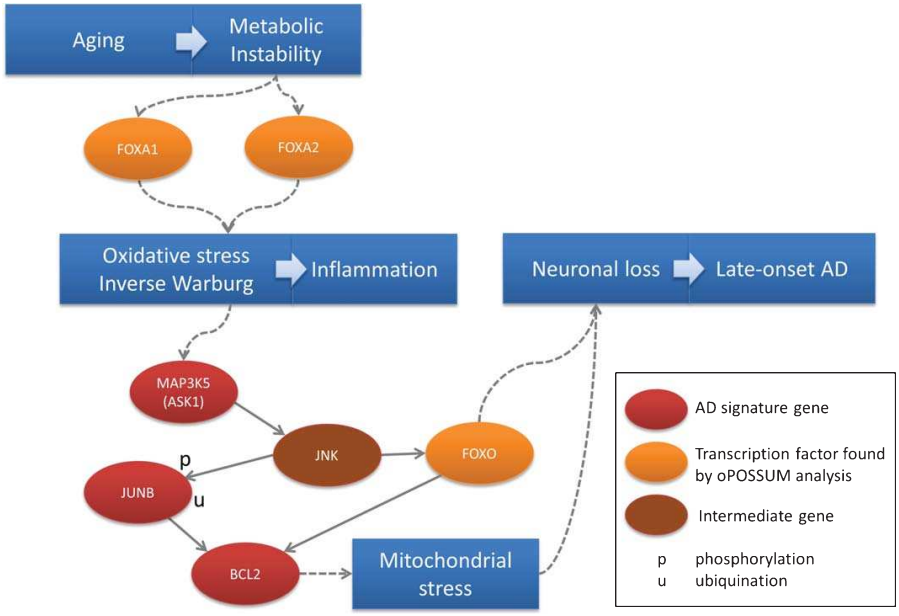


Fig. 7. FOXA and FOXO mediated induction of oxidative stress and consequent neuronal loss. FOXA transcription factors induce Oxidative Phosphorylation leading to oxidative stress and abnormal neuronal cell cycle entry. Different cellular localizations of FOXO1 together with its regulators JNK and PI3K-PKB in response to oxidative stress may cause cell cycle arrest. Eventually, cell cycle arrest results in neuronal loss.

p-value < 0.05 and at least two genes in the pathway. KEGG pathway Protein processing in endoplasmic reticulum (Supplementary Fig. 11) obviously is related to AD associated misfolded amyloid-β (Aβ) and tau proteins. Furthermore, within this pathway involvement of the ubiquitin ligase complex via FBP (*FBXO2*) and induction of apoptosis (*BCL2* and *ASK1:MAP3K5*) are fundamentals of AD.

Figure 5 shows the network resulting from a transcription factor analysis with the gene signature using the oPOSSUM-3 tool [38]. Factors from the FOX family (*FOXA1, FOXA2, FOXD1, FOXO3, FOXQ1*, *FOXI1*) have the highest significance (see also Supplementary Fig. 12, Supplementary Table 1). The graphs point to the Forkhead/winged helix transcription factors that regulate the gene signature. Transcription factors *FOXA1* and *FOXA2* and others from the Forkhead/winged helix family were most significant with a z-score greater than mean plus two standard deviations. Foxa1 and Foxa2 have been reported to regulate development and maintenance of mouse dopaminergic neurons [25, 26] and the dopaminergic system in midbrain has been associated with AD [23]. The FOXA1 and FOXA2 mediated regulation of dopaminergic neurons has been reported for the midbrain whereas the hippocampus is part of the forebrain. Thus the question emerged if the midbrain-located dopaminergic system also affects the hippocampus. Influences of the dopaminergic system on the hippocampus have been described [45]: dopaminergic neurons promote hippocampal reactivation and axons expressing dopaminergic markers have been found within the hippocampus. Our results suggest that the gene regulatory network orchestrated by FOXA1 and FOXA2 not only affects the dopaminergic neurons in the midbrain but also in the hippocampus where dopaminergic axons are located. However, this hypothesis needs further validation using an AD-iPSC cell model.

To further investigate if there is contribution of the FOX genes to AD, we analyzed expression of several FOX genes by qualitative real-time PCR analysis as shown in Fig. 6. We analyzed the genes encoding FOXA2 (see Fig. 5 and Supplementary Fig. 12), FOXD1 (involved in diencephalon development [46]; see Supplementary Fig. 12), FOXF1, and FOXG1 (associated with forebrain interneuronal development [47]). For a general overview of these FOX genes in neuronal patterning, we documented their expression in healthy adult and AD brain (normalized to fetal brain; all commercial RNA from Amsbio®), Fig. 6A. In the AD-affected brain, *FOXA2* is slightly downregulated, beyond that the medial ganglionic eminence-progenitor marker *FOXG1* is barely detectable as expressed in both healthy adult and AD brain. Further analysis in the iPSC models of Hossini et al. [22], reveals a general depression of all FOX genes in the neural network derived from the sAD patient with the exception of *FOXA2*, which seems to be upregulated (Fig. 6B).

Furthermore, in Supplementary Fig. 13, significant transcription factors FOXA1, FOXA2, FOXO3, FOXD1, and PAX6 were analyzed together with genes from the AD signature in the iPSC-based AD model from Hossini et al. [22]. FOXO3 shows the highest expression over all samples and clusters together with PAX6. FOXA1 and FOXA2 have high differences between AD and healthy control and are in the same cluster together with *GLIS3, CXCR4, NUPR1*, and *EMP1*.

Consideration of AD as metabolic disease

As described extensively in our introduction, it is evident that metabolism should be included when studying the etiology of AD. To this end, we carried out analyses targeting metabolic processes within the current datasets.

Table 5 lists the over-representation of central metabolic KEGG pathways in the analyzed AD datasets tested via hypergeometric test. We found TCA-cycle significant ($p < 0.05$) in the dataset from Hokama et al. [28] as well as oxidative phosphorylation which also is significant in Blalock et al. [44], while insulin signaling is significant in Hokama et al. [28] and also in the iPSC AD model from Kondo et al. [36]. Additionally, other metabolism-related pathways are detected as significant, for example, purine metabolism (Supplementary Table 2). Furthermore, gene ontology over-representation analysis demonstrates that response to stress, regulation of cellular metabolic processes, and reactive oxygen species are significantly over-represented in the AD gene signature (Supplementary Table 3). In summary, these findings show that several metabolic processes and particularly oxidative stress are associated and intimately involved in the pathogenesis of AD.

DISCUSSION

With our meta-analysis we found a gene signature of 34 distinct genes. For most of them an association with AD has been reported (Table 2). Trying to

Table 3
Gene ontology overrepresentation analysis of the AD gene signature

GOID	*p* value	Term	*g*enes
0014015	0.00003	positive regulation of gliogenesis	CXCR4,GFAP,HES1
0008283	0.00008	cell proliferation	BCL2,CXCR4,EMP1,GFAP,HES1,NUPR1,S100A6,SPARC,TGFBR3, WFDC1,YAP1
0035295	0.00016	tube development	BCL2,CXCR4,HES1,SMAD9,SPARC,YAP1
0023056	0.00020	positive regulation of signaling	ASXL2,BCL2,CXCR4,GFAP,HES1,MAP3K5,TGFBR3,YAP1
0014013	0.00020	regulation of gliogenesis	CXCR4,GFAP,HES1
0010647	0.00020	positive regulation of cell communication	ASXL2,BCL2,CXCR4,GFAP,HES1,MAP3K5,TGFBR3,YAP1
0010720	0.00024	positive regulation of cell development	BCL2,CXCR4,GFAP,HES1
0021783	0.00026	preganglionic parasympathetic nervous system development	HES1,NAV2
0042127	0.00028	regulation of cell proliferation	BCL2,GFAP,HES1,NUPR1,S100A6,SPARC,TGFBR3,WFDC1,YAP1
0045597	0.00035	positive regulation of cell differentiation	ASXL2,BCL2,CXCR4,GFAP,HES1,SMAD9
0048486	0.00035	parasympathetic nervous system development	HES1,NAV2
0060251	0.00041	regulation of glial cell proliferation	GFAP,HES1
0002320	0.00046	lymphoid progenitor cell differentiation	BCL2,HES1
0016049	0.00064	cell growth	BCL2,EMP1,NUPR1,TGFBR3,WFDC1
0030856	0.00065	regulation of epithelial cell differentiation	BCL2,HES1,YAP1
0014009	0.00070	glial cell proliferation	GFAP,HES1
0035265	0.00072	organ growth	BCL2,TGFBR3,YAP1
0045595	0.00075	regulation of cell differentiation	ASXL2,BCL2,CXCR4,GFAP,HES1,SMAD9,TGFBR3,YAP1
0009628	0.00084	response to abiotic stimulus	BCL2,CXCR4,GMPR,SMAD9,SPARC,TGFBR3,YAP1
0048713	0.00085	regulation of oligodendrocyte differentiation	CXCR4,HES1

The 20 most significant terms from gene ontology over-representation analysis of the AD gene signature point to a major role of regulation of glial cells and development of nervous system. The analysis had been conducted for all three Gene Ontology top categories but the most significant Cellular Component term was "glial cell projection" with only the gene GFAP and a *p*-value of 0.0034, which was above the *p*-values of the top 20 Biological Processes terms. The same holds for the Molecular Function terms where "coreceptor activity" was the most significant term with a *p*-value of 0.000873, which also was above the *p*-values of the top 20 Biological Processes terms.

find common functionality in this gene set via literature search identified the TGFβ pathway, which is represented by the genes *SMAD9* and *TGFBR3* and the genes *BCL2*, *FBXO2*, and *MAP3K5* related to protein processing. Tesseur et al. described that a deficiency in TGFβ signaling can initiate AD pathology [48]. Paradis et al. found that Aβ peptide of AD downregulates Bcl-2 [49]. Furthermore the astrocytic calcium/zinc binding gene *S100A6* which is part of our AD signature has been reported to be overexpressed in AD and in PS1/APP transgenic mouse models [50]. Another striking gene from the AD signature was *GFAP*. Its expression differed from the brain-specific- to the iPSC-derived AD models. It seems that neuronal loss and the inflamed region of the AD-affected brain induces *GFAP* gene expression. Typically, an increased expression of *GFAP* represents astroglia activation and gliosis, especially during neurodegeneration [51]. In the iPSC-derived and AD neuronal network models, *GFAP* seemed to be expressed at a lower level than in the brain biopsies. However, in almost all AD models, e.g., Kondo

Table 4
KEGG pathway analysis of the AD gene signature

KEGGid	KEGG_name	p_hyper	q_hyper	Genes
hsa04141	Protein processing in endoplasmic reticulum	0.0017	0.3285	FBXO2,MAP3K5, BCL2
hsa05014	Amyotrophic lateral sclerosis	0.0029	0.3285	MAP3K5,BCL2
hsa04722	Neurotrophin signaling pathway	0.0151	1.0000	MAP3K5,BCL2
hsa04950	Maturity onset diabetes of the young	0.0377	1.0000	HES1

Results of KEGG pathway over-representation analysis via hypergeometric test (p_hyper: *p*-value from hypergeometric test, q_hyper: *p*-value adjusted for multiple testing). KEGG pathway over-representation analysis finds Protein processing in endoplasmatic reticulum, amyotrophic lateral sclerosis, and neurotrophin signaling pathway as significant with a *p*-value < 0.05 and at least two genes in the pathway.

Table 5
Overrepresentation of central metabolic KEGG pathways in the analyzed datasets

Dataset	Literature	Dataset_description	KEGG_name	p	FDR
GSE26927	Durrenberger et al., [30]	entorhinal cortex	Glycolysis / Gluconeogenesis	0.77	1.00
GSE28146	Blalock et al., [44]	hippocampus	Glycolysis / Gluconeogenesis	0.21	0.70
GSE28379	Yagi et al., [35]	iPSC AD model EOAD, fAD: PSEN2	Glycolysis / Gluconeogenesis	1.00	1.00
GSE36980	Hokama et al., [28]	hippocampus	Glycolysis / Gluconeogenesis	*0.09*	0.27
GSE43326	Kondo et al., [36]	iPSC AD model fAD: APP	Glycolysis / Gluconeogenesis	0.16	0.45
GSE4757	Dunckley et al., [29]	entorhinal cortex	Glycolysis / Gluconeogenesis	1.00	1.00
GSE42492	Hossini et al., [22]	iPSC AD model LOAD, sAD	Glycolysis / Gluconeogenesis	1.00	1.00
GSE29378	Miller et al., [27]	hippocampus CA1 and CA3	Glycolysis / Gluconeogenesis	0.96	1.00
GSE26927	Durrenberger et al., [30]	entorhinal cortex	Citrate cycle (TCA cycle)	1.00	1.00
GSE28146	Blalock et al., [44]	hippocampus	Citrate cycle (TCA cycle)	0.18	0.65
GSE28379	Yagi et al., [35]	iPSC AD model EOAD, fAD: PSEN2	Citrate cycle (TCA cycle)	1.00	1.00
GSE36980	Hokama et al., [28]	hippocampus	Citrate cycle (TCA cycle)	**1.2E-05**	**4.5E-04**
GSE43326	Kondo et al., [36]	iPSC AD model fAD: APP	Citrate cycle (TCA cycle)	0.41	0.68
GSE4757	Dunckley et al., [29]	entorhinal cortex	Citrate cycle (TCA cycle)	1.00	1.00
GSE42492	Hossini et al., [22]	iPSC AD model LOAD, sAD	Citrate cycle (TCA cycle)	1.00	1.00
GSE29378	Miller et al., [27]	hippocampus CA1 and CA3	Citrate cycle (TCA cycle)	0.40	1.00
GSE26927	Durrenberger et al., [30]	entorhinal cortex	Oxidative phosphorylation	0.10	1.00
GSE28146	Blalock et al., [44]	hippocampus	Oxidative phosphorylation	**0.03**	0.20
GSE28379	Yagi et al., [35]	iPSC AD model EOAD, fAD: PSEN2	Oxidative phosphorylation	1.00	1.00
GSE36980	Hokama et al., [28]	hippocampus	Oxidative phosphorylation	**1.2E-06**	**6.9E-05**
GSE43326	Kondo et al., [36]	iPSC AD model fAD: APP	Oxidative phosphorylation	0.99	1.00
GSE4757	Dunckley et al., [29]	entorhinal cortex	Oxidative phosphorylation	1.00	1.00
GSE42492	Hossini et al., [22]	iPSC AD model LOAD, sAD	Oxidative phosphorylation	0.65	1.00
GSE29378	Miller et al., [27]	hippocampus CA1 and CA3	Oxidative phosphorylation	0.94	1.00
GSE26927	Durrenberger et al., [30]	entorhinal cortex	Insulin signaling pathway	0.84	1.00
GSE28146	Blalock et al., [44]	hippocampus	Insulin signaling pathway	*0.08*	0.38
GSE28379	Yagi et al., [35]	iPSC AD model EOAD, fAD: PSEN2	Insulin signaling pathway	1.00	1.00
GSE36980	Hokama et al., [28]	hippocampus	Insulin signaling pathway	**1.2E-05**	**4.5E-04**
GSE43326	Kondo et al., [36]	iPSC AD model fAD: APP	Insulin signaling pathway	**2.2E-04**	**7.3E-03**
GSE4757	Dunckley et al., [29]	entorhinal cortex	Insulin signaling pathway	0.76	1.00
GSE42492	Hossini et al., [22]	iPSC AD model LOAD, sAD	Insulin signaling pathway	0.67	1.00
GSE29378	Miller et al., [27]	hippocampus CA1 and CA3	Insulin signaling pathway	0.17	1.00

$P < 0.1$ and FDR < 0.1 are marked by italic fonts, $p < 0.05$ and FDR < 0.05 additionally in bold.

et al. (Fig. 4), the expression of *GFAP* is higher in AD patients compared to healthy control individuals.

Our results clearly point to an important role of FOXA1 and FOXA2 in regulating biological processes, which are significantly dysregulated in AD. This finding is in corroboration with reported observations implying that Foxa1 and Foxa2 regulate development of dopaminergic neurons and that the dopaminergic system plays a major role in the development of AD. Thus, the results from this analysis have unveiled a gene regulatory network controlled by FOXA1 and FOXA2. This gene network is involved in development and maintenance of dopaminergic neurons and is associated with AD. The influence of FOXA1 and FOXA2 on the dopaminergic neurons has been described only for the midbrain but not for hippocampus where we found the dysregulation of these transcription factors. However, McNamara et al. previously discussed the influence of dopaminergic neurons on the hip-

pocampus [45] revealing that dopaminergic neurons promote hippocampal reactivation and that axons expressing dopaminergic markers have been found in hippocampus. Furthermore we have elucidated the FOX gene expression in the AD-affected brain and in our iPSC model. FOXA1/FOXA2 are regulating ventral midbrain neural development [25], but beyond that FOXA2 directly induce sonic hedgehog, a key player in forebrain development of inhibitory interneurons. Interneurons express FOXG1 upon progenitor state [47] and are impaired in AD patients [52].

Our results highlight the important role of metabolic processes, particularly oxidative phosphorylation and oxidative stress in the pathology of AD. Furthermore, the FOX transcription factors have already been reported to play a major role in metabolism and induction of oxidative stress. Interestingly, Gao et al. find an inverse regulation of metabolism-related and neuronal genes in Foxa1 and Foxa2 compound knockdown mouse beta cells [53]. Kittappa et al. already discussed that impairment of mitochondrial function via oxidative stress induced by FOXA2 leads to neuronal loss [54]. However, this study was based on dopaminergic neurons in PD but similar mechanisms may also exist in AD. An important difference to PD might be the contribution of other FOX family transcription factors such as the FOXG and FOXO families.

For FOXO transcription factors which we also found significantly enriched in the Opossum analysis, Manolopoulos et al. reported an involvement in the pathogenesis of AD and in insulin resistance via oxidative stress [55]. They hypothesize that ROS activating c-Jun N-terminal kinases (JNK) and inhibiting Wnt-signaling may lead to amyloid-β plaques and tau protein phosphorylation. Additionally, inhibition of Wnt-signaling may further activate FOXO proteins thus leading to a feedback loop of oxidative stress, insulin resistance, ROS, and neurodegeneration. In summary, we hypothesize that the FOXA- and FOXO-mediated mechanisms leading to neuronal loss in AD (Fig. 7) might be as follows: FOXA is involved in the aging-induced dysregulation of metabolic processes including glucose homeostasis [56], carbohydrate metabolism [53], and oxidative phosphorylation [54, 57]. This metabolic dysregulation leads to oxidative stress, the abundant increase of ROS. ROS activates JNK through MAP3K5 (ASK1) [58, 59]. JNK regulates several processes including a cascade of JUNB and BCL2 connected to mitochondrial stress and FOXO-mediated processes involved in neuronal loss: (i) FOXO3 regulates the pro-apoptotic BCL2 [60]. (ii) Aberrant neuronal cell cycle re-entry and dysregulation by FOXO together with its regulators JNK and PI3K-PKB [61] eventually results in neuronal loss [10, 62, 63].

Finally, we propose an AD-iPSC cell model to test our hypothesis, which implies an important role of FOXA1, FOXA2, and FOXO gene regulatory networks in the etiology of AD.

ACKNOWLEDGMENTS

We acknowledge support from the European Union funding/FP7 (FP7/2007-2013)/Grant Agreement n° 305299 (AgedBrainSYSBIO) and from the medical faculty of the Heinrich-Heine Universität Düsseldorf.

Authors' disclosures available online (http://www.j-alz.com/manuscript-disclosures/15-0733r2).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/JAD150733.

REFERENCES

[1] James OG, Doraiswamy PM, Borges-Neto S (2015) PET imaging of tau pathology in Alzheimer's disease and tauopathies. *Front Neurol* **6**, 38.

[2] Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP (2013) The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimers Dement* **9**, 63–75.e2.

[3] Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database. *Nat Genet* **39**, 17-23.

[4] Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921-923.

[5] Brink TC, Demetrius L, Lehrach H, Adjaye J (2009) Age-related transcriptional changes in gene expression in different organs of mice support the metabolic stability theory of aging. *Biogerontology* **10**, 549-564.

[6] Brink TC, Regenbrecht C, Demetrius L, Lehrach H, Adjaye J (2009) Activation of the immune response is a key feature of aging in mice. *Biogerontology* **10**, 721-734.

[7] de la Monte SM, Wands JR (2008) Alzheimer's disease is type 3 diabetes–evidence reviewed. *J Diabetes Sci Technol* **2**, 1101-1113.

[8] Ngo ST, Steyn FJ (2015) The interplay between metabolic homeostasis and neurodegeneration: Insights into the neurometabolic nature of amyotrophic lateral sclerosis. *Cell Regen (Lond)* **4**, 5.

[9] Pistollato F, Cavanaugh SE, Chandrasekera PC (2015) A human-based integrated framework for Alzheimer's disease research. *J Alzheimers Dis* **47**, 857-868.

[10] Demetrius LA, Driver J (2013) Alzheimer's as a metabolic disease. *Biogerontology* **14**, 641-649.

[11] Pellerin L, Magistretti PJ (2004) Neuroenergetics: Calling upon astrocytes to satisfy hungry neurons. *Neuroscientist* **10**, 53-62.

[12] Haroutunian V, Schnaider-Beeri M, Schmeidler. J, Wysocki M, Purohit DP, Perl DP, Libow LS, Lesser GT, Maroukian M, Grossman HT (2008) Role of the neuropathology of Alzheimer disease in dementia in the oldest-old. *Arch Neurol* **65**, 1211-1217.

[13] Savva GM, Wharton SB, Ince PG, Forster G, Matthews FE, Brayne C (2009) Age, neuropathology, and dementia. *N Engl J Med* **360**, 2302-2309.

[14] Yoshii F, Barker WW, Chang JY, Loewenstein D, Apicella A, Smith D, Boothe T, Ginsberg MD, Pascal S, Duara R (1988) Sensitivity of cerebral glucose metabolism to age, gender, brain volume, brain atrophy, and cerebrovascular risk factors. *J Cereb Blood Flow Metab* **8**, 654-661.

[15] Mosconi L, Sorbi S, de Leon MJ, Li Y, Nacmias B, Myoung PS, Tsui W, Ginestroni A, Bessi V, Fayyazz M, Caffarra P, Pupi A (2006) Hypometabolism exceeds atrophy in presymptomatic early-onset familial Alzheimer's disease. *J Nucl Med* **47**, 1778-1786.

[16] Demetrius LA, Magistretti PJ, Pellerin L (2015) Alzheimer's disease: The amyloid hypothesis and the Inverse Warburg effect. *Front Physiol* **5**, 522.

[17] Beach TG, Monsell SE, Phillips LE, Kukull W (2012) Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. *J Neuropathol Exp Neurol* **71**, 266-273.

[18] Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, DeStafano AL, Bis JC, Beecham GW, Grenier-Boley B, Russo G, Thorton-Wells TA, Jones N, Smith AV, Chouraki V, Thomas C, Ikram MA, Zelenika D, Vardarajan BN, Kamatani Y, Lin CF, Gerrish A, Schmidt H, Kunkle B, Dunstan ML, Ruiz A, Bihoreau MT, Choi SH, Reitz C, Pasquier F, Cruchaga C, Craig D, Amin N, Berr C, Lopez OL, De Jager PL, Deramecourt V, Johnston JA, Evans D, Lovestone S, Letenneur L, Morón FJ, Rubinsztein DC, Eiriksdottir G, Sleegers K, Goate AM, Fiévet N, Huentelman MW, Gill M, Brown K, Kamboh MI, Keller L, Barberger-Gateau P, McGuiness B, Larson EB, Green R, Myers AJ, Dufouil C, Todd S, Wallon D, Love S, Rogaeva E, Gallacher. J, St George-Hyslop P, Clarimon J, Lleo A, Bayer A, Tsuang DW, Yu L, Tsolaki M, Bossù P, Spalletta G, Proitsi P, Collinge J, Sorbi S, Sanchez-Garcia F, Fox NC, Hardy J, Deniz Naranjo MC, Bosco P, Clarke R, Brayne C, Galimberti D, Mancuso M, Matthews F, European Alzheimer's Disease Initiative (EADI), Genetic and Environmental Risk in Alzheimer's Disease, Alzheimer's Disease Genetic Consortium, Cohorts for Heart and Aging Research in Genomic Epidemiology, Moebus S, Mecocci P, Del Zompo M, Maier W, Hampel H, Pilotto A, Bullido M, Panza F, Caffarra P, Nacmias B, Gilbert JR, Mayhaus M, Lannefelt L, Hakonarson H, Pichler S, Carrasquillo MM, Ingelsson M, Beekly D, Alvarez V, Zou F, Valladares O, Younkin SG, Coto E, Hamilton-Nelson KL, Gu W, Razquin C, Pastor P, Mateo I, Owen MJ, Faber KM, Jonsson PV, Combarros O, O'Donovan MC, Cantwell LB, Soininen H, Blacker D, Mead S, Mosley TH, Bennett DA, Harris TB, Fratiglioni L, Holmes C, de Bruijn RF, Passmore P, Montine TJ, Bettens K, Rotter JI, Brice A, Morgan K, Foroud TM, Kukull WA, Hannequin D, Powell JF, Nalls MA, Ritchie K, Lunetta KL, Kauwe JS, Boerwinkle E, Riemenschneider M, Boada M, Hiltuenen M, Martin ER, Schmidt R, Rujescu D, Wang LS, Dartigues JF, Mayeux R, Tzourio C, Hofman A, Nöthen MM, Graff C, Psaty BM, Jones L, Haines JL, Holmans PA, Lathrop M, Pericak-Vance MA, Launer LJ, Farrer LA, van Duijn CM, Van Broeckhoven C, Moskvina V, Seshadri S, Williams J, Schellenberg GD, Amouyel P (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-1458.

[19] Swardfager W, Lanctôt K, Rothenburg L, Wong A, Cappell J, Herrmann N (2010) A meta-analysis of cytokines in Alzheimer's disease. *Biol Psychiatry* **68**, 930-941.

[20] Sitzer DI, Twamley EW, Jeste DV (2006) Cognitive training in Alzheimer's disease: A meta-analysis of the literature. *Acta Psychiatr Scand* **114**, 75-90.

[21] Antonell A, Lladó A, Altirriba J, Botta-Orfila T, Balasa M, Fernández M, Ferrer I, Sánchez-Valle R, Molinuevo JL (2013) A preliminary study of the whole-genome expression profile of sporadic and monogenic early-onset Alzheimer's disease. *Neurobiol Aging* **34**, 1772-1778.

[22] Hossini AM, Megges M, Prigione A, Lichtner B, Toliat MR, Wruck W, Schröter F, Nuernberg P, Kroll H, Makrantonaki E, Zouboulis CC, Zouboulis CC, Adjaye J (2015) Induced pluripotent stem cell-derived neuronal cells from a sporadic Alzheimer's disease donor as a model for investigating AD-associated gene regulatory networks. *BMC Genomics* **16**, 84.

[23] Trillo L, Das D, Hsieh W, Medina B, Moghadam S, Lin B, Dang V, Sanchez MM, De Miguel Z, Ashford JW, Salehi A (2013) Ascending monoaminergic systems alterations in Alzheimer's disease translating basic science into clinical care. *Neurosci Biobehav Rev* **37**, 1363-1379.

[24] Borroni B, Agosti C, Archetti S, Costanzi C, Bonomi S, Ghianda D, Lenzi GL, Caimi L, Luca MD, Padovani A (2004) Catechol-O-methyltransferase gene polymorphism is associated with risk of psychosis in Alzheimer Disease. *Neurosci Lett* **370**, 127-129.

[25] Ferri ALM, Lin W, Mavromatakis YE, Wang JC, Sasaki H, Whitsett JA, Ang S-L (2007) Foxa1 and Foxa2 regulate multiple phases of midbrain dopaminergic neuron development in a dosage-dependent manner. *Dev Camb Engl* **134**, 2761-2769.

[26] Stott SRW, Metzakopian E, Lin W, Kaestner KH, Hen R, Ang S-L (2013) Foxa1 and foxa2 are required for the maintenance of dopaminergic properties in ventral midbrain neurons at late embryonic stages. *J Neurosci* **33**, 8022-8034.

[27] Miller JA, Woltjer RL, Goodenbour JM, Horvath S, Geschwind DH (2013) Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Med* **5**, 48.

[28] Hokama M, Oka S, Leon J, Ninomiya T, Honda H, Sasaki K, Iwaki T, Ohara T, Sasaki T, LaFerla FM, Kiyohara Y, Nakabeppu Y (2014) Altered expression of diabetes-related genes in Alzheimer's disease brains: The Hisayama study. *Cereb Cortex* **24**, 2476-2488.

[29] Dunckley T, Beach TG, Ramsey KE, Grover A, Mastroeni D, Walker DG, LaFleur BJ, Coon KD, Brown KM, Caselli R, Kukull W, Higdon R, McKeel D, Morris JC, Hulette C, Schmechel D, Reiman EM, Rogers J, Stephan DA (2006) Gene expression correlates of neurofibrillary tangles in Alzheimer's disease. *Neurobiol Aging* **27**, 1359-1371.

[30] Durrenberger PF, Fernando FS, Magliozzi R, Kashefi SN, Bonnert TP, Ferrer I, Seilhean D, Nait-Oumesmar B,

Schmitt A, Gebicke-Haerter PJ, Falkai P, Grünblatt E, Palkovits M, Parchi P, Capellari S, Arzberger T, Kretzschmar H, Roncaroli F, Dexter DT, Reynolds R (2012) Selection of novel reference genes for use in the human central nervous system: A BrainNet Europe Study. *Acta Neuropathol (Berl)* **124**, 893-903.

[31] Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J (2004) Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80.

[32] Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307-315.

[33] Du P, Kibbe WA, Lin SM (2008) lumi: A pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547-1548.

[34] Carvalho BS, Irizarry RA (2010) A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363-2367.

[35] Yagi T, Ito D, Okada Y, Akamatsu W, Nihei Y, Yoshizaki T, Yamanaka S, Okano H, Suzuki N (2011) Modeling familial Alzheimer's disease with induced pluripotent stem cells. *Hum Mol Genet* **20**, 4530-4539.

[36] Kondo T, Asai M, Tsukita K, Kutoku Y, Ohsawa Y, Sunada Y, Imamura K, Egawa N, Yahata N, Okita K, Takahashi K, Asaka I, Aoi T, Watanabe A, Watanabe K, Kadoya C, Nakano R, Watanabe D, Maruyama K, Hori O, Hibino S, Choshi T, Nakahata T, Hioki H, Kaneko T, Naitoh M, Yoshikawa K, Yamawaki S, Suzuki S, Hata R, Ueno S-I, Seki T, Kobayashi K, Toda T, Murakami K, Irie K, Klein WL, Mori H, Asada T, Takahashi R, Iwata N, Yamanaka S, Inoue H (2013) Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular Aβ and differential drug responsiveness. *Cell Stem Cell* **12**, 487-496.

[37] Israel MA, Yuan SH, Bardy C, Reyna SM, Mu Y, Herrera C, Hefferan MP, Van Gorp S, Nazor KL, Boscolo FS, Carson CT, Laurent LC, Marsala M, Gage FH, Remes AM, Koo EH, Goldstein LSB (2012) Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. *Nature* **482**, 216-220.

[38] Kwon AT, Arenillas DJ, Worsley Hunt R, Wasserman WW (2012) oPOSSUM-3: Advanced analysis of regulatory motif over-representation across genes or ChIP-Seq datasets. *G3 (Bethesda)* **2**, 987-1002.

[39] Butts C (2008) network: A package for managing relational data in R. *J Stat Softw* **24**, doi: 10.18637/jss.v024.i02

[40] Ihaka R, Gentleman R (1996) R: A language for data analysis and graphics. *J Comput Graph Stat* **5**, 299-314.

[41] Storey JD (2002) A direct approach to false discovery rates. *J R Stat Soc Ser B Stat Methodol* **64**, 479-498.

[42] Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. *Bioinformatic* **23**, 257-258.

[43] Bertram L, Lill CM, Tanzi RE (2010) The genetics of Alzheimer disease: Back to the future. *Neuron* **68**, 270-281.

[44] Blalock EM, Buechel HM, Popovic J, Geddes JW, Landfield PW (2011) Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. *J Chem Neuroanat* **42**, 118-126.

[45] McNamara CG, Tejero-Cantero Á, Trouche S, Campo-Urriza N, Dupret D (2014) Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. *Nat Neurosci* **17**, 1658-1660.

[46] Herrera E, Marcus R, Li S, Williams SE, Erskine L, Lai E, Mason C (2004) Foxd1 is required for proper formation of the optic chiasm. *Dev Camb Engl* **131**, 5727-5739.

[47] Nicholas CR, Chen J, Tang Y, Southwell DG, Chalmers N, Vogt D, Arnold CM, Chen Y-JJ, Stanley EG, Elefanty AG, Sasai Y, Alvarez-Buylla A, Rubenstein JLR, Kriegstein AR (2013) Functional maturation of hPSC-derived forebrain interneurons requires an extended timeline and mimics human neural development. *Cell Stem Cell* **12**, 573-586.

[48] Tesseur I, Zou K, Esposito L, Bard F, Berber E, Can JV, Lin AH, Crews L, Tremblay P, Mathews P, Mucke L, Masliah E, Wyss-Coray T (2006) Deficiency in neuronal TGF-beta signaling promotes neurodegeneration and Alzheimer's pathology. *J Clin Invest* **116**, 3060-3069.

[49] Paradis E, Douillard H, Koutroumanis M, Goodyer C, LeBlanc A (1996) Amyloid beta peptide of Alzheimer's disease downregulates Bcl-2 and upregulates bax expression in human neurons. *J Neurosci* **16**, 7533-7539.

[50] Boom A, Pochet R, Authelet M, Pradier L, Borghgraef P, Van Leuven F, Heizmann CW, Brion J-P (2004) Astrocytic calcium/zinc binding protein S100A6 over expression in Alzheimer's disease and in PS1/APP transgenic mice models. *Biochim Biophys Acta* **1742**, 161-168.

[51] Heppner FL, Ransohoff RM, Becher B (2015) Immune attack: The role of inflammation in Alzheimer disease. *Nat Rev Neurosci* **16**, 358-372.

[52] Levenga J, Krishnamurthy P, Rajamohamedsait H, Wong H, Franke TF, Cain P, Sigurdsson EM, Hoeffer CA (2013) Tau pathology induces loss of GABAergic interneurons leading to altered synaptic plasticity and behavioral impairments. *Acta Neuropathol Commun* **1**, 34.

[53] Gao N, Le Lay J, Qin W, Doliba N, Schug J, Fox AJ, Smirnova O, Matschinsky FM, Kaestner KH (2010) Foxa1 and Foxa2 maintain the metabolic and secretory features of the mature beta-cell. *Mol Endocrinol* **24**, 1594-1604.

[54] Kittappa R, Chang WW, Awatramani RB, McKay RDG (2007) The foxa2 gene controls the birth and spontaneous degeneration of dopamine neurons in old age. *PLoS Biol* **5**, e325.

[55] Manolopoulos KN, Klotz L-O, Korsten P, Bornstein SR, Barthel A (2010) Linking Alzheimer's disease to insulin resistance: The FoxO response to oxidative stress. *Mol Psychiatry* **15**, 1046-1052.

[56] Friedman JR, Kaestner KH (2006) The Foxa family of transcription factors in development and metabolism. *Cell Mol Life Sci* **63**, 2317-2328.

[57] Vatamaniuk MZ, Gupta RK, Lantz KA, Doliba NM, Matschinsky FM, Kaestner KH (2006) Foxa1-deficient mice exhibit impaired insulin secretion due to uncoupled oxidative phosphorylation. *Diabetes* **55**, 2730-2736.

[58] Tobiume K, Matsuzawa A, Takahashi T, Nishitoh H, Morita K, Takeda K, Minowa O, Miyazono K, Noda T, Ichijo H (2001) ASK1 is required for sustained activations of JNK/p38 MAP kinases and apoptosis. *EMBO Rep* **2**, 222-228.

[59] Kadowaki H, Nishitoh H, Urano F, Sadamitsu C, Matsuzawa A, Takeda K, Masutani H, Yodoi J, Urano Y, Nagano T, Ichijo H (2005) Amyloid beta induces neuronal cell death through ROS-mediated ASK1 activation. *Cell Death Differ* **12**, 19-24.

[60] Dijkers PF, Medema RH, Lammers JW, Koenderman L, Coffer PJ (2000) Expression of the pro-apoptotic Bcl-2 fam-

ily member Bim is regulated by the forkhead transcription factor FKHR-L1. *Curr Biol* **10**, 1201-1204.

[61] Eijkelenboom A, Burgering BMT (2013) FOXOs: Signalling integrators for homeostasis maintenance. *Nat Rev Mol Cell Biol* **14**, 83-97.

[62] Yuan Z, Becker EBE, Merlo P, Yamada T, DiBacco S, Konishi Y, Schaefer EM, Bonni A (2008) Activation of FOXO1 by Cdk1 in cycling cells and postmitotic neurons. *Science* **319**, 1665-1668.

[63] Zhu X, Lee H, Perry G, Smith MA (2007) Alzheimer disease, the two-hit hypothesis: An update. *Biochim Biophys Acta* **1772**, 494-502.

[64] Orcholski ME, Zhang Q, Bredesen DE (2011) Signaling via amyloid precursor-like proteins APLP1 and APLP2. *J Alzheimers Dis* **23**, 689-699.

[65] Demetrius LA, Simon DK (2013) The inverse association of cancer and Alzheimer's: A bioenergetic mechanism. *J R Soc Interface* **10**, 20130006-20130006.

[66] Ghani M, Sato C, Lee JH, Reitz C, Moreno D, Mayeux R, St George-Hyslop P, Rogaeva E (2013) Evidence of recessive Alzheimer disease loci in a Caribbean Hispanic data set: Genome-wide survey of runs of homozygosity. *JAMA Neurol* **70**, 1261-1267.

[67] Jun G, Moncaster JA, Koutras C, Seshadri S, Buros J, McKee AC, Levesque G, Wolf PA, St George-Hyslop P, Goldstein LE, Farrer LA (2012) δ-Catenin is genetically and biologically associated with cortical cataract and future Alzheimer-related structural and functional brain changes. *PloS One* **7**, e43728.

[68] Atkin G, Hunt J, Minakawa E, Sharkey L, Tipper N, Tennant W, Paulson HL (2014) F-box only protein 2 (Fbxo2) regulates amyloid precursor protein levels and processing. *J Biol Chem* **289**, 7038-7048.

[69] Tseveleki V, Rubio R, Vamvakas S-S, White J, Taoufik E, Petit E, Quackenbush J, Probert L (2010) Comparative gene expression analysis in mouse models for multiple sclerosis, Alzheimer's disease and stroke for identifying commonly regulated and disease-specific gene changes. *Genomics* **96**, 82-91.

[70] Gentil BJ, Delphin C, Mbele GO, Deloulme JC, Ferro M, Garin (3261) J, Baudier J (2001) The giant protein AHNAK is a specific target for the calcium- and zinc-binding S100B protein: Potential implications for Ca2+homeostasis regulation by S100B. *J Biol Chem* **276**, 23253-23252.

[71] Liang D, Han G, Feng X, Sun J, Duan Y, Lei H (2012) Concerted perturbation observed in a hub network in Alzheimer's disease. *PloS One* **7**, e40498.

[72] Mann MB, Wu S, Rostamkhani M, Tourtellotte W, MacMurray J, Comings DE (2001) Phenylethanolamine N-methyltransferase (PNMT) gene and early-onset Alzheimer disease. *Am J Med Genet* **105**, 312-316.

[73] Sekar S, McDonald J, Cuyugan L, Aldrich J, Kurdoglu A, Adkins J, Serrano G, Beach TG, Craig DW, Valla J, Reiman EM, Liang WS (2015) Alzheimer's disease is associated with altered expression of genes involved in immune response and mitochondrial processes in astrocytes. *Neurobiol Aging* **36**, 583-591.

[74] Cruchaga C, Kauwe JSK, Harari O, Jin SC, Cai Y, Karch CM, Benitez BA, Jeng AT, Skorupa T, Carrell D, Bertelsen S, Bailey M, McKean D, Shulman JM, De Jager PL, Chibnik L, Bennett DA, Arnold SE, Harold D, Sims R, Gerrish A, Williams J, Van Deerlin VM, Lee VM-Y, Shaw LM, Trojanowski JQ, Haines JL, Mayeux R, Pericak-Vance MA, Farrer LA, Schellenberg GD, Peskind ER, Galasko D, Fagan AM, Holtzman DM, Morris JC, GERAD Consortium, Alzheimer's Disease Neuroimaging Initiative (ADNI), Alzheimer Disease Genetic Consortium (ADGC), Goate AM (2013) GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron* **78**, 256-268.

[75] Parachikova A, Cotman CW (2007) Reduced CXCL12/CXCR4 results in impaired learning and is downregulated in a mouse model of Alzheimer disease. *Neurobiol Dis* **28**, 143-153.

[76] Yan J, Kim S, Nho K, Chen R, Risacher SL, Moore JH, Saykin AJ, Shen L, Alzheimer's Disease Neuroimaging Initiative (2015) Hippocampal transcriptome-guided genetic analysis of correlated episodic memory phenotypes in Alzheimer's disease. *Front Genet* **6**, 117.

[77] Pey P, Pearce RKB, Kalaitzakis ME, Griffin WST, Gentleman SM (2014) Phenotypic profile of alternative activation marker CD163 is different in Alzheimer's and Parkinson's disease. *Acta Neuropathol Commun* **2**, 21.

[78] Wilhelmus MMM, Otte-Höller I, Wesseling P, de Waal RMW, Boelens WC, Verbeek MM (2006) Specific association of small heat shock proteins with the pathological hallmarks of Alzheimer's disease brains. *Neuropathol Appl Neurobiol* **32**, 119-130.

[79] Chacón PJ, Rodríguez-Tébar A (2012) Increased expression of the homologue of enhancer-of-split 1 protects neurons from beta amyloid neurotoxicity and hints at an alternative role for transforming growth factor beta1 as a neuroprotector. *Alzheimers Res Ther* **4**, 31.

[80] Anderson AJ, Cummings BJ, Cotman CW (1994) Increased immunoreactivity for Jun- and Fos-related proteins in Alzheimer's disease: Association with pathology. *Exp Neurol* **125**, 286–295.

2.1.3 Concise Review: Current Status and Future Directions on Research Related to Nonalcoholic Fatty Liver Disease.

Stem Cells. 2017 Jan;35(1):89-96. doi: 10.1002/stem.2454. Epub 2016 Jul 13.

Wruck W(1), Graffmann N(1), Kawala MA(1), Adjaye J(1).

Author information: (1)Medical Faculty, Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, Düsseldorf, Germany.

Considered a feature of the metabolic syndrome, nonalcoholic fatty liver disease (NAFLD), is associated with insulin resistance, type 2 diabetes, obesity and drug toxicity. Its prevalence is estimated at about 30% in western countries mainly due to sedentary life styles and high fat diets. Genome-wide association studies have identified polymorphisms in several genes, for example, PNPLA3, and TM6SF2 which confer susceptibility to NAFLD. Here, we review recent findings in the NAFLD field with a particular focus on published transcriptomics datasets which we subject to a meta-analysis. We reveal a common gene signature correlating with the progression of the disease from steatosis and steatohepatitis and reveal that lipogenic and cholesterol metabolic pathways are main actors in this signature. We propose the use of disease-in-a-dish models based on hepatocyte-like cells derived from patient-specific induced pluripotent stem cells (iPSC). These will enable investigations into the contribution of genetic background in the progression from NALFD to non-alcoholic steatohepatitis. Furthermore, an iPSC-based approach should aid in the elucidation of the function of new biomarkers, thus enabling better diagnostic tests and validation of potential drug targets. Stem Cells 2017;35:89-96.

DOI: 10.1002/stem.2454

PMID: 27374784 [Indexed for MEDLINE]

URL: https://stemcellsjournals.onlinelibrary.wiley.com/doi/full/10.1002/stem.2454

Contribution by Wasco Wruck: Manuscript writing, data analysis and interpretation.

STEM CELLS

CONCISE REVIEWS

Current Status and Future Directions on Research Related to Nonalcoholic Fatty Liver Disease

WASCO WRUCK, NINA GRAFFMANN, MARIE-ANN KAWALA, JAMES ADJAYE

Key Words. Nonalcoholic fatty liver disease • Nonalcoholic steatohepatitis • Meta-analysis • Steatosis • microRNA • Induced pluripotent stem cells

Medical Faculty, Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, Düsseldorf, Germany

Correspondence: Prof. Dr. James Adjaye, Medical Faculty, Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, 40225 Düsseldorf, Germany. Telephone: +49 (0) 211 81 – 08191; Fax: +49 (0) 211 81 – 17858; e-mail: James.Adjaye@med.uni-duesseldorf.de

Received March 14, 2016; accepted for publication June 18, 2016; first published online in STEM CELLS *EXPRESS* July 4, 2016.

1066-5099/2016/$30.00/0

http://dx.doi.org/10.1002/stem.2454

ABSTRACT

Considered a feature of the metabolic syndrome, nonalcoholic fatty liver disease (NAFLD), is associated with insulin resistance, type 2 diabetes, obesity and drug toxicity. Its prevalence is estimated at about 30% in western countries mainly due to sedentary life styles and high fat diets. Genome-wide association studies have identified polymorphisms in several genes, for example, *PNPLA3*, and *TM6SF2* which confer susceptibility to NAFLD. Here, we review recent findings in the NAFLD field with a particular focus on published transcriptomics datasets which we subject to a meta-analysis. We reveal a common gene signature correlating with the progression of the disease from steatosis and steatohepatitis and reveal that lipogenic and cholesterol metabolic pathways are main actors in this signature. We propose the use of disease-in-a-dish models based on hepatocyte-like cells derived from patient-specific induced pluripotent stem cells (iPSC). These will enable investigations into the contribution of genetic background in the progression from NALFD to non-alcoholic steatohepatitis. Furthermore, an iPSC-based approach should aid in the elucidation of the function of new biomarkers, thus enabling better diagnostic tests and validation of potential drug targets. STEM CELLS *2016; 00:000–000*

SIGNIFICANCE STATEMENT

The continuously increasing prevalence of nonalcoholic fatty liver disease (NAFLD) estimated at approximately 30% in western countries and the paucity of available publications incorporating epidemiology, genome-wide association studies, epigenetics and meta-analysis of transcriptome data, served as impetus for this review article. We reveal a NAFLD-gene signature which correlates with the progression of the disease from steatosis to non-alcoholic steatohepatitis, fibrosis, cirrhosis, and eventually liver cancer. We also propose the implementation of patient-specific disease models based on induced pluripotent stem cells differentiated into hepatocytes and challenged with oleic acid to enable a better understanding of the molecular mechanisms underlying the etiology of NAFLD.

INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD) has an increasingly large impact on public health as it is estimated at about 30% prevalence in western countries. This is in contrast with developing countries where sedentary life style and lack of exercise are not yet widespread. Burden to public health is further underlined by the fact that NAFLD is the second most common reason for liver transplantation [1]. As a consequence, there is an urgent need for non-invasive diagnostic tests which are complication-free unlike liver biopsies-based diagnosis. Imaging techniques have potential but also some limitations in non-invasive detection of NAFLD [2]: while ultrasonography (US) is cost-effective but lacking accuracy magnetic resonance (MR) spectroscopy (MRS) and imaging (MRI) are at the moment the most accurate imaging techniques for detection of steatosis. US elastography and MR elastography are promising methods for diagnosis of liver fibrosis.

The etiology of NAFLD and its progression is caused by a complex interplay of genetic and environmental factors. The involvement of genetic background is assessed by genome-wide association studies (GWAS) [3, 4] while a plethora of gene expression, proteomics and metabolomics studies explore environmental factors and their impact on disease-associated molecular pathways. An emerging theme is the function of Peroxisome proliferator-activated receptor α (PPARα) in lipid homeostasis and its lower expression in NAFLD patients. Additionally Montagner et al.

reported that hepatocyte-specific deletion of PPARα is sufficient to promote NAFLD in mice [5]. Ratziu et al. [6] demonstrated the resolution of non-alcoholic steatohepatitis (NASH) by the PPARα agonist Elafibranor. Another aspect gaining attention in the NAFLD community is the role of polyunsaturated fatty acids (PUFAs) in the pathogenesis of NAFLD and NASH [7]. Arendt et al. found lower hepatic PUFAs in patients with NASH compared to simple steatosis—possibly due to overexpression of FADS1 and FADS2 [8].

Disease-models for NAFLD and NASH have been proposed [9] but a better understanding of the complex mechanisms involved in NAFLD and its progression will only be possible when patient-specific differences are accounted for. This can be achieved by disease-models based on hepatocyte-like cells derived from induced pluripotent stem cells (iPSCs) [10].

Here, we review and analyse the progress of NAFLD from steatosis to NASH employing previously published gene expression datasets [11–13] and examine which of the biomarkers, signaling and metabolic pathways found in this meta-analysis are manifested in the progression from low- to high-grade steatosis [13].

Genome-Wide Association Studies

GWAS has identified polymorphisms in numerous genes such as patatin-like phospholipase domain containing protein (*PNPLA3*) [14], and Transmembrane 6 Superfamily Member 2 (*TM6SF2*) [4] which correlate with NAFLD. The relevant *PNPLA3* rs738409 (I148M) variant has a minor allele frequency (MAF) of 49% in Hispanics, 23% in those of European ancestry and 17% in African Americans [15] which is much higher than MAF for the relevant *TM6SF2* polymorphism rs58542926 (E167K) of 7% in Europeans, 4% in Hispanics, and 2% in African Americans [16]. Individuals with these variants have a higher risk of developing NAFLD (*PNPLA3* rs738409: odds ratio(OR) = 3.26 (95% CI: 2.11–7.21) [3]; TM6SF2 rs58542926: OR = 2.13 (95% CI: 1.36–3.30) [17]). However, the variants are only one piece of the jigsaw in the development of NAFLD as there are many individuals carrying the variants but are disease-free. The mechanisms of *PNPLA3* action still needs to be elucidated in detail. Recently, Pirazzi et al. reported promising results showing high level expression of *PNPLA3* in hepatic stellate cells (HSCs) and reduction of *PNPLA3* enzymatic lipase activity in retinyl palmitate lysis into retinol and palmitic acid in the I148M variant [18]. Eichmann et al. further showed that Perilipin 2 (*PLIN2*), adipose TG lipase (*ATGL*; alias *PNPLA2*) and co-activator comparative gene identification-58 (CGI-58) are components of lipid droplets in rat stellate cells and that also *ATGL* effectively hydrolises triglycerides and retinol esters [19]. *TM6SF2* has a dichotomic characteristic: the minor allele appears to increase lipid droplet area and decrease hepatic very-low-density lipoprotein (VLDL) excretion while the predominant allele appears to increase hepatic VLDL excretion and a consequence of this is the risk of cardiovascular disease [20].

Sources of Liver Fat

Surplus energy from a plethora of food not matched by energy expenditure via physical activity results in higher fat levels in the body. Fat can be stored mainly as triglycerides in adipose tissue or in liver. Browning et al. refer to the conventional explanation for hepatic triglyceride accumulation stating that obesity and insulin resistance result in increased release of free fatty acids (FFAs) through hydrolysis from adipocytes [21]. Uptake of FFAs by the liver is proportional to FFA concentrations in the blood [22]. In the liver, FFAs are then either catabolized by β-oxidation in the mitochondria or esterified to triglycerides. Steatosis results from impairment in one of these pathways.

To complement the conventional explanation, hepatic lipogenesis has to be added to these fluxes to and from the liver which according to Donnelly et al. 59.0 ± 9.9% originate from non-esterified fatty acids (FAs) in adipose tissue, 26.1 ± 6.7%, from de novo lipogenesis (DNL) and 14.9 ± 7.0%, from diet [23]. However, a detailed analysis of the origin of liver fat revealed that individuals with high liver fat had more than threefold higher DNL than individuals with low liver fat while having similar levels of adipose flux of FFAs or production of VLDL triglycerides from FFAs [24]. Thus, DNL emerges as the major contributor to the development of steatosis besides lipolysis.

Progression from NAFLD to NASH and Fibrosis

Progression of non-alcoholic steatosis to steatohepatitis has been described by the "two-hit" hypothesis: triglyceride accumulation in the liver is the "first hit" and oxidative stress or induction of cytokines the "second hit," finally resulting in steatohepatitis and in some cases fibrosis. However, related mechanisms may be involved in the two hits because FFA can increase oxidative stress and triglycerides might accumulate in the liver to protect against this [25].

HSCs play a major role in the progression to fibrosis. In healthy tissue, HSCs store lipid droplets containing about 80% of retinoids and do not proliferate. In injured liver fat and retinoids are removed from HSCs and they start to proliferate. Endoplasmic reticulum (ER) stress and oxidative stress can trigger this state through autophagy [26]. This state is defined by the marker, α-smooth-muscle actin and results in the trans-differentiation of HSCs into myofibroblasts and generation of fibrotic connective tissue (fibrogenesis) [27]. The autophagy model is confirmed by Gonzalez-Rodriguez et al. who report a pattern of the autophagy markers p62 and increased LC3-II/LC3-I ratio during the progression of NAFLD and hypothesize that elevated ER stress disrupts the autophagic flux leading to accumulation of unfolded proteins and damaged organelles, triggering liver injury the and apoptotic cell death [28].

Carpino et al. describe that, in the lipotoxicity-mediated progression of NAFLD, prolonged hepatocyte apoptosis and cell cycle arrest induced by oxidative stress can trigger the proliferation and activation of hepatic stem/progenitor cells (HPCs) [29]. A compartment of HPCs which can differentiate into hepatocytes and cholangiocytes resides within the intrahepatic biliary tree. HPCs are activated and recruited only in cases of massive injuries. The activation of HPCs correlates with ductular reaction initiating fibrogenesis either through production of chemotactic agents for inflammatory cells activating HSCs or through ductular epithelial-mesenchymal

transition contributing to the portal myofibroblast pool. Liver macrophages, the Kupffer cells, are involved in progression to NASH via transformation into foam cells due to excessive accumulation of lipids. Adipokines such as Adiponectin and Resistin usually secreted by adipose tissue but also by HPCs mediate the cross-talk among HSCs, HPCs and liver macrophages.

There is an urgent need for non-invasive biomarkers of NASH and fibrosis because liver biopsy as the diagnostic "gold standard" can result in undesirable complications, furthermore, imaging techniques can detect steatosis but not steatohepatitis. Tests for fibrosis have better sensitivity and specificity while for NASH, cytokeratin 18 (CK18/KRT18) seems to perform better [30]. Despite slight improvements by new assays or algorithm tests combining CK18 with other markers or parameters, tests for NASH still lack sensitivity and specificity. For diagnosing fibrosis satisfactory test panels exist, for example, the FibroTest [31] can distinguish advanced fibrosis from not-advanced stages quite accurately but lacks this accuracy when it comes to distinguishing the non-advanced stages into moderate, mild or no fibrosis. Ratziu et al. claim that FibroTest provides a continuous quantitative assessment of liver fibrosis in 100% of patients without indeterminate cases [31]. However, Angulo et al. interprete the study by Ratziu et al., that in 33% of the patients the FibroTest value was in between the proposed cutoffs of 0.30 and 0.70, and thus, unable to predict the presence or absence of advanced fibrosis [32]. Moreover, they demonstrate superior performance of their own simple fibrosis score based on routine clinical parameters. New putative biomarkers have been proposed by Moylan et al. [33] which might improve diagnosis. The importance of classifying fibrosis is emphasized by Angulo et al., they found that fibrosis stage and to a lesser extent liver histological feature of NASH are associated with the long-term outcome of NAFLD patients [34].

Meta-Analysis of NAFLD Progression

To enable better molecular understanding of the progression of NAFLD, a meta-analysis (methods described in Supporting Information Materials and Methods) of publicly accessible NAFLD datasets was performed (Fig. 1). Progression from NAFLD steatosis to NASH was investigated in datasets from du Plessis et al. [35], Ahrens et al. [12] and Horvath et al. [11], this was then compared to the progression from low- to high-grade steatosis utilizing the dataset from Wruck et al. [13]. This strategy was tailored to find early detectable biomarkers of disease progression to NASH. The analysis was constrained to NASH and did not include fibrosis-related datasets, for example, the dataset of Moylan et al. [33], which have a distinct fibrosis characteristic. Moylan et al. revealed that the severity of fibrosis is reflected in the liver transcriptome and can be reduced to a gene signature defining severe NAFLD. Our meta-analysis focused on microarray-based datasets from liver biopsies, for a more comprehensive view of the disease, a multi-omics approach is needed (Fig. 1). This comprises analyses of serum samples, liver biopsies and iPSC-derived hepatocyte-like cells challenged with oleic acid to mimic steatosis in vitro and analysed at the metabolomic, proteomic, lipidomic and secretomic level. Figure 2A shows a venn diagram of genes significantly ($p < .05$) associated with the progression of NAFLD to NASH based on datasets from du Plessis et al. [35], Horvath et al. [11], Ahrens et al. [12] and Wruck et al. [13]. Figure 2B lists the DAVID terms associated with subsets of genes exclusively significant in the dedicated datasets (marked with corresponding colored frames) from the venn diagram (complete tables of gene correlations, venn diagram subsets and DAVID analyses can be found in Supporting Information Tables 1 and 2). Based on the intersection of all datasets, a signature of 22 genes (marked with a blue frame) was unveiled. The most significant pathways found via DAVID functional annotation are cholesterol, fatty acid, lipid synthesis, and metabolism (Fig. 2C). Interestingly, oxido-reductase is among the top categories, this enzyme is associated with oxidative stress and inflammation which lead to NASH. The common 22-gene signature mostly associated with metabolic pathways is listed in Figure 2D. HPRT1 is central to purine metabolism, SQLE and CYP51A1 are central enzymes in cholesterol biosynthesis [36, 37], FASN catalyzes the synthesis of palmitate and long-chain saturated FAs, and FADS1 and FADS2 regulate unsaturation of FAs.

The correlation of cholesterol synthesis and metabolism with the severity of NAFLD has been described in numerous studies, for example, Min et al. reported an association of hepatic synthesis and dysregulation of cholesterol metabolism with the severity of NAFLD [38]. Musso et al. identified a pivotal role of cholesterol content in the mitochondrial membrane and the membrane of the ER in the induction of NASH [39]: when cholesterol content in the mitochondrial membrane is high, transport of glutathione to the mitochondria is reduced thus impairing the regulation of reactive oxygen species (ROS). De-regulation of ROS levels leads to oxidative stress and sensitization of hepatocytes to the action of pro-inflammatory signals. Within the ER membrane, the ratio of free cholesterol to phospholipids is needed to maintain its fluidity and changes result in sarco(endo)plasmic reticulum Ca2+-ATPase (SERCA) inhibition, a low Ca^{2+} concentration in the ER and ER stress, unfolded protein response and eventually inflammation.

Clinical Trials

Since diagnostic methods for NAFLD still remain imprecise, many research groups are embarking on large efforts to identify accurate biomarkers. Recently, Gorden et al., published an "omics" approach enabling distinguishing between steatosis and NASH [40]. They describe a panel of 20 plasma metabolites for the detection of a specific disease signature. Especially ether-linked phospholipids, certain neutral lipids and FAs show suggestive differences between NASH and steatosis. Furthermore, they suggest NOD-like receptor pyrin domain containing 3 as a marker for fibrosis. Due to the lack of effective biomarkers there are few definite pharmacotherapy treatment options available to NAFLD patients. Current therapy recommendations include weight reduction by diet, physical exercise or weight reduction drugs (e.g., orlistat, sibutramine), lipid metabolism-

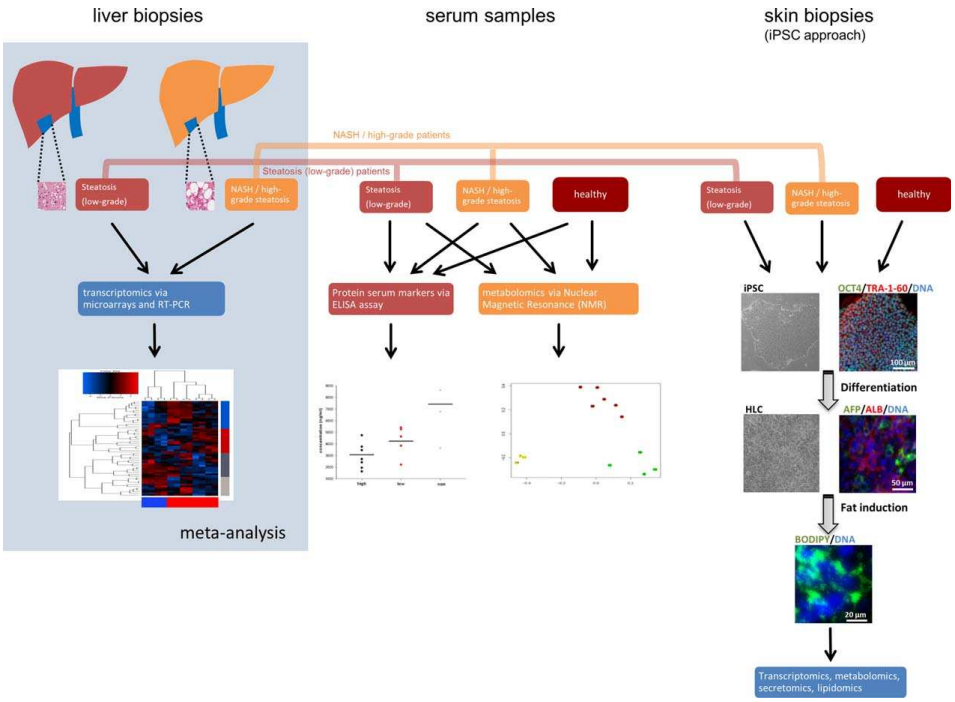
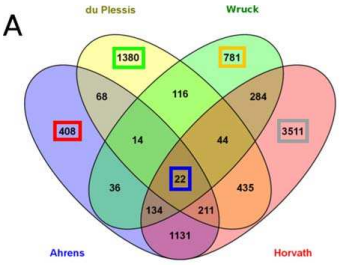


Figure 1. Scheme of experiments for multi-*omics* comparison of steatosis grades. (adapted from Wruck et al. [13]). In the figure, a multi-*omics* experiment design for stratification of nonalcoholic fatty liver disease severity is shown. Patients from du Plessis et al. [35], Ahrens et al. [12] and Horvath et al. [11] classified as NASH and steatosis and patients from Wruck et al. [13] classified by pathologists as low-grade (5–33% steatosis area) and high-grade (>33% steatosis area) are compared based on gene expression microarrays and RT-PCR from liver biopsies. Here these datasets are further investigated via meta-analysis. Furthermore, in our previous study Wruck et al. [13], ELISA assays (proteome) and NMR-based metabolome were analyzed from patient and healthy control serum samples. Fibroblasts from skin biopsies were reprogrammed into iPSCs and differentiated into HLCs [73, 74]. Pluripotency in iPSCs were confirmed by the expression of OCT4 and TRA-1-60. HLCs express ALBUMIN and AFP. After stimulus with for example, oleic acid they incorporate fat as lipid droplets (Graffmann et al. [75]). Scale bars: 100 µm/50 µm/20 µm. Abbreviations: HLC, hepatocyte like cell; iPSCs, induced pluripotent stem cells; NASH, non-alcoholic steatohepatitis; NMR, Nuclear Magnetic Resonance.

regulating drugs and bariatric surgery [41]. As weight loss is difficult to achieve and maintain, numerous clinical trials employing distinct drugs have been conducted [42]. Most extensively evaluated drugs are glitazones (TZDs) and antioxidants to improve insulin sensitivity and reduce oxidative stress [42]. Past trials demonstrated that TZD medication leads to significant weight gain as well as higher risks of myocardial infarction and bladder cancer, and therefore their impact is still controversial [42]. The PIVENS (https://clinicaltrials.gov/ct2/show/NCT00063622) trial showed a comparable effect of vitamin E and pioglitazone [43]. However, recent studies have not been able to detect superior efficacy compared to placebo [44]. Many studies also involve combined treatment with the antioxidants vitamin E and vitamin C. Although improved insulin sensitivity, transaminases and liver histology were detected after supplementation [45, 46], recent studies linked the risk of developing prostate cancer and increased mortality to daily administration of vitamin E [47, 48]. Betaine is a nutrient, which leads to an up-regulation of PPARα and LXRα in the liver, resulting in reduced lipid accumulation and improved insulin resistance [49]. Furthermore, fibrates (peroxisome proliferator-activated receptor (PPAR) α ligands) are commonly used for treatment of hypertriglyceridemia, suggesting a beneficial effect for NAFLD patients by reducing serum lipid levels [50]. Nevertheless, there are just few studies targeting PPARα in human at the present time. While long-term treatments with fibrates lead to hepatic cancer in rodent models, Nseir et al. [51] concluded that they are safe to use in NAFLD patients. Another strategy might be the treatment with Omega-3 FAs, which are already approved for hypertriglyceridemia. Therefore, a dose of 2–4 g/day of eicosapentaenoic acid (EPA) plus docohexaenoic acid (DPA) is recommended. In the WELCOME study combined treatment of EPA and DPA showed a trend toward the improvement in liver fat percentage in NAFLD patients. However, no improvement in fibrosis scores was detected [52]. Pentoxifyllin and several ACE-inhibitors have

A



C

Term	PValue	Benjamini
GO:0008610~lipid biosynthetic process	7.38E-07	2.01E-04
GO:0016126~sterol biosynthetic process	7.07E-06	9.61E-04
GO:0008203~cholesterol metabolic process	1.31E-04	0.0088378
GO:0006695~cholesterol biosynthetic process	3.67E-04	0.01651487
Fatty acid biosynthesis	9.98E-04	0.04821588
GO:0008202~steroid metabolic process	0.001307302	0.04956097
hsa03030:DNA replication	0.002575117	0.06242728
oxidoreductase	0.002904675	0.0915303
PIRSF015921:fatty acid desaturase/sphingolipid desaturase	0.003646674	0.03938995
SM00350:MCM	0.004947764	0.0389034

B

Category	Term	PValue	Bonferroni	source
SP_PIR_KEYWORDS	acetylation	1.04E-05	0.0037853	Ahrens
SP_PIR_KEYWORDS	isopeptide bond	1.37E-04	0.0483937	Ahrens
SP_PIR_KEYWORDS	alternative splicing	5.31E-16	3.25E-13	du Plessis
UP_SEQ_FEATURE	splice variant	7.28E-16	2.50E-12	du Plessis
SP_PIR_KEYWORDS	nucleus	2.06E-12	1.20E-09	du Plessis
UP_SEQ_FEATURE	mutagenesis site	8.38E-09	2.70E-05	du Plessis
SP_PIR_KEYWORDS	phosphoprotein	1.02E-07	5.98E-05	du Plessis
GOTERM_BP_FAT	GO:0006974~response to DNA damage stimulus	1.59E-07	5.05E-04	du Plessis
GOTERM_CC_FAT	GO:0031981~nuclear lumen	7.00E-07	3.63E-04	du Plessis
SP_PIR_KEYWORDS	ubl conjugation pathway	1.84E-06	0.00107339	du Plessis
SP_PIR_KEYWORDS	cytoplasm	2.11E-06	0.00123392	du Plessis
GOTERM_BP_FAT	GO:0042770~DNA damage response, signal transduction	3.36E-06	0.01058736	du Plessis
GOTERM_CC_FAT	GO:0070013~intracellular organelle lumen	1.01E-05	0.00522297	du Plessis
GOTERM_BP_FAT	GO:0033554~cellular response to stress	1.49E-05	0.04600347	du Plessis
GOTERM_CC_FAT	GO:0005654~nucleoplasm	1.49E-05	0.00771834	du Plessis
GOTERM_CC_FAT	GO:0043233~organelle lumen	1.64E-05	0.00849708	du Plessis
SP_PIR_KEYWORDS	rna-binding	2.47E-05	0.01433257	du Plessis
GOTERM_CC_FAT	GO:0031974~membrane-enclosed lumen	3.39E-05	0.01746092	du Plessis
GOTERM_MF_FAT	GO:0003723~RNA binding	4.59E-05	0.0454494	du Plessis
SP_PIR_KEYWORDS	endoplasmic reticulum	9.97E-11	5.32E-08	Wruck
GOTERM_CC_FAT	GO:0005783~endoplasmic reticulum	6.99E-09	2.93E-06	Wruck
SP_PIR_KEYWORDS	acetylation	1.13E-07	6.03E-05	Wruck
SP_PIR_KEYWORDS	er-golgi transport	9.54E-05	0.04968409	Wruck
SP_PIR_KEYWORDS	phosphoprotein	8.98E-09	6.75E-06	Horvath
SP_PIR_KEYWORDS	ribonucleoprotein	7.00E-08	5.26E-05	Horvath
SP_PIR_KEYWORDS	ribosomal protein	7.24E-07	5.45E-04	Horvath
GOTERM_CC_FAT	GO:0033279~ribosomal subunit	2.74E-06	0.00180809	Horvath
GOTERM_CC_FAT	GO:0005840~ribosome	4.61E-06	0.00303852	Horvath
GOTERM_MF_FAT	GO:0003735~structural constituent of ribosome	2.32E-05	0.03469443	Horvath

D

symbol	horvath_cor.p	wruck_p	duplessis_cor.p	ahrens_cor.p
LRRC31	5.73E-08	0.00023128	0.009587548	0.000135264
HPRT1	5.04E-09	0.00070555	0.004159158	0.000152029
SQLE	1.61E-05	0.03513908	0.001795979	0.000520544
FADS2	0.010776373	0.03512344	0.002576204	0.00129214
WIPI1	2.42E-05	0.00410375	0.031193939	0.001349224
FADS1	0.007415506	0.0086579	3.56E-06	0.001491239
PPIF	0.0020393	0.00086748	0.031193939	0.001828382
NIPAL1	6.67E-06	0.0018436	0.027848189	0.002292821
MET	6.88E-08	0.00158345	0.027212972	0.002993365
SHBG	1.58E-09	3.15E-07	0.031897874	0.004709214
MCM6	7.86E-05	0.04592997	0.017102331	0.008028648
IDI1	0.040687722	0.03985392	0.001516814	0.008456354
CYP51A1	0.028853025	0.04868852	0.000624684	0.009804184
FASN	0.002893765	0.01806343	0.02897705	0.011713233
NPC1L1	0.010510651	0.03017396	0.001962763	0.016585654
MCM2	6.00E-05	0.01032887	0.031712961	0.016672329
DISP1	0.005677848	3.76E-05	0.019423098	0.017022881
TMEM14A	0.000822925	1.97E-05	0.048633384	0.01965584
C3orf58	0.00758831	0.00664822	0.024199888	0.021096094
CD48	0.049586409	0.02176865	0.030501762	0.034342057
POLE4	0.008128169	0.03607999	0.009036368	0.038337584
FAM171A1	7.56E-07	0.00182706	0.027212972	0.045238207

Figure 2. A gene signature associated with nonalcoholic fatty liver disease (NAFLD) progression highlights involvement of lipid and cholesterol metabolic processes and oxido-reductase/oxidative stress in the development of non-alcoholic steatohepatitis (NASH). (A): The venn diagram shows the overlap between genes significantly ($p < .05$) correlated with NAFLD progression to NASH from four datasets from du Plessis et al. [35], Wruck et al. [13], Ahrens et al. [12] and Horvath et al. [11]. The colored frames correspond to frames in (B–D). (B): The functional annotations from the DAVID web tool associated with the genes expressed exclusively in the early-stage dataset from Wruck et al. [13], shows involvement of *endoplasmatic reticulum (ER* and *ER-Golgi transport* and furthermore *acetylation* as overlap with the NASH dataset from Ahrens et al. [12] while du Plessis et al. [35] and Horvath et al. [11] overlap in the term *phosphoprotein*. (C): The functional annotations from the DAVID web tool associated with the common gene signature highlights lipid and cholesterol synthesis and metabolism as major processes involved in disease progression. Furthermore, the term *oxido-reductase* points at oxidative stress as the basis for inflammation. (D): The common steatosis- signature consists of 22 genes which include HPRT1, central to purine metabolism, SQLE, the rate-limiting enzyme in sterol biosynthesis, CYP51A1, central to cholesterol biosynthesis [37], FASN, catalyzing the synthesis of palmitate and long-chain saturated fatty acids, and FADS1 and FADS2, regulating unsaturation of fatty acids. The color scale marks lower *p*-values in green and higher *p*-values in red.

been suggested to prevent fibrosis and improve NASH due to the inhibition of proinflammatory cytokines, however more trials are needed as their effect was mainly shown in animal studies [53]. To date several promising drugs such as metformin, with beneficial effects in rodent models have generated controversial results in human trials [44]. Recently, a promising clinical trial of the farnesoid X nuclear receptor (FXR) activator obeticholic acid showed improvement in the histological features of NASH [54]. However, its safety needs further clarification. The results of our meta-analysis and literature on the role of cholesterol in the progression to NASH [38, 39] are in line with the outcome of this clinical trial on an activator of FXR which is known to accelerate retrograde cholesterol transport by increasing the clearance of HDL [54]. Evidence presented suggest we cannot extrapolate findings in rodent models to human and further trials are needed to identify feasible medications for NAFLD patients and medication should be currently restricted to patients with related comorbidities.

Epigenetics and microRNAs in NAFLD

As in many other diseases not only genetic factors play a role in the development of NAFLD and NASH, but also epigenetic alterations are involved. The term epigenetics describes heritable changes in gene expression patterns that are not encoded directly within the DNA but are rather determined by associated factors such as DNA methylation or histone modifications [55]. Although heritable, epigenetic modifications can be altered during a cell's lifespan in response to environmental conditions [55]. As the liver is the main metabolic organ it is exposed to nutrition derived factors that can influence its epigenetic signature [56]. With acetyl-CoA—the input to fatty-acid synthesis—and S-adenosylmethionine, two main factors that are required for epigenetic modifications of histones and DNA are directly linked to glucose or methionine metabolism, respectively [56]. It has been demonstrated that histone acetylation contributes to NAFLD via activation of lipogenic and glycolytic genes, while deviations from the ideal

level of S-adenosylmethionine are linked to lipogenesis, accumulation of hepatic triglycerides and NAFLD [56]. However, epigenetic factors are only suitable as biomarkers when they can be determined in easily accessible lymphocytes. In this regard, a study with obese adolescents made promising progress showing a high correlation between body mass index loss and DNA methylation changes in selected genes such as *AQP9, DUSP22, HIPK3, TNNT1*, and *TNNI3* [57].

Another easily accessible epigenetic parameter are microRNAs (miRNAs) which can be measured directly in blood samples. miRNAs are short RNA molecules of about 22 nucleotides that regulate mRNA stability and thus transcription levels [58]. Several studies accessed their role in steatosis with somewhat diverse results, probably depending on differences of diagnostic methods, staging and miRNA measurement [59]. As miRNA levels can be directly analysed in blood, they are also suitable as biomarkers for the distinct stages of steatosis. For example, the expression miRNA-122, which accounts for 70% of the total liver miRNAs is highly upregulated in the blood of NAFLD patients [60, 61]. Additionally, blood levels of miRNA-34a and 16 are also upregulated in NAFLD patients compared to healthy controls. The expression levels of miRNA-122 and 34a could even serve to distinguish NAFLD and NASH [60]. Another study demonstrated that in addition to miRNA-122, expression levels of miRNA-192 and 375 correlate with the severity of the disease [61]. However, as the relationship between miRNAs and regulated genes is highly dynamic and multifactorial, careful additional studies are needed before they can be reliably used as diagnostic biomarkers.

Role of the Gut Microbiome in NAFLD

The gut microbiome is highly sensitive toward nutritional changes and adapts in mice and humans after switches from healthy to unhealthy diets and vice versa within a day [62, 63]. Dysbiosis, meaning a disturbed balance within the bacterial populations, plays a role in the development of NAFLD. The ratio of *Firmicutes* to *Bacteroidetes*, the two main gut-residing bacterial phyla, increases with obesity [64] and there is an inverse correlation between *Bacteroidetes* percentage and steatohepatitis [65]. In addition, an overall reduction of bacterial diversity has been associated with obesity [66]. Interestingly, transplanting the gut microbiome of mice with NAFLD into wildtype mice resulted in an increase in the disease in the latter animals, demonstrating that NAFLD is transplantable [67]. Dysbiosis is also responsible for metabolic endotoxemia where increased circulating Lipopolysaccharids (LPS) trigger inflammatory reactions also in the liver [68]. In addition, the gut barrier gets leaky because tight-junctions between gut cells get disrupted under high fat diets [69]. This leakiness allows LPS to enter the blood system and to contribute to the systemic low grade inflammation typically associated with NAFLD [69]. Besides these phenomena it has also been observed that a couple of other metabolism related pathways, which are influenced by the gut microbiome, are changed in NAFLD as for example choline, fatty and bile acid metabolism or the production of endogenous ethanol for a detailed review on these processes please refer to Kirpich et al. [65] and Houghton et al. [70].

The Prospect of *Disease-in-a-Dish* Models with Induced Pluripotent Stem Cells (iPSCs)

In vitro models of steatosis have been proposed by Gómez-Lechón et al. [71] and a sophisticated co-culture model of hepatocytes and HSCs for NASH has been described [9]. As we showed recently, iPSC derived hepatocyte-like cells are a reliable cellular tool for modeling NAFLD (Graffmann et al. [75]). iPSCs have the advantage that they can be derived from distinct genetic backgrounds and the parental somatic cells are accessible with minimal invasive methods or can be even taken from blood and urine [72]. The HLCs derived from these iPSCs can then be cultured mimicking different nutritional conditions in order to monitor donor-specific effects on metabolism and LD incorporation. In contrast to primary hepatocytes, the amount of HLCs is not limited, which enables larger and more detailed studies. In addition, HLCs can be cultured longer than primary hepatocytes before they dedifferentiate, which increases the time period for analyses (unpublished observation).

With HLCs, the impact of a given genetic background on hepatocyte TG storage can be analysed under standardized conditions. In combination with CRISPR/cas9 it is even possible to introduce or correct single mutations and thus directly assess their role during disease development. Importantly, in vitro models are also suitable for drug testing and for experimentally verifying the action and function of biomarkers [Wruck et al. Strategies for identifying predictive biomarkers of non-alcoholic fatty liver disease. Drug target Review 2016, 2. (17 March 2016)].

Conclusion

The importance of finding practicable diagnostic tests and therapies for the distinct stages of NAFLD is underscored by the fact it is becoming an economic burden and also an epidemic of the 21st century. While there are appropriate tests for fibrosis which may be even improved by integration of novel research findings diagnostic tests for NASH suffer from a lack of sensitivity and specificity. Although it has been reported that the long-term outcome of the disease depends only on the fibrosis stage, early diagnosis of NASH or even steatosis would enable intervention at early disease phases which are reversible. Our meta-analysis unveiled a 22-gene-signature which mainly represents cholesterol and lipid metabolic pathways thus confirming publications pinpointing cholesterol as major player in disease progression. Recent promising results of a clinical trial based on obeticholic acid—a FXR agonist accelerating cholesterol clearance—corroborates our findings and highlight that cholesterol transport and metabolism are potential targets for diagnostic markers and therapies. We propose the use of disease-in-a-dish-models based on hepatocytes derived from patient-specific iPSCs which takes into account individual differences in this complex metabolic disease. This approach could be combined with iPSC-derived HSCs from the same patient in order to model the important interactions between these two cell types which are responsible for progression to NASH and fibrosis.

Acknowledgments

James Adjaye acknowledges support from the Medical Faculty of the Heinrich Heine University Düsseldorf, Germany.

Author Contributions

W.W.: Manuscript writing, data analysis and interpretation; N.G.: Manuscript writing; M.K.: Manuscript writing; J.A.: Conception and design, manuscript writing, final approval of manuscript.

Disclosure of Potential Conflicts of Interest

The authors declare no conflict of interest.

References

1 Baffy G, Brunt EM, Caldwell SH. Hepatocellular carcinoma in non-alcoholic fatty liver disease: An emerging menace. J Hepatol 2012;56:1384–1391.

2 Lee SS. Radiologic evaluation of nonalcoholic fatty liver disease. World J Gastroenterol 2014;20:7392.

3 Speliotes EK, Yerges-Armstrong LM, Wu J et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLOS Genet 2011;7:e1001324.

4 Kozlitina J, Smagris E, Stender S et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2014;46:352–356.

5 Montagner A, Polizzi A, Fouché E et al. Liver PPARα is crucial for whole-body fatty acid homeostasis and is protective against NAFLD. Gut 2016;65:1202–1214. doi: 10.1136/gutjnl-2015-310798.

6 Ratziu V, Harrison SA, Francque S et al. Elafibranor, an agonist of the peroxisome proliferator−activated receptor−α and −δ, induces resolution of nonalcoholic steatohepatitis without fibrosis worsening. Gastroenterology 2016;150:1147–1159.e5.

7 Lopez-Vicario C, Gonzalez-Periz A, Rius B et al. Molecular interplay between 5/6 desaturases and long-chain fatty acids in the pathogenesis of non-alcoholic steatohepatitis. Gut 2014;63:344–355. doi: 10.1136/gutjnl-2012-303179.

8 Arendt BM, Comelli EM, Ma DWL et al. Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids: Steatohepatitis/metabolic liver disease. Hepatology 2015;61:1565–1578.

9 Barbero-Becerra VJ, Giraudi PJ, Chávez-Tapia NC et al. The interplay between hepatic stellate cells and hepatocytes in an in vitro model of NASH. Toxicol In Vitro 2015;29:1753–1758.

10 Jozefczuk J, Kashofer K, Ummanni R et al. A systems biology approach to deciphering the etiology of steatosis employing patient-derived dermal fibroblasts and iPS cells. Front Physiol 2012;3:339.

11 Horvath S, Erhart W, Brosch M et al. Obesity accelerates epigenetic aging of human liver. Proc Natl Acad Sci USA 2014;111:15538–15543.

12 Ahrens M, Ammerpohl O, Schönfels WV et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. Cell Metab 2013;18:296–302.

13 Wruck W, Kashofer K, Rehman S et al. Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes. Sci Data 2015;2:150068.

14 Romeo S, Kozlitina J, Xing C et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2008;40:1461–1465.

15 Anstee QM, Day CP. The genetics of nonalcoholic fatty liver disease: Spotlight on PNPLA3 and TM6SF2. Semin Liver Dis 2015;35:270–290.

16 Palmer ND, Musani SK, Yerges-Armstrong LM et al. Characterization of European ancestry nonalcoholic fatty liver disease-associated variants in individuals of African and Hispanic descent. Hepatol Baltim MD 2013;58:966–975.

17 Pirola CJ, Sookoian S. The dual and opposite role of the *TM6SF2* -rs58542926 variant in protecting against cardiovascular disease and conferring risk for nonalcoholic fatty liver: A meta-analysis. Hepatology 2015;62:1742–1756.

18 Pirazzi C, Valenti L, Motta BM et al. PNPLA3 has retinyl-palmitate lipase activity in human hepatic stellate cells. Hum Mol Genet 2014;23:4077–4085.

19 Eichmann TO, Grumet L, Taschler U et al. ATGL and CGI-58 are lipid droplet proteins of the hepatic stellate cell line HSC-T6. J Lipid Res 2015;56:1972–1984.

20 Kahali B, Liu Y-L, Daly AK et al. TM6SF2: Catch-22 in the fight against nonalcoholic fatty liver disease and cardiovascular disease? Gastroenterology 2015;148:679–684.

21 Browning JD, Horton JD. Molecular mediators of hepatic steatosis and liver injury. J Clin Invest 2004;114:147–152.

22 Wahren J, Sato Y, Ostman J et al. Turnover and splanchnic metabolism of free fatty acids and ketones in insulin-dependent diabetics at rest and in response to exercise. J Clin Invest 1984;73:1367–1376.

23 Donnelly KL, Smith CI, Schwarzenberg SJ et al. Sources of fatty acids stored in liver and secreted via lipoproteins in patients with nonalcoholic fatty liver disease. J Clin Invest 2005;115:1343–1351.

24 Lambert JE, Ramos–Roman MA, Browning JD et al. Increased de novo lipogenesis is a distinct characteristic of individuals with nonalcoholic fatty liver Disease. Gastroenterology 2014;146:726–735.

25 Yamaguchi K, Yang L, McCall S et al. Inhibiting triglyceride synthesis improves hepatic steatosis but exacerbates liver damage and fibrosis in obese mice with nonalcoholic steatohepatitis. Hepatol Baltim MD 2007;45:1366–1374.

26 Hernández-Gea V, Hilscher M, Rozenfeld R et al. Endoplasmic reticulum stress induces fibrogenic activity in hepatic stellate cells through autophagy. J Hepatol 2013;59:98–104.

27 Gressner AM, Weiskirchen R. Modern pathogenetic concepts of liver fibrosis suggest stellate cells and TGF-beta as major players and therapeutic targets. J Cell Mol Med 2006;10:76–99.

28 González-Rodríguez Á, Mayoral R, Agra N et al. Impaired autophagic flux is associated with increased endoplasmic reticulum stress during the development of NAFLD. Cell Death Dis 2014;5:e1179.

29 Carpino G, Renzi A, Onori P et al. Role of hepatic progenitor cells in nonalcoholic fatty liver disease development: Cellular cross-talks and molecular networks. Int J Mol Sci 2013;14:20112–20130.

30 Feldstein AE, Wieckowska A, Lopez AR et al. Cytokeratin-18 fragment levels as noninvasive biomarkers for nonalcoholic steatohepatitis: A multicenter validation study. Hepatology 2009;50:1072–1078.

31 Ratziu V, Massard J, Charlotte F et al. Diagnostic value of biochemical markers (FibroTest-FibroSURE) for the prediction of liver fibrosis in patients with non-alcoholic fatty liver disease. BMC Gastroenterol 2006;6:6.

32 Angulo P, Hui JM, Marchesini G et al. The NAFLD fibrosis score: A noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology 2007;45:846–854.

33 Moylan CA, Pang H, Dellinger A et al. Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. Hepatol Baltim MD 2014;59:471–482.

34 Angulo P, Kleiner DE, Dam-Larsen S et al. Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. Gastroenterology 2015;149:389–397.e10.

35 Plessis Jd, Pelt Jv, Korf H, et al. Association of adipose tissue inflammation with histologic severity of nonalcoholic fatty liver disease. Gastroenterology 2015;149:635–648.e14.

36 Ikonen E. Cellular cholesterol trafficking and compartmentalization. Nat Rev Mol Cell Biol 2008;9:125–138.

37 Lorbek G, Perše M, Jeruc J et al. Lessons from hepatocyte-specific Cyp51 knockout mice: Impaired cholesterol synthesis leads to oval cell-driven liver injury. Sci Rep 2015;5:8777.

38 Min H-K, Kapoor A, Fuchs M et al. Increased hepatic synthesis and dysregulation of cholesterol metabolism is associated with the severity of nonalcoholic fatty liver disease. Cell Metab 2012;15:665–674.

39 Musso G, Gambino R Cassader M. Cholesterol metabolism and the pathogenesis of non-alcoholic steatohepatitis. Prog Lipid Res 2013;52:175–191.

40 Gorden DL, Myers DS, Ivanova PT et al. Biomarkers of NAFLD progression: A

lipidomics approach to an epidemic. J Lipid Res 2015;56:722–736.
41 Peng L, Wang J, Li F. Weight reduction for non-alcoholic fatty liver disease. Cochrane Database Syst Rev 2011;6:CD003619.
42 Polyzos SA, Kountouras J, Zavos C et al. Nonalcoholic fatty liver disease: Multimodal treatment options for a pathogenetically multiple-hit disease. J Clin Gastroenterol 2012;46:272–284.
43 Chalasani NP, Sanyal AJ, Kowdley KV et al. Pioglitazone versus vitamin E versus placebo for the treatment of non-diabetic patients with non-alcoholic steatohepatitis: PIVENS trial design. Contemp Clin Trials 2009;30:88–96.
44 Schattenberg JM, Schuppan D. Nonalcoholic steatohepatitis: The therapeutic challenge of a global epidemic. Curr Opin Lipidol 2011;22:479–488.
45 Nobili V, Manco M, Devito R et al. Effect of vitamin E on aminotransferase levels and insulin resistance in children with non-alcoholic fatty liver disease. Aliment Pharmacol Ther 2006;24:1553–1561.
46 Williams DB, Wan Z, Frier BC et al. Dietary supplementation with vitamin E and C attenuates dexamethasone-induced glucose intolerance in rats. Am J Physiol Regul Integr Comp Physiol 2012;302:R49–R58.
47 Klein EA, Thompson IM, Tangen CM et al. Vitamin E and the risk of prostate cancer: The selenium and vitamin E cancer prevention trial (SELECT). JAMA 2011;306:1549.
48 Bjelakovic G, Nikolova D Gluud C. Antioxidant supplements and mortality. Curr Opin Clin Nutr Metab Care 2013;1.
49 Ge C-X, Yu R, Xu M-X et al. Betaine prevented fructose-induced NAFLD by regulating LXRα/PPARα pathway and alleviating ER stress in rats. Eur J Pharmacol 2016;770:154–164.
50 Tailleux A, Wouters K, Staels B. Roles of PPARs in NAFLD: Potential therapeutic targets. Biochim Biophys Acta 2012;1821:809–818.
51 Nseir W, Mograbi J, Ghali M. Lipid-lowering agents in nonalcoholic fatty liver disease and steatohepatitis: Human studies. Dig Dis Sci 2012;57:1773–1781.
52 Scorletti E, Bhatia L, McCormick KG et al. Effects of purified eicosapentaenoic and docosahexaenoic acids in nonalcoholic fatty liver disease: Results from the WELCOME* study. Hepatology 2014;60:1211–1221.
53 Hossain N, Kanwar P, Mohanty SR. A comprehensive updated review of pharmaceutical and nonpharmaceutical treatment for NAFLD. Gastroenterol Res Pract 2016;2016:1–17.
54 Neuschwander-Tetri BA, Loomba R, Sanyal AJ et al. Farnesoid X nuclear receptor ligand obeticholic acid for non-cirrhotic, non-alcoholic steatohepatitis (FLINT): A multicentre, randomised, placebo-controlled trial. Lancet 2015;385:956–965.
55 Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. Cell 2007;128:669–681.
56 Lee JH, Friso S, Choi S-W. Epigenetic mechanisms underlying the link between non-alcoholic fatty liver diseases and nutrition. Nutrients 2014;6:3303–3325.
57 Moleres A, Campión J, Milagro FI et al. Differential DNA methylation patterns between high and low responders to a weight loss intervention in overweight or obese adolescents: The EVASYON study. FASEB J 2013;27:2504–2512.
58 Bartel DP. MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell 2004;116:281–297.
59 Baffy G. MicroRNAs in nonalcoholic fatty liver disease. J Clin Med 2015;4:1977–1988.
60 Cermelli S, Ruggieri A, Marrero JA et al. Circulating microRNAs in patients with chronic hepatitis C and non-alcoholic fatty liver disease. PLoS One 2011;6:e23937.
61 Pirola CJ, Fernández Gianotti T, Castaño GO et al. Circulating microRNA signature in non-alcoholic fatty liver disease: From serum non-coding RNAs to liver histology and disease pathogenesis. Gut 2015;64:800–812.
62 Turnbaugh PJ, Ridaura VK, Faith JJ et al. The effect of diet on the human gut microbiome: A metagenomic analysis in humanized gnotobiotic mice. Sci Transl Med 2009;1:6ra14–6ra14.
63 Wu GD, Chen J, Hoffmann C et al. Linking long-term dietary patterns with gut microbial enterotypes. Science 2011;334:105–108.
64 Ley RE, Turnbaugh PJ, Klein S et al. Microbial ecology: Human gut microbes associated with obesity. Nature 2006;444:1022–1023.
65 Kirpich IA, Marsano LS, McClain CJ. Gut–liver axis, nutrition, and non-alcoholic fatty liver disease. Clin Biochem 2015;48:923–930.
66 Le Chatelier E, Nielsen T, Qin J et al. Richness of human gut microbiome correlates with metabolic markers. Nature 2013;500:541–546.
67 Le Roy T, Llopis M, Lepage P et al. Intestinal microbiota determines development of non-alcoholic fatty liver disease in mice. Gut 2013;62:1787–1794.
68 Cani PD, Amar J, Iglesias MA et al. Metabolic endotoxemia initiates obesity and insulin resistance. Diabetes 2007;56:1761–1772.
69 Brun P, Castagliuolo I, Leo VD et al. Increased intestinal permeability in obese mice: New evidence in the pathogenesis of nonalcoholic steatohepatitis. Am J Physiol Gastrointest Liver Physiol 2006;292:G518–G525.
70 Houghton D, Stewart C, Day C et al. Gut microbiota and lifestyle interventions in NAFLD. Int J Mol Sci 2016;17:447.
71 Gómez-Lechón MJ, Donato MT, Martínez-Romero A et al. A human hepatocellular in vitro model to investigate steatosis. Chem Biol Interact 2007;165:106–116.
72 Raab S, Klingenstein M, Liebau S et al. A comparative view on human somatic cell sources for iPSC generation. Stem Cells Int 2014;2014:1–12.
73 Jozefczuk J, Prigione A, Chavez L et al. Comparative analysis of human embryonic stem cell and induced pluripotent stem cell-derived hepatocyte-like cells reveals current drawbacks and possible strategies for improved differentiation. STEM CELLS DEV 2011;20:1259–1275.
74 Graffmann N, Matz P, Wruck W et al. The promise of induced pluripotent stem cells for liver disease, toxicology and drug discovery. DRUG TARGET REV 2015;2:8–11.
75 Graffmann N, Ring S, Kawala M-A et al. Modelling NAFLD with human pluripotent stem cell derived immature hepatocyte like cells reveals activation of PLIN2 and confirms regulatory functions of PPARα. STEM CELLS DEV. 2016 Jun 16. [Epub ahead of print]


See www.StemCells.com for supporting information available online.

2.1.4 Meta-analysis reveals up-regulation of cholesterol processes in non-alcoholic and down-regulation in alcoholic fatty liver disease.

World J Hepatol. 2017 Mar 18;9(8):443-454. doi: 10.4254/wjh.v9.i8.443.

Wruck W(1), Adjaye J(1).

Author information: (1)Wasco Wruck, James Adjaye, Medical Faculty, Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, 40225 Düsseldorf, Germany.

AIM: To compare transcriptomes of non-alcoholic fatty liver disease (NAFLD) and alcoholic liver disease (ALD) in a meta-analysis of liver biopsies. METHODS: Employing transcriptome data from patient liver biopsies retrieved from several public repositories we performed a meta-analysis comparing ALD and NAFLD. RESULTS: We observed predominating commonalities at the transcriptome level between ALD and NAFLD, most prominently numerous down-regulated metabolic pathways and cytochrome-related pathways and a few up-regulated pathways which include ECM-receptor interaction, phagosome and lysosome. However some pathways were regulated in opposite directions in ALD and NAFLD, for example, glycolysis was down-regulated in ALD and up-regulated in NAFLD. Interestingly, we found rate-limiting genes such as HMGCR, SQLE and CYP7A1 which are associated with cholesterol processes adversely regulated between ALD (down-regulated) and NAFLD (up-regulated). We propose that similar phenotypes in both diseases may be due to a lower level of the enzyme CYP7A1 compared to the cholesterol synthesis enzymes HMGCR and SQLE. Additionally, we provide a compendium of comparative KEGG pathways regulation in ALD and NAFLD. CONCLUSION: Our finding of adversely regulated cholesterol processes in ALD and NAFLD draws the focus to regulation of cholesterol secretion into bile. Thus, it will be interesting to further investigate CYP7A1-mediated cholesterol secretion into bile - also as possible drug targets. The list of potential novel biomarkers may assist differential diagnosis of ALD and NAFLD.

DOI: 10.4254/wjh.v9.i8.443 PMCID: PMC5355767 PMID: 28357032

Conflict of interest statement: Conflict-of-interest statement: The authors have no conflict of interest.

URL: https://dx.doi.org/10.4254/wjh.v9.i8.443

Contribution: Wasco Wruck performed transcriptomics analyses and the meta-analysis and wrote the manuscript.

World Journal of Hepatology

Submit a Manuscript: http://www.wjgnet.com/esps/ *World J Hepatol* 2017 March 18; 9(8): 443-454

DOI: 10.4254/wjh.v9.i8.443 ISSN 1948-5182 (online)

META-ANALYSIS

Meta-analysis reveals up-regulation of cholesterol processes in non-alcoholic and down-regulation in alcoholic fatty liver disease

Wasco Wruck, James Adjaye

Wasco Wruck, James Adjaye, Medical Faculty, Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, 40225 Düsseldorf, Germany

Author contributions: Wruck W performed transcriptomics analyses and the meta-analysis; Wruck W and Adjaye J wrote the manuscript; Adjaye J initiated and co-ordinated the work.

Supported by The Medical Faculty of the Heinrich Heine University Düsseldorf.

Conflict-of-interest statement: The authors have no conflict of interest.

Data sharing statement: No additional data are available.

Manuscript source: Unsolicited manuscript

Correspondence to: James Adjaye, PhD, Professor, Medical Faculty, Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, Moorenstrasse 5, 40225 Düsseldorf, Germany. james.adjaye@med.uni-duesseldorf.de
Telephone: +49-0211-8108191
Fax: +49-0211-8117858

Received: October 12, 2016
Peer-review started: October 17, 2016
First decision: November 14, 2016
Revised: November 29, 2016
Accepted: December 13, 2016
Article in press: December 14, 2016
Published online: March 18, 2017

Abstract

AIM

To compare transcriptomes of non-alcoholic fatty liver disease (NAFLD) and alcoholic liver disease (ALD) in a meta-analysis of liver biopsies.

METHODS

Employing transcriptome data from patient liver biopsies retrieved from several public repositories we performed a meta-analysis comparing ALD and NAFLD.

RESULTS

We observed predominating commonalities at the transcriptome level between ALD and NAFLD, most prominently numerous down-regulated metabolic pathways and cytochrome-related pathways and a few up-regulated pathways which include ECM-receptor interaction, phagosome and lysosome. However some pathways were regulated in opposite directions in ALD and NAFLD, for example, glycolysis was down-regulated in ALD and up-regulated in NAFLD. Interestingly, we found rate-limiting genes such as *HMGCR*, *SQLE* and *CYP7A1* which are associated with cholesterol processes adversely regulated between ALD (down-regulated) and NAFLD (up-regulated). We propose that similar phenotypes in both diseases may be due to a lower level of the enzyme CYP7A1 compared to the cholesterol synthesis enzymes HMGCR and SQLE. Additionally, we provide a compendium of comparative KEGG pathways regulation in ALD and NAFLD.

CONCLUSION

Our finding of adversely regulated cholesterol processes in ALD and NAFLD draws the focus to regulation of cholesterol secretion into bile. Thus, it will be interesting to further investigate CYP7A1-mediated cholesterol secretion into bile - also as possible drug targets. The list of potential novel biomarkers may assist differential diagnosis of ALD and NAFLD.

Key words: Non-alcoholic fatty liver disease; Alcoholic liver disease cholesterol; Bile; Alcohol dehydrogenase; CYP7A1

Core tip: With a meta-analysis of newly published liver biopsy-derived transcriptome datasets we identified multiple key genes and pathways in common and mutually exclusive in alcoholic liver disease (ALD) and non-alcoholic fatty liver disease (NAFLD). We provide a compendium of comparative regulation for all KEGG pathways in both diseases and propose a list of biomarkers distinguishing both diseases. One surprising finding was that cholesterol metabolism was up-regulated in NAFLD and down-regulated in ALD although leading to the same steatosis phenotype which might be explained by an insufficient conversion rate to bile acids under both conditions.

Wruck W, Adjaye J. Meta-analysis reveals up-regulation of cholesterol processes in non-alcoholic and down-regulation in alcoholic fatty liver disease. *World J Hepatol* 2017; 9(8): 443-454 Available from: URL: http://www.wjgnet.com/1948-5182/full/v9/i8/443.htm DOI: http://dx.doi.org/10.4254/wjh.v9.i8.443

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) and alcoholic liver disease (ALD) have nearly identical symptoms and in the first report non-alcoholic steatohepatitis (NASH) was described as histologically mimicking alcoholic hepatitis[1]. While the cause of ALD is excessive alcohol, the cause of NAFLD is excessive fat resulting from an imbalance between diet and physical activity often associated with insulin resistance and obesity.

We are working on the hypothesis that alcohol is metabolized to fat and beyond this pathway both diseases share a common phenotype. Therefore we place special emphasis on alcohol metabolism which naturally plays a crucial role in ALD. Associations of variants in alcohol and aldehyde dehydrogenases with alcoholism have already been proposed[2]. Most variants protective against alcoholism result in a higher acetaldehyde level either by accelerating alcohol dehydrogenase (most common variants in *ADH1B*) metabolizing alcohol to acetaldehyde or by reducing aldehyde dehydrogenase (most common variants in *ALDH2*) metabolizing acetaldehyde to acetic acid. Acetaldehyde is a carcinogen and causes severe reactions such as flushing, accelerated heart rate and nausea. These severe reactions will impose on most carriers of these variants to abstain from alcohol and thus reduce their risk of becoming alcohol addicts. Furthermore, it has been reported that aldehyde dehydrogenases are down-regulated in alcoholics[3] or animals continually exposed to alcohol had lower ethanol elimination rates[4]. However, this is a matter of debate as no significant down-regulation of aldehyde dehydrogenases was reported by Vidal *et al*[5] but instead a down-regulation in cirrhotic livers independent of alcoholism. Acetic acid - the product of ethanol metabolism, can be further metabolized by acyl-CoA synthetases (ACSS1 and ACSS2) to acetyl-CoA, the substrate for fatty acid synthesis[6]. The expression and activity of Acyl-CoA synthetases in turn are controlled by the sterol regulatory element-binding protein which has been reported to be activated by ethanol[7].

The progression of NAFLD from mild steatosis up to severe NASH or from ALD to alcoholic hepatitis varies widely between individual patients. Oxidative stress and dysregulation of cytokines as a basis for inflammation appear to foster progression to NASH[8] as well as alcoholic hepatitis (AH)[9]. A two-hit progression from simple steatosis to steatohepatitis and fibrosis has been proposed[10], and suggests that after fat accumulation in the liver, lipids are peroxidized by oxidative stress induced by factors such as CYP2E1. The microsomal enzyme CYP2E1 metabolizes ethanol to acetaldehyde under conditions of alcohol dehydrogenase overload and generates oxidative stress as a by-product, however fatty acids also can be a substrate of CYP2E1[9].

Recently the role of the gut has attracted attention. Under alcoholic or high-fat conditions lipopolysaccharides can pass the border of the intestine to the portal vein and circulate to the liver where they trigger inflammation in ALD[11] and in NAFLD[12].

Some studies have already compared ALD and NAFLD[13], *e.g.*, Wilfred de Alwis and Day[14] compared the genetics of both diseases addressing the question why only a small percentage of heavy drinkers and obese people progress from steatosis to severe liver disease. Here, we provide an analytical comparison of transcriptomic and metabolic processes involved in the progression of ALD and NAFLD. Employing transcriptome data derived from patient liver biopsies retrieved from several public repositories we performed a meta-analysis and report a signature of biomarkers distinguishing AH from NASH samples. Furthermore, we found predominating commonalities between both diseases at the level of biological pathways thus implying a large mechanistic similarity between both diseases.

MATERIALS AND METHODS

Transcriptome data analysis

Datasets of microarray gene expression data from liver biopsies were downloaded from the public repositories at NCBI GEO and EBI Array-Express. The compendium consisted of the ALD datasets GSE28619[15] and E-MTAB-2664[16] and the NAFLD datasets GSE61260[17], GSE59045[18], GSE48452[19] and GSE46300[12]. Illumina data was processed *via* R/Bioconductor[20] and packages lumi[21], limma[22] and qvalue[23]. Background-corrected log2-transformed data was normalized *via* quantile

normalization from the lumi package. Affymetrix data was processed *via* R/Bioconductor and packages affy[24], limma, qvalue employing the *rma* normalization method.

Measurements from the multiple platforms were brought together in terms of mean ratios between ALD cases and controls and between NAFLD cases and controls. As controls, healthy liver biopsies or liver biopsies with a low grade of fat accumulation were used. For details we refer to the methods sections of the publications associated with the employed datasets[12,15-19]. Heterogeneity of the datasets was assessed *via* the meta-analysis R package metafor[25] generating forest and funnel plots (supplementary Figure 1A and B). The ratios were transformed to a log2 scale and normalized *via* quantile normalization. The results were again assessed with forest and funnel plots (supplementary Figure 1C and D).

Pathway analysis

In order to disentangle commonalities and differences between ALD and NAFLD, KEGG pathways[26] were analysed with respect to common pathways, up- and down-regulation and discordant up- and down-regulation. The ratios between ALD and control and NAFLD and control were employed to count the numbers of up- and down-regulated genes for each pathway. A pathway was considered up-regulated when it contained more up- than down-regulated genes. Genes with a ratio > t were termed up-regulated and genes with a ratio < 1/t were termed down-regulated. The threshold t was determined at the 95-quantile of the mean ratios between ALD and NAFLD *vs* control and was set accordingly to t = 4/5. Up- and down-regulation of a pathway was determined *via* the ratio of numbers of up-and down-regulated genes and *via* a binomial test assuming an equal probability of $P = 0.5$ for a gene to be up- or down-regulated.

$$n_{up,pw,case} = |\{g|(\frac{x_{g,case}}{x_{g,control}} > t) \wedge (g \in g_{pw})\}|, \text{ case} \in \{\text{ALD, NAFLD}\} \quad (1)$$

$$n_{down,pw,case} = |\{g|(\frac{x_{g,case}}{x_{g,control}} < 1/t) \wedge (g \in g_{pw})\}|, \text{ case} \in \{\text{ALD, NAFLD}\} \quad (2)$$

$$n_{pw,case} = n_{up,pw,case} + n_{down,pw,case} \quad (3)$$

$$r_{pw,case} = \frac{n_{up,pw,case}}{n_{down,pw,case}} \quad (4)$$

Here, $n_{up,pw,case}$ and $n_{down,pw,case}$ are the numbers of up- and down-regulated genes in a pathway *pw*, g_{pw} are the gene symbols associated with a pathway, $x_{g,case}$ is the gene expression value in a case which can be *ALD* or *NAFLD*, $x_{g,control}$ is the gene expression value in the control case, $r_{pw,case}$ is the ratio indicating up-regulation ($r_{pw,case} > 1$) or down-regulation ($r_{pw,case} < 1$) of pathway *pw*. Significance of up- or down-regulation of a pathway is assessed *via* the Binomial test with the Null hypothesis $H_0: p \leqslant p_0$ and the test statistic $B(p_0, n_{pw,case})$. Because of assumed equal distribution of up- and down-regulation the probability for the binomial distribution is set to $p_0 = 0.5$.

Pathway charts of KEGG pathways indicating up- and down-regulation of genes in ALD and NAFLD were generated *via* the R/Bioconductor package pathview[27].

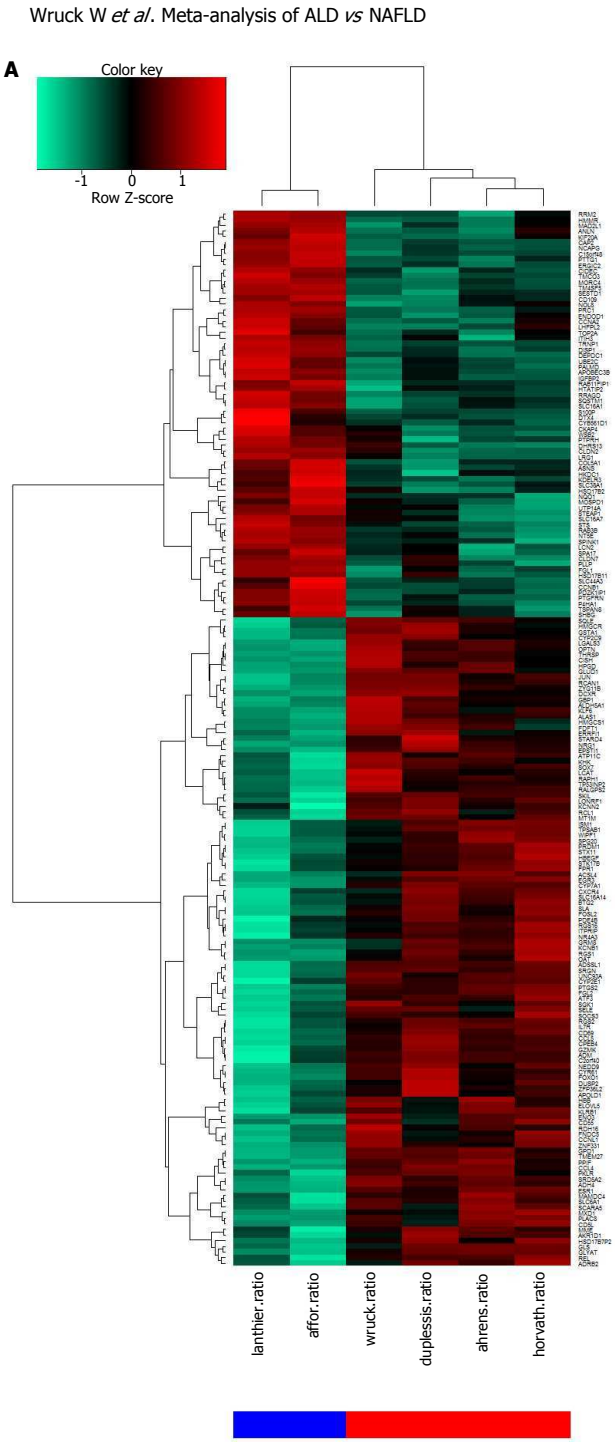
RESULTS

A gene signature distinquishes ALD from NAFLD

The differences between ALD and NAFLD at the transcriptome level could be condensed to a signature of 187 genes which are differentially expressed between both conditions with a *P*-value < 0.01 from the limma test and a ratio > 3/2 or a ratio < 2/3. The heatmap in Figure 1A shows a cluster analysis of this signature of gene expression data from ALD liver biopsies (blue bar) and NAFLD liver biopsies (red bar). The table in Figure 1B shows the 20 most up-regulated and 20 most-down-regulated genes from the signature indicating their log2-ratios and their *P*- and *Q*-values for the comparison ALD *vs* NAFLD. The most up-regulated gene between ALD and NAFLD was SPINK1. SPINK1 is secreted in the pancreatic juice to reversibly inhibit activated trypsin thus preventing pancreatic auto-digestion[28] and variants in this gene have been associated with pancreatitis[29]. Obesity and more prominent alcohol abuse are other causative factors for pancreatitis[28] which by its effects on insulin may contribute to liver disease. Lanthier *et al*[16] revealed the association of SPINK1 with inflammation and proliferation *via* correlation with the inflammatory macrophage marker CD68 and the cell cycle markers Cdk1 and CyclinB1. At the lower part of the table in Figure 1B two RGS (regulator of G-protein signalling) encoding genes, RGS1 and RGS2 are down-regulated in ALD but up-regulated in NAFLD. Nunn *et al*[30] reported reduced fat deposits, decreased serum lipids, and low Leptin levels in RGS2 deficient mice.

Genes regulated in common between ALD and NAFLD

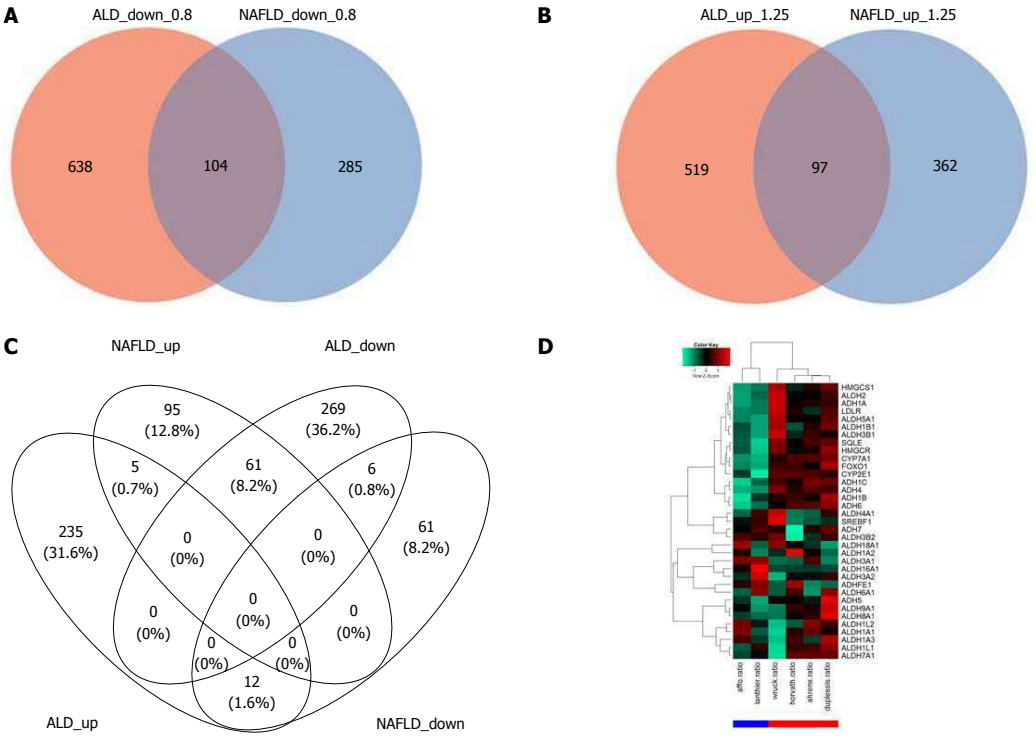
Analysis of the common genes between ALD and NAFLD was subdivided into analysis of down- and up-regulated genes. Figure 2A shows that 104 genes are down-regulated in ALD and NAFLD (ratio < 0.8) while 638 genes are exclusively down-regulated in ALD and 285 in NAFLD. Figure 2B shows that 97 genes are up-regulated in ALD and NAFLD (ratio > 1.25) while 519 genes are exclusively up-regulated in ALD and 362 in NAFLD. There are more distinctly expressed than overlapping genes - in contrary to the KEGG pathways where most pathways overlap (Figure 2E and F). Gene regulation was further restricted with a threshold for the limma test for differential expression of $P < 0.05$. Figure 2C shows a venn diagram of the four resulting sets of up/down-regulated genes in ALD and NAFLD. Here most genes are exclusively regulated but interestingly from the genes regulated in both diseases more genes are oppositely than commonly regulated: 61 genes are up-regulated in NAFLD but down-regulated in ALD and 12



B

Gene	log2ratio_ ALD_NAFLD	limma_p	limma_q
SPINK1	3.1120	0.00086043	0.18404683
IGFBP2	2.0876	0.00021163	0.11364535
S100P	2.0475	0.00076244	0.1814015
LCN2	1.8578	0.00157791	0.20177258
P4HA1	1.6796	6.30E-5	0.09050201
PDZK1IP1	1.5916	8.43E-5	0.09050201
HKDC1	1.4752	0.0099653	0.27847194
NQO1	1.4712	0.00077208	0.1814015
SHBG	1.4032	0.003424	0.22085368
TRNP1	1.3509	1.28E-5	0.09050201
RAB3B	1.3441	0.0001067	0.09050201
RRM2	1.3100	0.0005424	0.15563552
CLDN2	1.2816	0.0023388	0.21111468
STEAP1	1.2667	0.00207474	0.20878205
C15orf48	1.2369	5.77E-5	0.09050201
SLC16A7	1.1967	0.00684988	0.25828645
TMCO3	1.1675	9.41E-5	0.09050201
SQSTM1	1.1231	0.00140252	0.20177258
FGL1	1.0544	0.00104332	0.18745735
CD109	1.0035	0.0081723	0.1814015
SGK1	-1.4274	0.00386065	0.23007365
HBEGF	-1.4634	0.00616426	0.25461827
THRSP	-1.5061	0.00141011	0.20177258
IL7R	-1.5111	0.00126129	0.19647757
ACSL4	-1.5548	0.00302525	0.2146298
ENO3	-1.6119	0.00193989	0.20878205
CD69	-1.6346	0.00050895	0.15352702
JUN	-1.6946	0.00050895	0.11364535
KCNN2	-1.7040	0.0019417	0.27677322
CXCR4	-1.7057	0.00958215	0.27677322
SQLE	-1.8663	0.00937092	0.18404683
PTGS2	-1.8778	6.27E-5	0.09050201
ATF3	-1.9327	0.00026863	0.12722548
SRD5A2	-2.0482	3.49E-5	0.09050201
NR4A3	-2.0615	0.00436311	0.23545747
CYR61	-2.0653	0.00104554	0.18745735
CYP7A1	-2.0965	1.68E-5	0.09050201
RGS2	-2.1605	0.00069501	0.17342654
HMGCS1	-2.1724	0.0057275	0.25288411
RGS1	-2.2757	0.00128147	0.19647757

Figure 1 A gene signature distinquishes alcoholic liver disease from non-alcoholic fatty liver disease. A: The heatmap shows a cluster analysis of logarithmic ratios of gene expression data from ALD liver biopsies *vs* control (blue bar) and NAFLD liver biopsies *vs* control (red bar); B: The table shows the 20 most up-regulated and 20 most-down-regulated genes from the signature indicating their log2-ratios and their *P*- and *Q*-values for the comparison ALD *vs* NAFLD. The full list of these genes can be found in Supplementary Table 2. ALD: Alcoholic liver disease; NAFLD: Non-alcoholic fatty liver disease.



kegg_name	ALD_r_updown	NAFLD_r_updown
Metabolic pathways	0.2890625	0.910714286
Retinol metabolism	0	0.25
Metabolism of xenobiotics by cytochr	0.047619048	0.090909091
Drug metabolism - cytochrome P450	0.090909091	0.272727273
Glycine, serine and threonine metabo	0	0.4
Fatty acid degradation	0	0.75
Jak-STAT signaling pathway	0.0625	0.285714286
Tryptophan metabolism	0.083333333	0.5
Arginine and proline metabolism	0	0.166666667
Steroid hormone biosynthesis	0.090909091	0.833333333
Cysteine and methionine metabolism	0	0.25
Tyrosine metabolism	0.111111111	0
Drug metabolism - other enzymes	0.111111111	0.5
Linoleic acid metabolism	0.125	0.2
Alanine, aspartate and glutamate met	0.125	0.25
Circadian rhythm	0	0

kegg_name	ALD_r_updown	NAFLD_r_updown
ECM-receptor interaction	5.666666667	Inf
Lysosome	12	Inf
Phagosome	4.25	2
Vibrio cholerae infection	Inf	Inf
Focal adhesion	2.875	4.333333333
Regulation of actin cytoskeleton	3.4	2
Gap junction	4	2
Cell cycle	1.75	8
Small cell lung cancer	2	Inf
Cell adhesion molecules (CAMs)	1.714285714	10
Leukocyte transendothelial migra	1.6	7
Pathways in cancer	1.25	2.666666667

Figure 2 Most biological pathways are regulated in the same direction in alcoholic liver disease and non-alcoholic fatty liver disease but a subset of metabolism-associated genes are oppositely regulated. A: Compares ALD and NAFLD in terms of down-regulated genes (ratio < 0.8); B: In terms of up-regulated genes (ratio > 1.25). There are more distinct than overlapping genes - in contrary to the KEGG pathways where most pathways overlap (E and F); C: Interestingly, when regulation is further restricted with a *P*-value < 0.05 more genes are oppositely than commonly regulated - but most are exclusively regulated. Many of the oppositely regulated genes are associated with cholesterol processes, *e.g.*, HMGCR, SQLE and CYP7A1, and are co-expressed with alcohol (ADH) and aldehyd dehydrogeneases (ALDH) as seen in the heatmap (ALD: Blue bar, NAFLD: Red bar) (D). A pathway is considered down-regulated (E) when it contains more down- than up-regulated genes as tested by the binomial test and the ratio, analogously up-regulated pathways are determined (F). The table of common down-regulated pathways includes metabolic, retinol, cytochrome and fatty acid degradation pathways, the up-regulated include ECM-receptor, lysosome and phagosome. ALD: Alcoholic liver disease; NAFLD: Non-alcoholic fatty liver disease.

are up-regulated in ALD and down-regulated in NAFLD while only 5 were commonly up and 6 commonly down-regulated. Supplementary Table 1 shows the corresponding gene sets. The genes up-regulated in

NAFLD but down in ALD refer to major players in cholesterol processes such as *HMGCS1, HMGCR, SQLE, CYP7A1* and *LDLR*. This would confirm the involvement of cholesterol biological processes in the etiology of NAFLD as we previously reported[31] and which distinguish it from the etiology of ALD. The opposite regulation of cholesterol processes as down in ALD and up in NAFLD can also be observed in the corresponding KEGG pathways Steroid biosynthesis, Primary bile acid biosynthesis and Terpenoid backbone biosynthesis (Supplementary file 1, p22, 34 and 84). These findings are in line with reports of a 29% decrease in HMGCR and a 56% decrease in cholesterol 7α-hydroxylase alias CYP7A1 by Lakshmanan *et al*[32], they suggested that increased ethanol leads to a reduced rate of cholesterol degradation to bile acids and accumulation of cholesterol in the liver. We also found (Supplementary Table 2) a stronger down-regulation of *CYP7A1* (log2-ratio = -0.95) than of the upstream cholesterol genes *HMGCR* (log2-ratio = -0.429) and *SQLE* (log2-ratio = -0.33) in ALD while in NAFLD, *CYP7A1* (log2-ratio = 1.15) was weaker up-regulated than *HMGCR* (log2-ratio = 1.57) and *SQLE* (log2-ratio = 1.53). Thus although oppositely regulated in ALD and NAFLD in both diseases more cholesterol is produced than can be secreted by the bile *via* CYP7A1.

Amongst the genes up-regulated in ALD but down in NAFLD are *TNFSF14* in line with the major role of TNF-alpha in ALD[11] and *SPINK1* which was described above in "a gene signature distinguishes ALD from NAFLD".

To further investigate the mechanisms by which ethanol induces these changes in cholesterol processes we analysed expression clusters of genes involved in ethanol and cholesterol related processes. The analysis revealed a cluster of genes down-regulated in ALD and up-regulated in NAFLD including among others the genes encoding for ALDH2, ADH1A, LDLR, SQLE, HMGCR, CYP7A1, CYP2E1 and FOXO1 (Figure 2D). FOXO Transcription factors such as FOXO1, whose expression has been reported to be altered by ethanol[33] and may play a role in the regulation of several genes from this cluster. Interestingly, the heatmap (Figure 2D) shows a much higher degree of co-regulation of *FOXO1* with the rate-limiting cholesterol synthesis enzymes *HMGCR* and *SQLE* than of *SREBF1* which is known as the main regulator of cholesterol[34].

The five genes up-regulated in common between ALD and NAFLD include two collagen encoding genes - *COL1A1* and *COL3A1*, thus demonstrating overlapping disease pathology in the development of fibrotic tissue. The six down-regulated genes in ALD and NAFLD include *HPRT1* which has been reported to be down-regulated in severe liver disease[35].

Pathway analysis

Most biological pathways are regulated in the same direction in ALD and NAFLD. A pathway is considered down-regulated (Figure 2E) when it contains more down- than up-regulated genes as tested by the binomial test and the ratio is less than 1. Up-regulated pathways are determined accordingly (Figure 2F). The table of common down-regulated pathways includes metabolic, retinol, cytochrome and fatty acid degradation pathways, the up-regulated pathways include ECM-receptor, lysosome and phagosome.

Common pathways down-regulated in ALD and NAFLD

Sixteen common pathways are down-regulated in ALD and NAFLD. A pathway with high relevance to both diseases is Fatty acid degradation which is down-regulated in ALD and NAFLD but more so in ALD. The KEGG graph in Figure 3A shows down-regulation (green) in nearly all genes for ALD (left part of the gene boxes) while for NAFLD (right part of the gene boxes) there are up-regulated genes such as *ACSL1* and *ACAT1* but more are down-regulated. Interestingly, in the alcohol metabolism at the bottom of the chart, genes are down-regulated in ALD. At the bottom of Figure 3A, alcohol metabolism is shown in a schematic view. In a more detailed view we examined the behaviour of the alcohol dehydrogenase (ADH) encoding genes in the heatmap in Figure 3B and in the aldehyde dehydrogenase genes in Figure 3C. This resulted in a clear image for the ADHs which were down-regulated in ALD. The heatmap for the ALDHs (Figure 3C) looked more complex showing consistently ALD-down-regulated ALDHs only in a cluster at the top including *ALDH2* while most genes were heterogeneously regulated between ALD and NAFLD.

Common pathways up-regulated in ALD and NAFLD

Few pathways (12) were up-regulated in ALD and NAFLD. One of these is ECM-receptor interaction (Supplementary file 1, p. 142). Up-regulation of this pathway might indicate the onset of fibrosis which is accompanied by excessive accumulation of extracellular matrix proteins including collagen[36]. Here, the involvement of the collagen *COL1A1* is shown.

Pathways oppositely regulated in ALD and NAFLD

Of the oppositely regulated pathways, sixteen were down-regulated in ALD and up-regulated in NAFLD while only one was up-regulated in ALD and down in NAFLD (Supplementary Table 3). The Glycolysis pathway was down-regulated in ALD and up-regulated in NAFLD. The KEGG graph (Supplementary file 1, p. 11) shows more down- (green, *e.g.*, *PGM1, ENO1*) than up-regulated (red, *e.g.*, *PFKL*) genes for ALD (left part of gene boxes) while for NAFLD (right part of gene boxes) up-regulated genes predominate. Reduction of glycolysis by ethanol has been brought into context with consumption of oxygen for the alcohol metabolism and has been reported by several authors[37,38]. Berry *et al*[38] reported that ethanol oxidation inhibits glycolysis in rat hepatocytes *via* competition of the reducing equivalents generated during ethanol oxidation with those arising in glycolysis for transfer to the mitochondria.

Pathway-based functional gene annotation

In "genes regulated in common between ALD and NAFLD" we described that after filtering genes with a

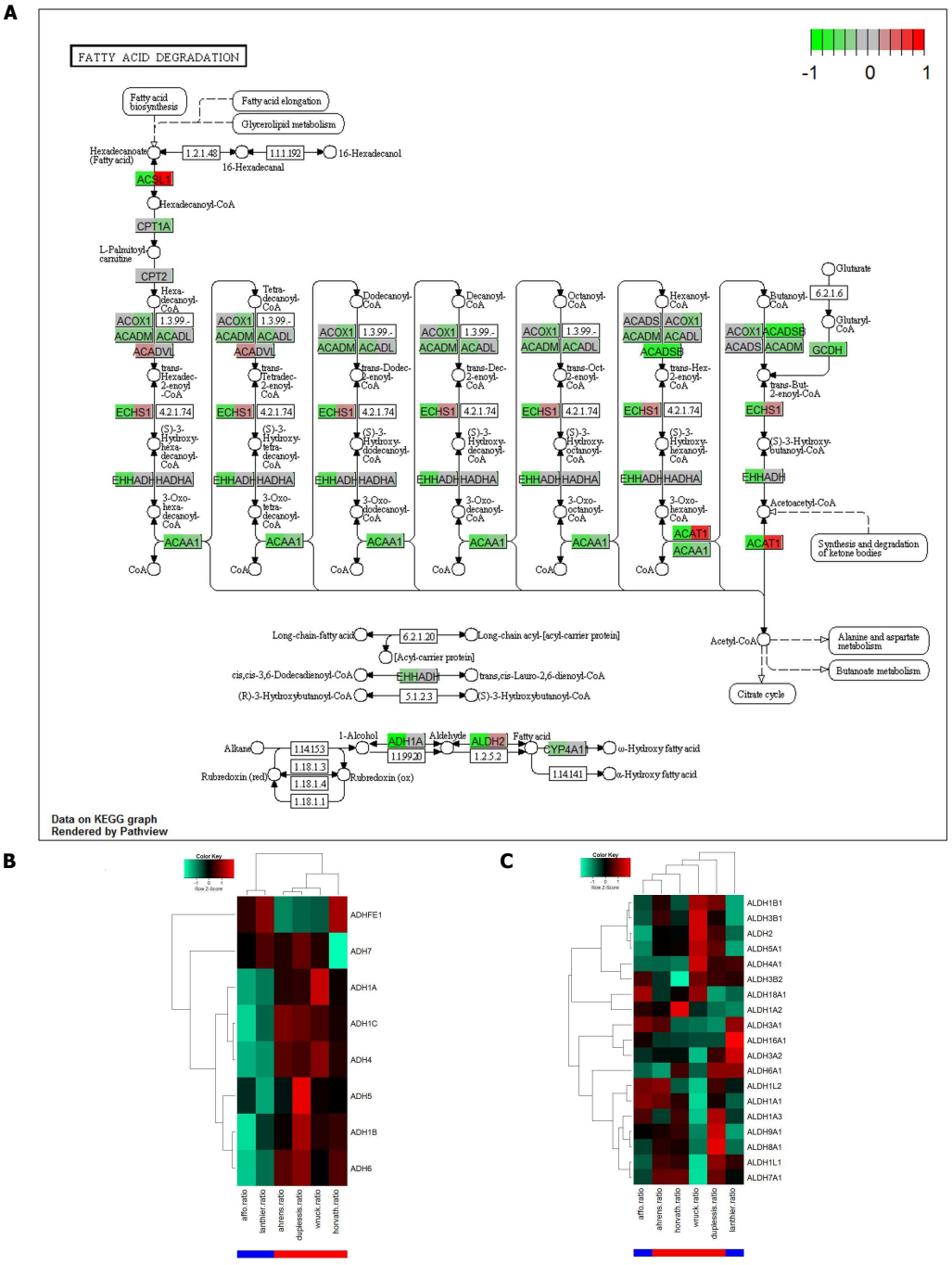


Figure 3 Fatty acid degradation is down-regulated in alcoholic liver disease and non-alcoholic fatty liver disease but more pronounced in alcoholic liver disease. A: The KEGG graph shows down-regulation (green) in nearly all genes for ALD (left part of the gene boxes) while for NAFLD (right part of the gene boxes) there are up-regulated genes such as ACSL1 and ACAT1 but more are down-regulated. Interestingly, in alcohol metabolism at the bottom of the chart, genes are down-regulated in ALD. Alcohol metabolism at the bottom of (A) is shown in detail in the alcohol dehydrogenase (ADH) genes in the heatmap in (B) and in the aldehyde dehydrogenase genes in (C). ADHs are down-regulated in ALD while only dedicated ALDHs, *e.g.*, ALDH2 are down-regulated in ALD. ALD: Alcoholic liver disease; NAFLD: Non-alcoholic fatty liver disease.

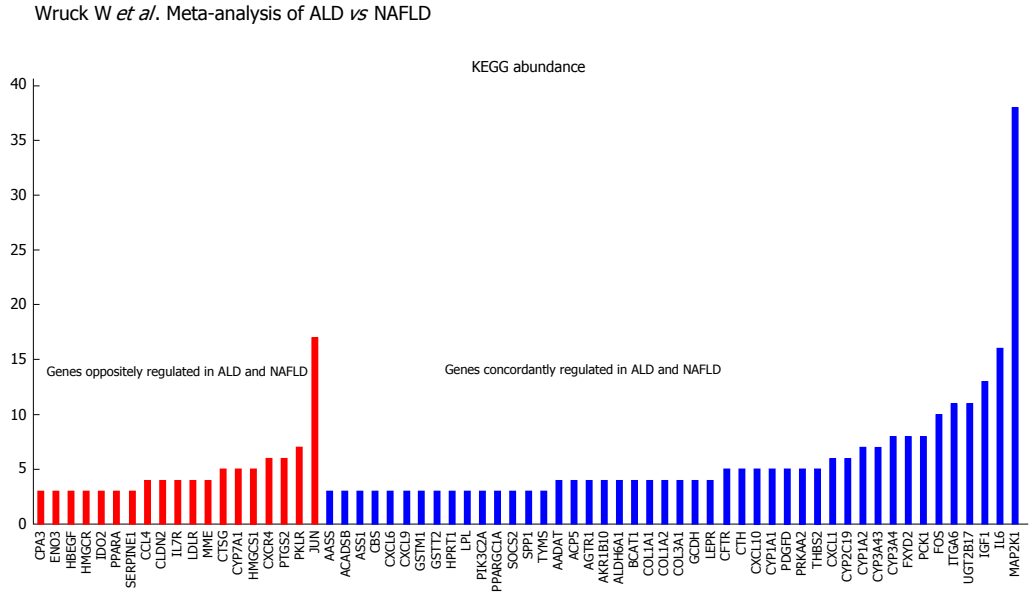


Figure 4 More genes are concordantly than oppositely regulated in alcoholic liver disease and non-alcoholic fatty liver disease. The chart shows the abundance of concordantly and oppositely regulated genes in KEGG pathways (for abundances > 3). The most abundant MAP2K1 (MEK1) refers to the MAPK/RAS-signalling module acting in many KEGG-pathways. JUN which is appearing in 17 KEGG pathways and is down-regulated in ALD and up-regulated in NAFLD shows that there are mechanistic differences in disease pathologies. ALD: Alcoholic liver disease; NAFLD: Non-alcoholic fatty liver disease.

P-value < 0.05 for differential expression more genes were oppositely than concordantly regulated in ALD and NAFLD. This filtering revealed the interesting genes described above but was very restrictive due to the low number of replicates in the condensed ratios - the *P*-values were relatively high. However, the condensed ratios were themselves based on numerous replicates so we consider them as reliable. In a second approach, we filtered genes only by fold change 1.25 and checked on the pathway-level if there were significantly more up- or down-regulated genes based on the binomial test. With this method more genes were concordantly than oppositely regulated in ALD and NAFLD. Figure 4 shows the abundance of concordantly and oppositely regulated genes in KEGG pathways (for abundances > 3). The most abundant MAP2K1 (MEK1) refers to the MAPK/RAS-signalling module acting in many KEGG-pathways. JUN which appears in 17 KEGG pathways and is down-regulated in ALD and up-regulated in NAFLD shows that there are mechanistic differences in molecular basis of these diseases. JUN which is directly connected to c-Jun N-terminal kinase (JNK) was down-regulated in ALD and up-regulated in NAFLD. The up-regulation of JUN in NAFLD is in line with reports from Samuel *et al*[39] showing that activated PKC-ε and JNK can induce insulin resistance *via* impaired IRS1 and IRS2 tyrosine phosphorylation in rats fed with high fat diet.

Pluripotent stem cell-based models of ALD and NAFLD

We recently described a disease-in-a-dish model of steatosis[40]. Pluripotent stem cells, both human embryonic stem cells and induced pluripotent stem cells were differentiated into hepatocyte-like cells and afterwards challenged with ethanol (E) and oleic acid. In order to test how close these models are to the modeled disease we applied our gene signature distinguishing ALD from NAFLD to gene expression data described in Graffmann *et al*[40]. Figure 5 demonstrates that our gene signature can clearly separate two clusters of the ALD and the NAFLD model in a heatmap generated from this gene expression dataset. Furthermore, relevant regulating or rate-limiting genes described above such as *CYP7A1, CYP2E1, HMGCS1, FOXO1* are down-regulated in the ALD-model and up-regulated in the NAFLD-model similar to the liver biopsy-derived dataset.

DISCUSSION

In this comparative analysis of gene expression in ALD and NAFLD liver biopsies we unveiled many commonalities in pathways regulated in the same direction in both diseases. However, there were also pathways regulated in the opposite direction and maybe even more important, essential rate-limiting or regulating genes were adversely regulated. This adverse effect was unexpected as in our working hypothesis, we stated that alcohol is metabolized to fat and beyond this pathway both diseases share a common phenotype. It could hardly be brought together with the common phenotype that of the genes significantly dysregulated between ALD and NAFLD there were more genes regulated in the opposite than in the same direction. One major complex within the adversely regulated genes were cholesterol-related processes including the rate-limiting genes *HMGCR, SQLE, CYP7A1* and *LDLR*. These

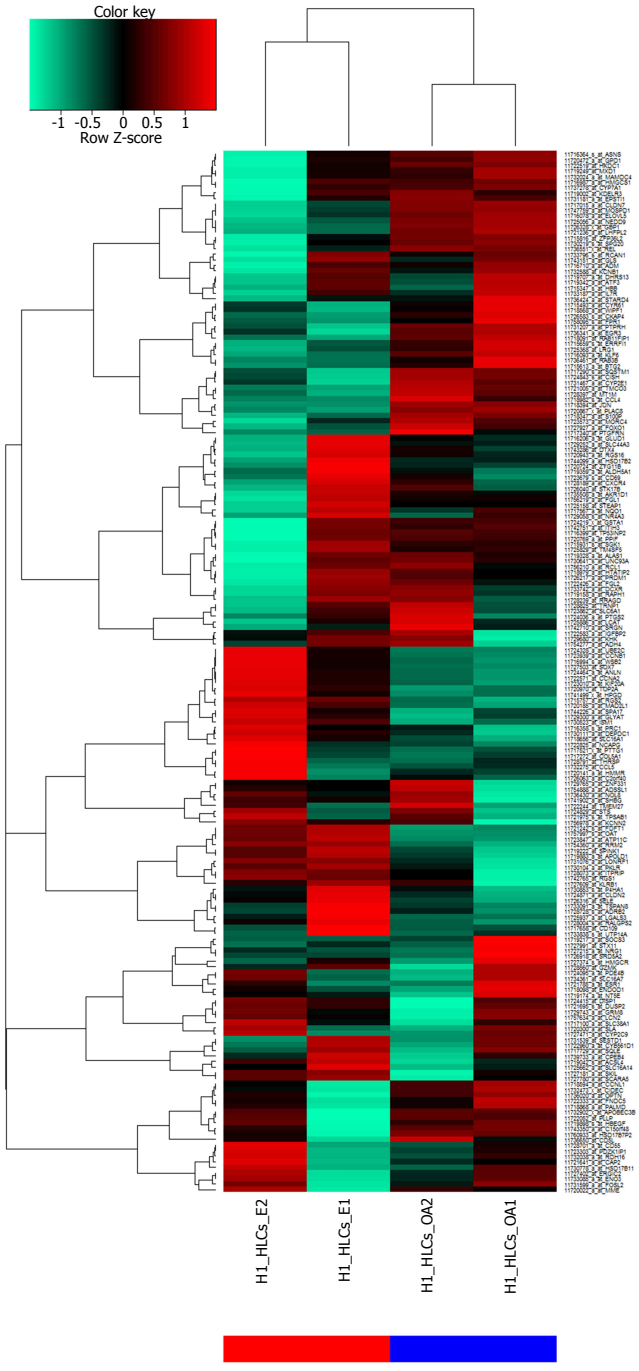


Figure 5 The pluripotent stem cell models of alcoholic liver disease and non-alcoholic fatty liver disease reflect the characteristics of the biopsy-derived gene signature. The gene signature condensed from the meta-analysis of multiple ALD and NAFLD gene expression datasets was applied to the steatosis-model by (Graffmann *et al*[401]) where pluripotent-stem-cell-derived hepatocyte-like cells (HLCs) were challenged with ethanol (E) and oleic acid (OA). The cluster analysis shows a clear separation into the ethanol model (red bar) and the oleic acid model (blue bar). ALD: Alcoholic liver disease; NAFLD: Non-alcoholic fatty liver disease.

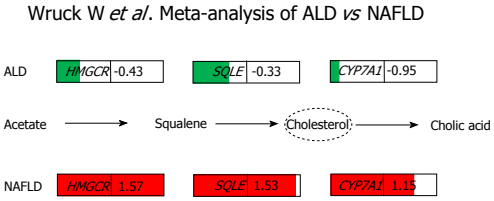


Figure 6 Rate-limiting genes of cholesterol metabolism are down-regulated in alcoholic liver disease and up-regulated in non-alcoholic fatty liver disease. This schematic figure shows the log2-ratios of *HMGCR*, *SQLE* and *CYP7A1* indicating down-regulation in ALD (green) and up-regulation in NAFLD (red). There was stronger down-regulation of *CYP7A1* (log2-ratio = -0.95) than of the upstream cholesterol genes *HMGCR* (log2-ratio = -0.429) and *SQLE* (log2-ratio = -0.33) in ALD while in NAFLD, *CYP7A1* (log2-ratio = 1.15) was weaker up-regulated than *HMGCR* (log2-ratio = 1.57) and *SQLE* (log2-ratio = 1.53). The size of the arrows points to a disequilibrium between cholesterol production and secretion into the bile *via* CYP7A1 in both diseases despite the opposite regulation in ALD and NAFLD. ALD: Alcoholic liver disease; NAFLD: Non-alcoholic fatty liver disease.

were down-regulated in ALD and up-regulated in NAFLD (each compared *vs* healthy control). However, we found in both cases that the gene encoding CYP7A1 - the enzyme responsible for cholesterol removal by catalysing the conversion of cholesterol to bile acids was regulated at a lower level than the genes encoding for the cholesterol synthesis determining enzymes *HMGCR* and *SQLE*. This would explain cholesterol accumulation in the liver because more cholesterol is produced than secreted into bile - regardless if the cholesterol processes are down-regulated in total (in ALD) or up-regulated (in NAFLD). Moreover, the strong down-regulation of *CYP7A1* in ALD might be a clue for the higher risk of cholestasis in ALD than in NAFLD[41]. Briefly, these findings emphasize the importance of cholesterol efflux from the liver *via* CYP7A1 and may suggest that the cause of the disease might be that the rate of cholesterol efflux is too low. Negative feedback loops down-regulating CYP7A1 by bile acids have already been described[42]: Bile acids can down-regulate CYP7A1 *via* (1) FXR and SHP; or (2) by interaction with liver macrophages (Kupffer cells) whose role in fibrosis has been established as they produce cytokines such as transforming growth factor beta leading to the transformation of stellate cells into myofibroblasts[43]. Furthermore, Kupffer cells secrete cytokines, *e.g.*, tumor necrosis factor (TNFα) and interleukin (IL-1β) which in turn induce protein kinase, c-Jun N-terminal kinase and thus inhibit hepatocyte nuclear factor and consequently CYP7A1[44,45]. This gives rise to the question if the lower CYP7A1 levels are a cause of steatosis or are a consequence of the profibrotic stage. Here, systems biology modelling of cholesterol fluxes in the liver including bile acids and regulatory mechanisms of CYP7A1 could be useful in determining under which condition efflux rates are too low.

Beside the differences in cholesterol processes we could also confirm effects which had been much disputed before such as the ethanol-mediated down-regulation of glycolysis and of alcohol and aldehyde dehydrogenases.

The common up-regulated pathways might provide synergies for research into ALD and NAFLD. We found similar mechanisms underlying the progression of both diseases and could identify the common up-regulated ECM-receptor interactions and also associated collagen encoding genes *COL1A1* and *COL3A1* which indicate development of fibrotic tissue.

Finally, we provide a comprehensive compendium displaying comparative regulation of all KEGG pathways in ALD *vs* NAFLD which may serve as an encyclopaedic tool to lookup regulation of dedicated pathways associated with ALD and NAFLD.

In the current study we performed a meta-analysis of gene expression data of liver-derived biopsies from ALD and NAFLD patients, and report a gene signature which clearly separates the transcriptomes of ALD and NAFLD derived liver biopsies. Furthermore, we uncovered predominating commonalities between both diseases at the level of biological pathways, *e.g.*, common down-regulation of the Fatty acid degradation pathway and common up-regulation of the ECM-receptor interaction pathway which may explain common progression of both diseases by cytokines being exchanged between hepatocytes, Kupffer cells and stellate cells at the fibrosis stage. This is confirmed by the common expression of *COL1A1* and *COL3A1* which are associated with fibrotic tissue.

Interestingly, we found rate-limiting genes of cholesterol processes such as *HMGCR*, *SQLE* and *CYP7A1* adversely regulated (Figure 6) between ALD (down-regulated) and NAFLD (up-regulated). The fact that both diseases have the same phenotype may be due to a lower level of the enzyme CYP7A1 compared to the cholesterol synthesis enzymes HMGCR and SQLE. Thus, it will be interesting to further investigate CYP7A1-mediated cholesterol secretion into bile - possibly by systems biology modeling of cholesterol fluxes in the liver. For future therapy, drugs able to adjust CYP7A1 to levels amenable with cholesterol synthesized in or transported to the liver will be useful.

COMMENTS

Background

Non-alcoholic fatty liver disease (NAFLD) and alcoholic liver disease (ALD) are highly prevalent liver diseases and in an increasing number of developed countries NAFLD is becoming the most common cause of liver disease. Although NAFLD and ALD have distinct etiologies the manifestation and the potential progression of both diseases to hepatitis, cirrhosis and cancer is similar.

Research frontiers

A two-hit hypothesis is the established explanation for disease progression to alcoholic hepatitis (AH) and non-alcoholic steatohepatitis (NASH). After steatotic fat accumulation due to metabolic disorders such as insulin resistance (NAFLD) or due to alcohol (ALD) oxidative stress and dysregulation of cytokines initiate inflammation and hence the progression to NASH as well as AH.

Innovations and breakthroughs

The authors found that rate-limiting enzymes of cholesterol metabolism such as HMGCR, SQLE and CYP7A1 are down-regulated in ALD and up-regulated in

NAFLD compared to a healthy control. However, in ALD and NAFLD CYP7A1 - associated with conversion of cholesterol into bile acids - is regulated at a lower level than HMGCR and SQLE. That might explain the accumulation of cholesterol by the reduced efflux into bile acids.

Applications

CYP7A1 is a potential drug target and the proposed gene signature distinguishing ALD from NAFLD consists of biomarkers which may be exploited for diagnostic tests. The compendium of KEGG pathway regulation in ALD and NAFLD and the finding of the adverse regulation of cholesterol metabolism in ALD and NAFLD are promising start points for future research.

Terminology

NAFLD is the disease related to fat accumulation (steatosis) in the liver in the absence of alcohol abuse (usually the threshold is set at 30 g/d of alcohol for men and 20 g/d for women). It ranges from the relatively benign steatosis to NASH, cirrhosis and hepatocellular carcinoma.

Peer-review

This manuscript was informative. The authors found commonalities between both ALD and NAFLD at the level of biological pathways implying some mechanistic similarity between both diseases.

REFERENCES

1 **Ludwig J**, Viggiano TR, McGill DB, Oh BJ. Nonalcoholic steatohepatitis: Mayo Clinic experiences with a hitherto unnamed disease. *Mayo Clin Proc* 1980; **55**: 434-438 [PMID: 7382552]
2 **Edenberg HJ**. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res Health* 2007; **30**: 5-13 [PMID: 17718394]
3 **Jenkins WJ**, Cakebread K, Palmer KR. Effect of alcohol consumption on hepatic aldehyde dehydrogenase activity in alcoholic patients. *Lancet* 1984; **1**: 1048-1049 [PMID: 6143977 DOI: 10.1016/S0140-6736(84)91453-3]
4 **Samson HH**, Morgan DC, Price CM, Tang M, Falk JL. Ethanol elimination rates in normal and ethanol dependent animals. *Pharmacol Biochem Behav* 1976; **5**: 335-341 [PMID: 1033570 DOI: 10.1016/0091-3057(76)90086-1]
5 **Vidal F**, Toda R, Gutiérrez C, Broch M, Fernández-Muixí F, Lorenzo A, Richart C. Influence of chronic alcohol abuse and liver disease on hepatic aldehyde dehydrogenase activity. *Alcohol* 1998; **15**: 3-8 [PMID: 9426831 DOI: 10.1016/S0741-8329(97)00073-6]
6 **Luong A**, Hannah VC, Brown MS, Goldstein JL. Molecular characterization of human acetyl-CoA synthetase, an enzyme regulated by sterol regulatory element-binding proteins. *J Biol Chem* 2000; **275**: 26458-26466 [PMID: 10843999 DOI: 10.1074/jbc.M004160200]
7 **You M**, Fischer M, Deeg MA, Crabb DW. Ethanol induces fatty acid synthesis pathways by activation of sterol regulatory element-binding protein (SREBP). *J Biol Chem* 2002; **277**: 29342-29347 [PMID: 12036955 DOI: 10.1074/jbc.M202411200]
8 **Marra F**, Gastaldelli A, Svegliati Baroni G, Tell G, Tiribelli C. Molecular basis and mechanisms of progression of non-alcoholic steatohepatitis. *Trends Mol Med* 2008; **14**: 72-81 [PMID: 18218340 DOI: 10.1016/j.molmed.2007.12.003]
9 **Lieber CS**. Alcoholic fatty liver: its pathogenesis and mechanism of progression to inflammation and fibrosis. *Alcohol* 2004; **34**: 9-19 [PMID: 15670660 DOI: 10.1016/j.alcohol.2004.07.008]
10 **Day CP**, James OF. Steatohepatitis: a tale of two "hits"? *Gastroenterology* 1998; **114**: 842-845 [PMID: 9547102 DOI: 10.1016/S0016-5085(98)70599-2]
11 **Lucey MR**, Mathurin P, Morgan TR. Alcoholic hepatitis. *N Engl J Med* 2009; **360**: 2758-2769 [PMID: 19553649 DOI: 10.1056/NEJMra0805786]
12 **Wruck W**, Kashofer K, Rehman S, Daskalaki A, Berg D, Gralka E, Jozefczuk J, Drews K, Pandey V, Regenbrecht C, Wierling C, Turano P, Korf U, Zatloukal K, Lehrach H, Westerhoff HV, Adjaye J. Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes. *Sci Data* 2015; **2**: 150068 [PMID: 26646939 DOI: 10.1038/sdata.2015.68]
13 **Sookoian S**, Pirola CJ. Systems biology elucidates common pathogenic mechanisms between nonalcoholic and alcoholic-fatty liver disease. *PLoS One* 2013; **8**: e58895 [PMID: 23516571 DOI: 10.1371/journal.pone.0058895]
14 **Wilfred de Alwis NM**, Day CP. Genetics of alcoholic liver disease and nonalcoholic fatty liver disease. *Semin Liver Dis* 2007; **27**: 44-54 [PMID: 17295176 DOI: 10.1055/s-2006-960170]
15 **Affò S**, Dominguez M, Lozano JJ, Sancho-Bru P, Rodrigo-Torres D, Morales-Ibanez O, Moreno M, Millán C, Loaeza-del-Castillo A, Altamirano J, García-Pagán JC, Arroyo V, Ginès P, Caballería J, Schwabe RF, Bataller R. Transcriptome analysis identifies TNF superfamily receptors as potential therapeutic targets in alcoholic hepatitis. *Gut* 2013; **62**: 452-460 [PMID: 22637703 DOI: 10.1136/gutjnl-2011-301146]
16 **Lanthier N**, Rubbia-Brandt L, Lin-Marq N, Clément S, Frossard JL, Goossens N, Hadengue A, Spahr L. Hepatic cell proliferation plays a pivotal role in the prognosis of alcoholic hepatitis. *J Hepatol* 2015; **63**: 609-621 [PMID: 25872168 DOI: 10.1016/j.jhep.2015.04.003]
17 **Horvath S**, Erhart W, Brosch M, Ammerpohl O, von Schönfels W, Ahrens M, Heits N, Bell JT, Tsai PC, Spector TD, Deloukas P, Siebert R, Sipos B, Becker T, Röcken C, Schafmayer C, Hampe J. Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci USA* 2014; **111**: 15538-15543 [PMID: 25313081 DOI: 10.1073/pnas.1412759111]
18 **du Plessis J**, van Pelt J, Korf H, Mathieu C, van der Schueren B, Lannoo M, Oyen T, Topal B, Fetter G, Nayler S, van der Merwe T, Windmolders P, Van Gaal L, Verrijken A, Hubens G, Gericke M, Cassiman D, Francque S, Nevens F, van der Merwe S. Association of Adipose Tissue Inflammation With Histologic Severity of Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2015; **149**: 635-648.e14 [PMID: 26028579 DOI: 10.1053/j.gastro.2015.05.044]
19 **Ahrens M**, Ammerpohl O, von Schönfels W, Kolarova J, Bens S, Itzel T, Teufel A, Herrmann A, Brosch M, Hinrichsen H, Erhart W, Egberts J, Sipos B, Schreiber S, Häsler R, Stickel F, Becker T, Krawczak M, Röcken C, Siebert R, Schafmayer C, Hampe J. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab* 2013; **18**: 296-302 [PMID: 23931760 DOI: 10.1016/j.cmet.2013.07.004]
20 **Gentleman RC**, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; **5**: R80 [PMID: 15461798 DOI: 10.1186/gb-2004-5-10-r80]
21 **Du P**, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 2008; **24**: 1547-1548 [PMID: 18467348 DOI: 10.1093/bioinformatics/btn224]
22 **Smyth GK**. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004; **3**: Article3 [PMID: 16646809 DOI: 10.2202/1544-6115.1027]
23 **Storey JD**. A direct approach to false discovery rates. *J R Stat Soc Ser B Stat Methodol* 2002; **64**: 479-498 [DOI: 10.1111/1467-9868.00346]
24 **Gautier L**, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004; **20**: 307-315 [PMID: 14960456 DOI: 10.1093/bioinformatics/btg405]
25 **Viechtbauer W**. Conducting meta-analyses in R with the metafor package. *J Stat Softw* 2010; **36**: 1-48 [DOI: 10.18637/jss.v036.i03]
26 **Kanehisa M**, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2016; **44**: D457-D462 [PMID: 26476454 DOI: 10.1093/nar/gkv1070]
27 **Luo W**, Brouwer C. Pathview: an R/Bioconductor package for

pathway-based data integration and visualization. *Bioinformatics* 2013; **29**: 1830-1831 [PMID: 23740750 DOI: 10.1093/bioinformatics/btt285]
28 **Mitchell RM**, Byrne MF, Baillie J. Pancreatitis. *Lancet* 2003; **361**: 1447-1455 [PMID: 12727412 DOI: 10.1016/S0140-6736(03)13139-X]
29 **Whitcomb DC**, Gorry MC, Preston RA, Furey W, Sossenheimer MJ, Ulrich CD, Martin SP, Gates LK, Amann ST, Toskes PP, Liddle R, McGrath K, Uomo G, Post JC, Ehrlich GD. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nat Genet* 1996; **14**: 141-145 [PMID: 8841182 DOI: 10.1038/ng1096-141]
30 **Nunn C**, Zhao P, Zou MX, Summers K, Guglielmo CG, Chidiac P. Resistance to age-related, normal body weight gain in RGS2 deficient mice. *Cell Signal* 2011; **23**: 1375-1386 [PMID: 21447383 DOI: 10.1016/j.cellsig.2011.03.020]
31 **Wruck W**, Graffmann N, Kawala MA, Adjaye J. Current Status and Future Directions on Research Related to Nonalcoholic Fatty Liver Disease. *Stem Cells* 2016; **35**: 89-96 [PMID: 27374784 DOI: 10.1002/stem.2454]
32 **Lakshmanan MR**, Veech RL. Short- and long-term effects of ethanol administration in vivo on rat liver HMG-CoA reductase and cholesterol 7alpha-hydroxylase activities. *J Lipid Res* 1977; **18**: 325-330 [PMID: 864324]
33 **Lieber CS**, Leo MA, Wang X, Decarli LM. Alcohol alters hepatic FoxO1, p53, and mitochondrial SIRT5 deacetylation function. *Biochem Biophys Res Commun* 2008; **373**: 246-252 [PMID: 18555008 DOI: 10.1016/j.bbrc.2008.06.006]
34 **Ikonen E**. Cellular cholesterol trafficking and compartmentalization. *Nat Rev Mol Cell Biol* 2008; **9**: 125-138 [PMID: 18216769 DOI: 10.1038/nrm2336]
35 **Congiu M**, Slavin JL, Desmond PV. Expression of common housekeeping genes is affected by disease in human hepatitis C virus-infected liver. *Liver Int* 2011; **31**: 386-390 [PMID: 21073651 DOI: 10.1111/j.1478-3231.2010.02374.x]
36 **Bataller R**. Liver fibrosis. *J Clin Invest* 2005; **115**: 1100-1100 [DOI: 10.1172/JCI200524282C1]
37 **Young TA**, Bailey SM, Van Horn CG, Cunningham CC. Chronic ethanol consumption decreases mitochondrial and glycolytic production of ATP in liver. *Alcohol Alcohol* 2006; **41**: 254-260 [PMID: 16571619 DOI: 10.1093/alcalc/agl017]
38 **Berry MN**, Gregory RB, Grivell AR, Phillips JW, Schön A. The capacity of reducing-equivalent shuttles limits glycolysis during ethanol oxidation. *Eur J Biochem* 1994; **225**: 557-564 [PMID: 7957170 DOI: 10.1111/j.1432-1033.1994.00557.x]
39 **Samuel VT**, Liu ZX, Qu X, Elder BD, Bilz S, Befroy D, Romanelli AJ, Shulman GI. Mechanism of hepatic insulin resistance in non-alcoholic fatty liver disease. *J Biol Chem* 2004; **279**: 32345-32353 [PMID: 15166226 DOI: 10.1074/jbc.M313478200]
40 **Graffmann N**, Ring S, Kawala MA, Wruck W, Ncube A, Trompeter HI, Adjaye J. Modeling Nonalcoholic Fatty Liver Disease with Human Pluripotent Stem Cell-Derived Immature Hepatocyte-Like Cells Reveals Activation of PLIN2 and Confirms Regulatory Functions of Peroxisome Proliferator-Activated Receptor Alpha. *Stem Cells Dev* 2016; **25**: 1119-1133 [PMID: 27308945 DOI: 10.1089/scd.2015.0383]
41 **Tannapfel A**, Denk H, Dienes HP, Langner C, Schirmacher P, Trauner M, Flott-Rahmel B. Histopathological diagnosis of non-alcoholic and alcoholic fatty liver disease. *Virchows Arch* 2011; **458**: 511-523 [PMID: 21442288 DOI: 10.1007/s00428-011-1066-1]
42 **Davis RA**, Miyake JH, Hui TY, Spann NJ. Regulation of cholesterol-7alpha-hydroxylase: BAREly missing a SHP. *J Lipid Res* 2002; **43**: 533-543 [PMID: 11907135]
43 **Kolios G**, Valatas V, Kouroumalis E. Role of Kupffer cells in the pathogenesis of liver disease. *World J Gastroenterol* 2006; **12**: 7413-7420 [PMID: 17167827 DOI: 10.3748/wjg.v12.i46.7413]
44 **Gupta S**, Stravitz RT, Dent P, Hylemon PB. Down-regulation of cholesterol 7alpha-hydroxylase (CYP7A1) gene expression by bile acids in primary rat hepatocytes is mediated by the c-Jun N-terminal kinase pathway. *J Biol Chem* 2001; **276**: 15816-15822 [PMID: 11278771 DOI: 10.1074/jbc.M010878200]
45 **Li T**, Jahan A, Chiang JY. Bile acids and cytokines inhibit the human cholesterol 7 alpha-hydroxylase gene via the JNK/c-jun pathway in human liver cells. *Hepatology* 2006; **43**: 1202-1210 [PMID: 16729332 DOI: 10.1002/hep.21183]

P- Reviewer: de Oliveira CPMS, Qu BG, Sanal MG, Tarantino G
S- Editor: Ji FF **L- Editor**: A **E- Editor**: Li D

2.1.5 Nijmegen Breakage Syndrome fibroblasts and iPSCs: cellular models for uncovering disease-associated signaling pathways and establishing a screening platform for anti-oxidants.

Sci Rep. 2017 Aug 8;7(1):7516. doi: 10.1038/s41598-017-07905-2.

Mlody B(1), Wruck W(2), Martins S(2), Sperling K(3), Adjaye J(4).

Author information: (1)Max-Delbrück-Centrum für Molekulare Medizin (MDC), 13092, Berlin, Germany. (2)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225, Düsseldorf, Germany. (3)Charité - Universitätsmedizin Berlin, Institute of Medical and Human Genetics, 13353, Berlin, Germany. (4)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225, Düsseldorf, Germany. James.Adjaye@med.uni-duesseldorf.de.

Nijmegen Breakage Syndrome (NBS) is associated with cancer predisposition, premature aging, immune deficiency, microcephaly and is caused by mutations in the gene coding for NIBRIN (NBN) which is involved in DNA damage repair. Dermal-derived fibroblasts from NBS patients were reprogrammed into induced pluripotent stem cells (iPSCs) in order to bypass premature senescence. The influence of antioxidants on intracellular levels of ROS and DNA damage were screened and it was found that EDHB-an activator of the hypoxia pathway, decreased DNA damage in the presence of high oxidative stress. Furthermore, NBS fibroblasts but not NBS-iPSCs were found to be more susceptible to the induction of DNA damage than their healthy counterparts. Global transcriptome analysis comparing NBS to healthy fibroblasts and NBS-iPSCs to embryonic stem cells revealed regulation of P53 in NBS fibroblasts and NBS-iPSCs. Cell cycle related genes were down-regulated in NBS fibroblasts. Furthermore, oxidative phosphorylation was down-regulated and glycolysis up-regulated specifically in NBS-iPSCs compared to embryonic stem cells. Our study demonstrates the utility of NBS-iPSCs as a screening platform for anti-oxidants capable of suppressing DNA damage and a cellular model for studying NBN de-regulation in cancer and microcephaly.

DOI: 10.1038/s41598-017-07905-2 PMCID: PMC5548734 PMID: 28790359 [Indexed for MEDLINE]

URL: https://www.nature.com/articles/s41598-017-07905-2

The data presented in this publication has been used also for the PhD thesis of Dr. Soraia Martins at the Heinrich Heine University Düsseldorf.

Contribution: Wasco Wruck wrote the manuscript and analyzed the data. Barbara Mlody and Wasco Wruck contributed equally to this work.

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Nijmegen Breakage Syndrome fibroblasts and iPSCs: cellular models for uncovering disease-associated signaling pathways and establishing a screening platform for anti-oxidants

Received: 20 March 2017
Accepted: 4 July 2017
Published online: 08 August 2017

Barbara Mlody[1], Wasco Wruck[2], Soraia Martins[2], Karl Sperling[3] & James Adjaye[2]

Nijmegen Breakage Syndrome (NBS) is associated with cancer predisposition, premature aging, immune deficiency, microcephaly and is caused by mutations in the gene coding for NIBRIN (NBN) which is involved in DNA damage repair. Dermal-derived fibroblasts from NBS patients were reprogrammed into induced pluripotent stem cells (iPSCs) in order to bypass premature senescence. The influence of antioxidants on intracellular levels of ROS and DNA damage were screened and it was found that EDHB-an activator of the hypoxia pathway, decreased DNA damage in the presence of high oxidative stress. Furthermore, NBS fibroblasts but not NBS-iPSCs were found to be more susceptible to the induction of DNA damage than their healthy counterparts. Global transcriptome analysis comparing NBS to healthy fibroblasts and NBS-iPSCs to embryonic stem cells revealed regulation of P53 in NBS fibroblasts and NBS-iPSCs. Cell cycle related genes were down-regulated in NBS fibroblasts. Furthermore, oxidative phosphorylation was down-regulated and glycolysis up-regulated specifically in NBS-iPSCs compared to embryonic stem cells. Our study demonstrates the utility of NBS-iPSCs as a screening platform for anti-oxidants capable of suppressing DNA damage and a cellular model for studying NBN de-regulation in cancer and microcephaly.

Nijmegen Breakage Syndrome (NBS) is a rare autosomal recessive genetic disorder, first described 1981 in Nijmegen, the Netherlands[1]. Characteristics of NBS include genomic instability (resulting in early onset of malignancies), premature aging, microcephaly and other growth retardations, immune deficiency, impaired puberty and infertility in females. The consequence of these manifestations is a severe decrease in average life span, caused by cancer or infection of the respiratory and urinary tracts[2].

On a molecular basis, NBS is caused by mutations in the gene coding for NIBRIN (*NBN*) which is involved in DNA damage repair[3]. Mutated versions of *NBN* cause accumulation of unrepaired DNA damage leading to cell cycle arrest, apoptosis[4] or accumulation of genomic point mutations and aberrations introduced by misregulated DNA repair[5]. Several cases of NBS with a variety of mutations in *NBN* exist but over 90% of the patients carry a 5 base pair deletion (657del5) within the *NBN* exon 6[6].

This hypomorphic mutation leads to a truncated 26 kD amino-terminal protein and a 70 kD carboxy-terminal protein due to alternative translation from a cryptic start site upstream of the deletion[7]. Mice *Nbn* null mutations are embryonic lethal and cells expressing only the truncated p26kD NBN fragment containing the FHA and the first BRCT domain, were nonviable[7]. The new splice form, p70 retains sufficient functionality to ensure

[1]Max-Delbrück-Centrum für Molekulare Medizin (MDC), 13092, Berlin, Germany. [2]Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225, Düsseldorf, Germany. [3]Charité - Universitätsmedizin Berlin, Institute of Medical and Human Genetics, 13353, Berlin, Germany. Barbara Mlody and Wasco Wruck contributed equally to this work. Correspondence and requests for materials should be addressed to J.A. (email: James.Adjaye@med.uni-duesseldorf.de)

survival by binding to MRE11 and ATM, which are essential components of DNA damage response[8]. The MRE11-RAD50-NBN (MRN) complex binds directly to DNA double-strand breaks (DSBs) and is involved in repair and signaling for homologous recombination (HR), non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ). Additionally, NBN is involved in telomere maintenance and therefore plays a role in the aging process[8].

Recent works indicate that NBN influences the repair pathway choice via 53BP1, which can shift the error-free HR-directed repair to the more error-prone NHEJ and MMEJ[9]. Apart from replication errors, mutagens and other external influences, endogenously, DNA damage is mostly caused by reactive oxygen species (ROS), which are byproducts of the respiratory chain reaction[10]. Cells counteract ROS by antioxidant production and enzymatic removal but ROS also have cellular signaling functions which must be maintained in a controlled balance[11].One strategy to minimize endogenous ROS levels is to regulate mitochondrial respiration, which plays a special role in stem cells.

Stem cell mitochondrial morphology is immature, rounded and with under-developed cristae. Consequently, they depend heavily on glycolysis for their ATP supply[12]. When cells differentiate and increase respiration, mitochondrial mass increases, their morphology then shifts to more matured and elongated tubular forms, with more defined cristae and increased mtDNA copy numbers[12].

When somatic cells are reprogrammed into induced pluripotent stem cells (iPSCs), they depend predominantly on glycolysis and their mitochondria become rejuvenated and transformed back to the immature form[13].

A key element in the reprogramming of metabolism is the HIF1-alpha pathway, which not only reacts in response to hypoxia, but also induces a shift from oxidative phosphorylation to glycolysis[14]. We have reported this "metabolic reprogramming" as an essential step in iPSC-generation, which precedes the activation of pluripotency-associated genes like OCT4 and NANOG[15].

The aim of this study was to use our previously published iPSC-based cellular model system for NBS and provide a screening platform for antioxidants capable of modulating genome stability. NBS-iPSCs may overcome several problems associated with NBS research such as: i) small patient numbers, ii) cell cultures limited to fibroblasts and lymphocytes, iii) premature senescence in cell culture due to high levels of ROS, iv) discovery of new NBS molecular mechanisms and v) provision of new and therapeutically relevant concepts.

There are several diseases like NBS which derive from mutated genes in repair pathways, examples of these include Fanconi Anemia (FA)[16], Ligase IV (LIG4) syndrome[17],Bloom syndrome[18],NBS-like disorder[19], Ataxia-Telangiectasia-Like Disorder (ATLD)[20], Nonhomologous end-joining factor 1 (NHEJ1) syndrome[21] and Seckel syndrome[22].

Our group recently published a study on modeling NBS by reprogramming[23]. Reprogrammed cells from patients with similar diseases like FA have been reported, though they could only be reprogrammed after genetic correction or with the aid of antioxidants[24]. In a study of patients with Cockayne syndrome (CS), a mutation in the repair pathway gene ERCC6 did not impair genetic reprogramming but exhibited elevated cell death rates and ROS production[25]. As NBS cells are hypersensitive to DNA damage[26], ROS may be detrimental to them under physiological conditions. Thus, it was hypothesized that antioxidants or the induction of pluripotency in NBS fibroblasts might suppress and maybe bypass ROS-mediated genome instability.

Microcephaly is a significant physical characteristic of NBS which can also be found in FA, LIG4 and NHEJ1 syndromes[16, 20, 21]. With recent cases in microcephaly which coincided with infections of the Zika virus[27], NBS-iPSCs and iPSC-derived neurons could serve as an excellent comparative model to study NBN-deregulation and associated molecular mechanisms underlying the onset of microcephaly.

In this study we present NBS fibroblasts and iPSCs as a screening platform for anti-oxidants and a model for studying NBN de-regulation in cancer and microcephaly. The screen for antioxidants capable of counteracting intracellular levels of ROS and DNA damage identified Ethyl-3,4-dihydroxybenzoate (EDHB) - an activator of the hypoxia pathway – as most potent antagonist of DNA damage in the presence of high oxidative stress in our NBS-model. Another finding was the higher susceptibility of NBS fibroblasts to induction of DNA damage compared to NBS-iPSCs. Furthermore, we found de-regulation of P53 in NBS fibroblasts and NBS-iPSCs, down-regulation of Cell cycle in NBS fibroblasts and down-regulation of oxidative phosphorylation and up-regulation of glycolysis in NBS-iPSCs compared to healthy embryonic stem cells.

Results

Roadblocks in reprogramming of NBS fibroblasts. Reprogramming of somatic cells towards pluripotent stem cells (PSCs) was reported to be negatively affected when P53 was activated[28]. Given the nature of NBS, which includes genomic instability and premature senescence, both of which are features known to lead to P53 activation[29], thus, hurdles for the reprogramming process were anticipated. To address these, we attempted reprogramming dermal fibroblast primary cultures from eight (8) clinically diagnosed patients with NBS.

Of the 8 NBS fibroblast lines (Table 1), 3 lines could not be cultured past passage 6 and were lost due to premature senescence. Four (4) lines were infected with retroviral reprogramming cocktail (O/S/K/M) but exhibited low infection efficiency (determined by O/S/K/M immuno-staining), senescence (by morphology), hardly showed changes in morphology (negative indicator for reprogramming) and did not yield any iPSC colonies (data not shown).

As previously reported and characterized, only one of the four NBS fibroblast cell lines (NBS-8) subjected to the reprogramming process was successful[23, 30]. This shows that fibroblasts from NBS patients can be reprogrammed to pluripotency despite genomic instability and premature senescence[30]. Sanger-sequencing of *NBN* exon 6 and Western Blotting confirmed the heterozygous mutation for 657del5 in NBS-8 fibroblasts (Supplementary Figure S1) and in the NBS-8-iPS cells (Supplementary Figure S2). As the 657del5 mutation leads to a truncation of NBN on the protein level (wt: 95 kDa; 657del5: 70 kDa), we detected NBN by western blot. Full-length NBN was not present in any of the NBS fibroblasts including NBS-8 and the NBS-iPSCs[30].

Cell line [NBS #]	Gender	Passage [#]	Premature senescence	NBN (657del5) Mutation
1	male	16	No	homozygous
2	male	5	Yes	homozygous
3	female	8	No	homozygous
4	female	14	Yes	homozygous
5	female	4	No	homozygous
6	female	10	Yes	homozygous
7	male	10	No	homozygous
8	male	3	No	heterozygous

Table 1. Fibroblasts lines from NBS patients and their behavior in reprogramming. n/a: not available; #: number.

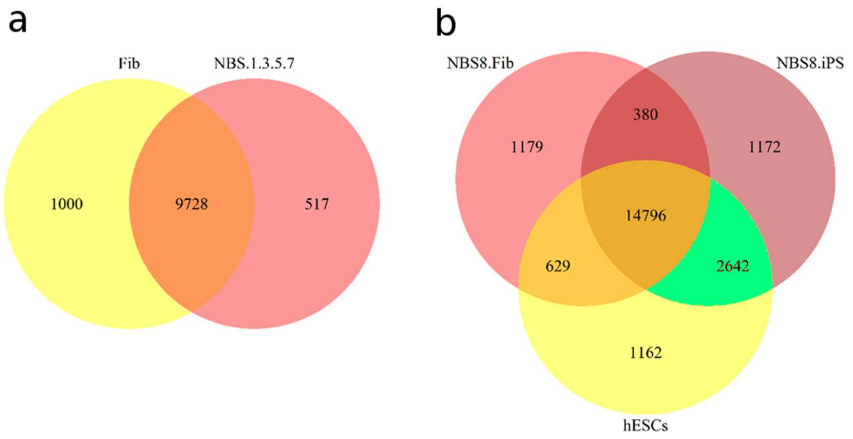
Transcription Regulator	Fold Change (Array data)	Predicted Activation State	Regulation z-score	Number of target molecules
TP53	1.48	Activated	4.32	175
CDKN2A	1.05	Activated	3.93	46
SMARCA4	−1.49	Activated	3.70	34
SMARCE1	1.24	Activated	2.45	6
TCF3	2.02	Activated	2.39	35
Rb (group)	n/a	Activated	2.26	14
GATA1	−1.09	Activated	2.24	11
TP63	−1.00	Activated	2.23	27
GLI3	1.65	Activated	2.11	4
SMAD7	−1.63	Inhibited	−2.13	16
SREBF1	−1.26	Inhibited	−2.31	22
RXRA	1.08	Inhibited	−2.46	17
SREBF2	1.16	Inhibited	−2.48	15
E2F1	1.36	Inhibited	−2.60	63
TBX2	−1.72	Inhibited	−2.77	24
MYCN	−1.07	Inhibited	−4.41	33

Table 2. Regulation changes in transcription factors in NBS fibroblasts (Ingenuity® Prediction Tool).

In vitro cultivation may introduce stress to the cells which may also elevate DNA double-strand breaks events. Accumulation of unrepaired damage DNA leads to the activation of P53, which is a known roadblock of induction of pluripotency as a result of cell cycle arrest, senescence or apoptosis[31]. Table 1 shows that the reprogrammable cell line NBS8 is heterozygous while the other cell lines are homozygous. It is probable that the homozygous nature of the mutated NBS gene increases the level of unrepaired damaged DNA hence drastically reducing the efficiency of inducing pluripotency. To examine this further, we performed transcriptome analysis of primary NBS fibroblasts cell cultures (1, 3, 5, 7 and 8).

Ingenuity® Pathway Analysis (IPA) was used to predict the status of transcription factors (TFs) using a list of differentially regulated genes between NBS and normal fibroblasts (Table 2). In the IPA results in Table 2 the fold change was determined by the expression data of the transcription factor itself but this can be different from the data inferred from the expression data of the de-regulated genes. The most significant, activated TF was P53 (TP53), well known to negatively interfere with reprogramming efficiency. We also found that the P53 pathway was significantly enriched in the transcriptomic analysis of the NBS fibroblasts (see "NBS-iPSCs as a model for studying molecular mechanisms associated with impaired DNA repair"). Among the most significant down-regulated or inactivated TFs was MYCN, which is known as transcriptional regulator in pluripotent stem cells[32]. This could also be a roadblock to reprogramming, as cells are required to proliferate continually during this process.

Which genes/factors contribute to bypassing of cell cycle arrest, senescence or apoptosis. To understand which mechanisms possibly enabled NBS fibroblasts to achieve pluripotency and therefore overcome cellular senescence, we compared transcriptomic data between NBS-8 fibs, NBS-8-iPSCs and hESCs. By identification of the overlaps for expressed genes (determined by expression p value < 0.01) we found 2642 genes to be commonly expressed in NBS-iPSCs and hESCs, but not in NBS fibroblasts (Fig. 1). This subset of genes may contain the distinct profile enabling NBS fibroblasts to reach pluripotency and therefore rejuvenation. We further analyzed the subset in an annotation database. Among the most over-represented results we found MAPK signaling pathway and genes that were specifically expressed in the brain (Fig. 1c). The MAPK signaling pathway is regulated by OCT4 and plays an important role in pluripotency and self-renewal[33].



c

	Term	Count	PValue	Benjamini
KEGG Pathway	Tight junction	31	0.0000	1.14E-04
	Leukocyte transendothelial migration	26	0.0000	0.00139394
	VEGF signaling pathway	16	0.0015	0.08301263
	Cell adhesion molecules (CAMs)	23	0.0018	0.07392233
	MAPK signaling pathway	38	0.0022	0.07281746
	Fc epsilon RI signaling pathway	14	0.0148	0.34659062
UP TISSUE	Brain	804	0.0000	5.33E-04
	Fetal brain	108	0.0000	0.00115483
	Hippocampus	63	0.0010	0.10176118
	Temporal cortex	4	0.0124	0.62532518
GOTERM BP FAT	neuron differentiation	73	0.0000	0.00327788
	cell projection organization	61	0.0000	0.01734537
	cell-cell adhesion	49	0.0000	0.01613593
	neuron projection development	46	0.0000	0.01692829
	homophilic cell adhesion	28	0.0001	0.03955928
	neuron development	54	0.0001	0.05773437
	neuron projection morphogenesis	38	0.0001	0.06519128
	axonogenesis	35	0.0002	0.07923858
	cell morphogenesis involved in neuron differentiation	37	0.0002	0.07357017
	cell adhesion	94	0.0002	0.0715399
	biological adhesion	94	0.0002	0.0682448
	axon guidance	23	0.0003	0.07933291
	cell projection morphogenesis	41	0.0003	0.07438399
	regulation of transmembrane transport	10	0.0005	0.11918867
	cell morphogenesis involved in differentiation	40	0.0005	0.11540241
	regulation of transporter activity	11	0.0006	0.12107772
	heart development	36	0.0007	0.13604176
	cell part morphogenesis	41	0.0007	0.12930809
	transmission of nerve impulse	52	0.0007	0.12971722
	regulation of ion transmembrane transport	9	0.0014	0.21937012

Figure 1. Statistics of Venn diagram analysis among NBS fibroblasts, NBS iPSCs and hESCs. The overlap of significantly expressed genes (detection p value < 0.01) in (**a**) Fibroblasts comparing the averaged group of NBS with unaffected cell lines and in (**b**) NBS fibroblasts versus NBS-iPSCs and unaffected hESCs. (**c**) Annotations for genes commonly expressed in NBS-iPSCs and hESCs but not NBS fibroblasts resulting from functional annotation analysis via the DAVID web tool. The output of the DAVID analysis was condensed to the *count* of genes annotated with the indicated category, *p-value* and *Benjamini-Hochberg*-correction for multiple-testing as calculated based on the Fisher-exact test.

Global transcriptome analysis of NBS-iPSCs. Global transcriptome analysis with NBS fibroblasts and NBS-iPSCs was performed to identify the problems interfering with reprogramming in the cell lines 1,3,5,7, determine NBS phenotypes or compensatory mechanisms in NBS-iPSCs derived from NBS-8 fibroblasts. In the cluster dendrogram (Fig. 2a), fibroblasts from NBS patients clearly clustered as a group and differed from normal fibroblasts, indicating a common transcriptional phenotype distinctive for NBS. Transcriptomes from NBS-iPSCs clustered closer to hESCs than to other fibroblasts-derived iPSCs (Fig. 2b). But, the parental fibroblast line NBS-8, clustered more distinct to the other fibroblasts. The pronounced gap in clustering between NBS-8-Fib-P8 and NBS-8-Fib-P15 indicates acquisition of mutations or aberrations since they only differ in passage number.

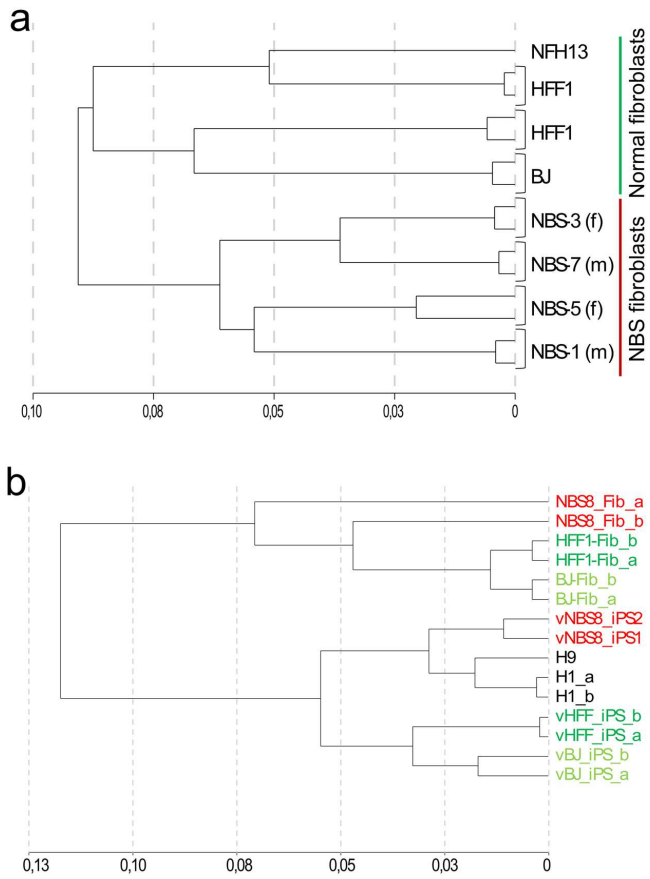


Figure 2. Global transcriptomic comparison of NBS fibroblasts and iPSCs to healthy controls. (**a**) Hierarchical cluster analysis depicting the distance of control (BJ, HFF1, NFH13) and patient (NBS-1, -3, -5, -7) fibroblasts global mRNA (Illumina 8-chip). (**b**) Hierarchical cluster analysis of total mRNA depicting the distance of control (BJ-Fib, HFF1-Fib) and patient (NBS8-Fib, passage number 8 and 15) fibroblasts, plus control (H1, H9, vHFF-iPS, vBJ-iPS) and patient (vNBS8-iPS, clone 1 and 2) pluripotent stem cells (Illumina 12-chip).

After selecting genes that were significantly de-regulated (p value < 0.05; fold change > 1.5) between the groups of NBS and normal fibroblasts, the list was subjected to "DAVID Annotation Tools"[34] to identify pathways which were most affected by mutated NBN (Fig. 3a). The same procedure was performed for the analysis of NBS-iPSCs in comparison with hESCs (Fig. 3b). There were different regulatory changes in both groups of analyses (NBS fibroblasts and NBS-8-iPSCs), but there was also overlap of pathways, indicating NBS specific traits in cell cycle and cancer. *Apoptosis* and *P53*, two of the safeguard mechanisms against cancer, were predominantly de-regulated in NBS fibroblasts than pluripotent NBS cells while *Mismatch repair*, another safeguard mechanism against cancer, was predominantly de-regulated in pluripotent NBS cells. *Mismatch repair* is illustrated in more detail in the heatmap in Fig. 3c (color bars: blue = NBS, red = healthy). In *Cell cycle* we observed a shift from down-regulation to equally balanced (Fig. 3a,b,d,e).

Interestingly, in NBS-iPSCs, the *Glycolysis*-pathway was significantly enriched. Most enzymes involved in glycolysis, including phosphofructokinase, muscle (PFKM, 2-fold) which catalyzes the rate-limiting step, were at least 1.5-fold up-regulated (Supplementary Figure S3). On the other hand, Fructose-1,6-bisphosphatase 1 (FBP1), a gluconeogenesis regulatory enzyme, was significantly down-regulated (2.6-fold). As previously reported, hESCs derive their energy from glycolysis rather than OXPHOS and have immature mitochondria[13,35]. It was also observed that cells acquire the same metabolic profile during the reprogramming process[13,35]. NBS-iPSCs in this case, depended even stronger on glycolysis than other PSCs.

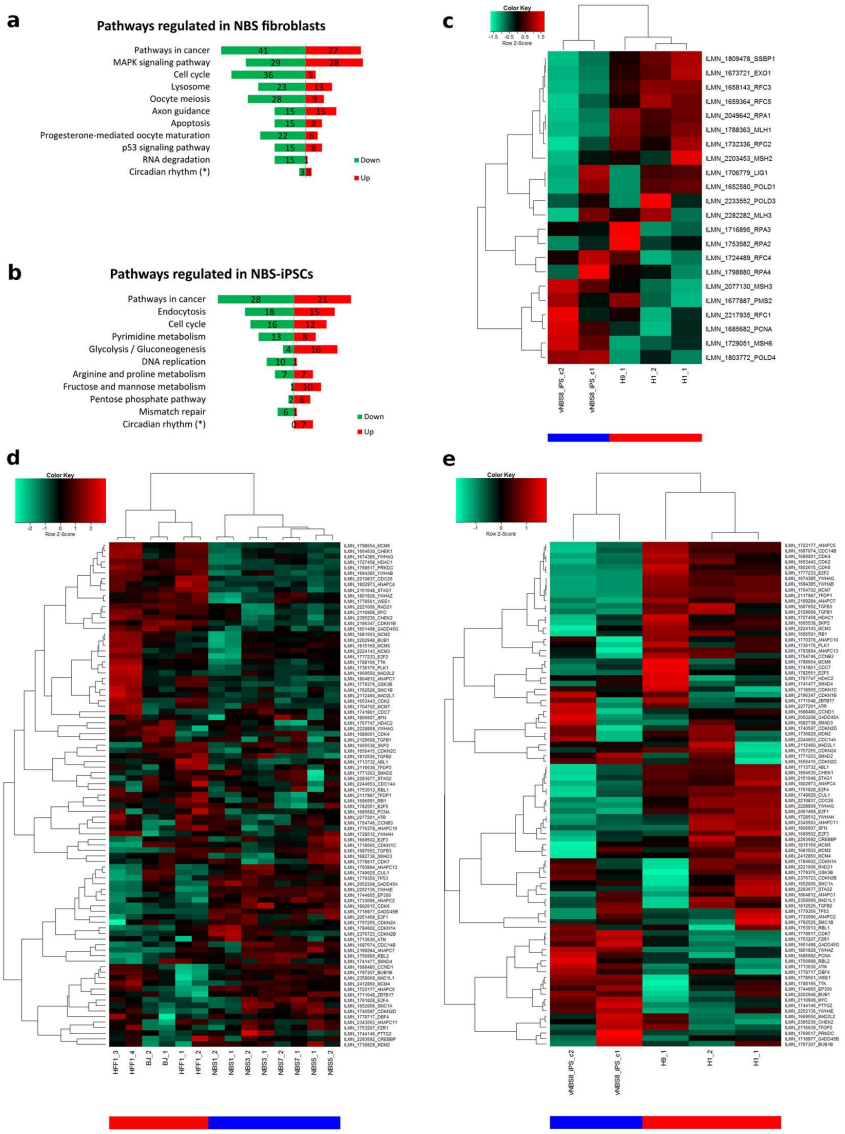


Figure 3. Functional transcriptomics study of NBS fibroblasts and iPSCs. Significantly de-regulated genes (differential p value < 0.05; fold-ratio > 1.5) between control and NBS-patient cells were analyzed by DAVID functional annotation tool (https://david.ncifcrf.gov/). The top 10 significantly (p value < 0.05) de-regulated KEGG pathways (http://www.genome.jp/kegg/pathway.html) are represented in the figure, numbers of significantly up-regulated genes are shown in red and down-regulated genes in green. (**a**) comparison between control (BJ, HFF1) and patient (NBS-1, -3, -5, -7) fibroblasts (**b**) comparison between control (H1, H9) and patient (vNBS8-iPS, clone 1 and 2) pluripotent stem cells. (*) p value of pathway > 0.05. The reprogramming procedure induced a shift of the Cell cycle pathway from down-regulated to nearly balanced. Furthermore, strong down-regulation of the Mismatch repair pathway was found in NBS-iPSCs. Cluster analysis and heatmaps of these de-regulated pathways are shown in (**c**–**e**): (**c**) depicts the Mismatch repair in NBS-iPSCs, (**d**) the Cell cycle in NBS fibroblasts and (**e**) the Cell cycle in NBS-iPSCs (Color bars: blue NBS, red control).

NBS-iPSCs as a model for studying molecular mechanisms associated with impaired DNA repair. In line with the known predisposition of NBS patients to cancer, we found *Pathways in cancer* as the most enriched pathway in NBS-fibroblasts compared to healthy fibroblasts as well as in NBS-iPSCs compared to embryonic stem cells via DAVID analysis[34] (Fig. 3a,b). With further cancer-related pathways, *Cell cycle* was found to be enriched in NBS-fibroblasts and NBS-iPSCs, while *Apoptosis* and *p53 signaling* were only enriched in NBS-fibroblasts. To further explore the relevance of the KEGG pathways in cancer[36] we performed a further DAVID analysis with the genes annotated with that pathway in differentially expressed genes in NBS-iPSCs compared to embryonic stem cells and this way could refine the functional annotation of the *pathways in cancer* (suppl. Table. S2). As expected numerous specific cancer types such as lung cancer and Melanoma are annotated with these genes but also pathways related to NBS emerge. *Cell cycle* (Supplementary Figure S4) and *p53 signaling* (Supplementary Figure S5) are known to be impaired by NBS and are described in more detail in the section "Establishing the antioxidant screening platform". The cluster analyses in Fig. 3 (c–e) and Supplementary Figure S6–S8 additionally provide a more detailed view of the dysregulation of dedicated genes in the pathways *Mismatch repair* (Fig. 3c), *Cell cycle* (Fig. 3d,e), *Glycolysis* (Supplementary Figure S6), *Oxidative phosphorylation* (Supplementary Figure S7) and *p53 signaling* (Supplementary Figure S8) between NBS and healthy states. The *Mismatch repair* (Fig. 3c) was predominantly down-regulated (6 of 7 genes) in the NBS-iPSCs. Impairment of *Mismatch repair* is associated with predisposition to cancer[37]. *Cell cycle* appears to shift from predominantly down-regulation (36 of 41 genes) in the NBS-fibroblasts to equal balance (16 genes down- and 12 genes up-regulated) in the NBS-iPSCs. Another shift along with the reprogramming took place from *Oxidative phosphorylation* in the NBS-fibroblasts to *Glycolysis* in the NBS-iPSCs (Supplementary Figures S6 and S7). This effect may have been induced by p53 down-regulation during reprogramming[38] and showed similarities to the Warburg effect in cancer cells which produced energy by Glycolysis[39]. In line with our previous publication[23], we found that essential genes such as TP53I3 in the p53 pathway shifted from up- to down-regulation during reprogramming. Supplementary Figure S8 depicts the *p53 signaling* in a comparison between NBS/WT PSCs and fibroblasts and demonstrates differences (including more down-regulation of the genes *P53* and *TP53I3)* in the only NBS line which could be reprogrammed (NBS8) compared to the other NBS lines.

Establishing the antioxidant screening platform. During cultivation of PSCs, which was performed at 5% oxygen levels, a temporary switch (12 h) to ambient (21%) oxygen resulted in apoptosis of NBS-iPSCs, but HFF1-iPSCs and hESCs were unaffected. We tested the effect of low (5%) and high (21%) oxygen quantitatively, by measuring gamma-H2AX, a marker for DNA double strand breaks in the presence or absence of the radiomimetic, Bleomycin. The result showed that low oxygen conditions greatly decreased the DNA damage under the influence of the mutagen. In addition, NBS Fibroblasts were more sensitive to DNA damage by mutagens than normal fibroblasts and the effect of low oxygen was less pronounced (Fig. 4a). NBS-iPSCs were then screened for various types of antioxidants to mimic or enhance the effect of low oxygen. The test measuring intracellular ROS levels revealed promising candidates to relieve NBS cells of oxidative stress, of which disulfiram (DSF) and EDHB were the most pronounced (Fig. 4b).

EDHB is utilized as a substrate analog and competitive inhibitor of prolyl 4-hydroxylases leading to specific inhibition of collagen synthesis[40] and to activation of the hypoxia inducible factor (HIF)[41]. We tested the effect of EDHB on intracellular ROS levels under stress conditions by supplementation with hydrogen peroxide (H_2O_2). EDHB decreased normal ROS levels and greatly decreased intracellular ROS levels in the presence of H_2O_2 in fibroblasts and PSCs (Fig. 4c,d). Interestingly, in fibroblasts, ROS levels were higher in EDHB treatment alone compared to treatment with EDHB plus hydrogen peroxide while in PSCs EDHB alone was lower than that of EDHB plus hydrogen peroxide treatment group. This effect was even more pronounced in NBS cells and needs further exploration. One possible explanation would be a change in ROS levels due to the shift in energy supply from oxidative phosphorylation to glycolysis along with the reprogramming which was described in the above paragraph "Global transcriptome analysis of NBS-iPSCs".

The effect of EDHB on DNA damage under stress conditions was also tested by supplementation with hydrogen peroxide or the radiomimetic Bleomycin. In fibroblasts, EDHB greatly decreased the DNA damage induced by hydrogen peroxide and moderately decreased the DNA damage caused by Bleomycin (Fig. 4e). In PSCs, EDHB decreased the DNA damage induced by H_2O_2 by 50%, but did not alter the DNA damage caused by Bleomycin (Fig. 4f). In addition, the same effects of EDHB, H_2O_2 and Bleomycin on DNA damage in PSCs could also be detected using western blotting (Fig. 4h).

As cells from patients with NBS are known to be affected by abnormal cell cycle checkpoints, e.g. failure of intra-S checkpoint after radiation[42], we determined the influence of DNA damage (administered by Bleomycin-treatment) and HIF-Pathway activation by EDHB on the status of CHEK1 and CHEK2, which are usually phosphorylated upon activation[43]. CHEK1 (S345) phosphorylation is mostly facilitated by ATR and required for the G2/M DNA damage checkpoint[44,45]. Upon DNA damage, CHEK1 becomes activated, it phosphorylates and inhibits CDC25C, thereby preventing activation of the cyclin B/CDK2 complex responsible for mitotic entry[46]. HFF1 cells only showed a slight increase in P-CHEK1 after treatment with bleomycin, but NBS-8 fibroblasts on the other hand exhibited CHEK1 activation without any treatment (Fig. 4g). This was decreased, but not eliminated upon treatment with EDHB. Bleomycin did not cause stronger activation than in control and this was not challenged by EDHB treatment. Interestingly, the supplementation with H_2O_2 completely diminished CHEK1 phosphorylation but not in the presence of EDHB.NBS iPSCs were found to have slightly higher base-level of phosphorylated CHEK1 in comparison to unaffected PSCs (H1, Fig. 4h). Treatment with EDHB reduced P-CHEK1 levels in NBS iPSCs under radiomimetic stress (Bleomycin) conditions, compared to the control.

CHEK2 is known to be phosphorylated (T68) and activated in an ATM-dependent manner in response to ionizing radiation[47]. In our NBS-context it is essential that the MRN complex regulates the activation of ATM[43,48] and

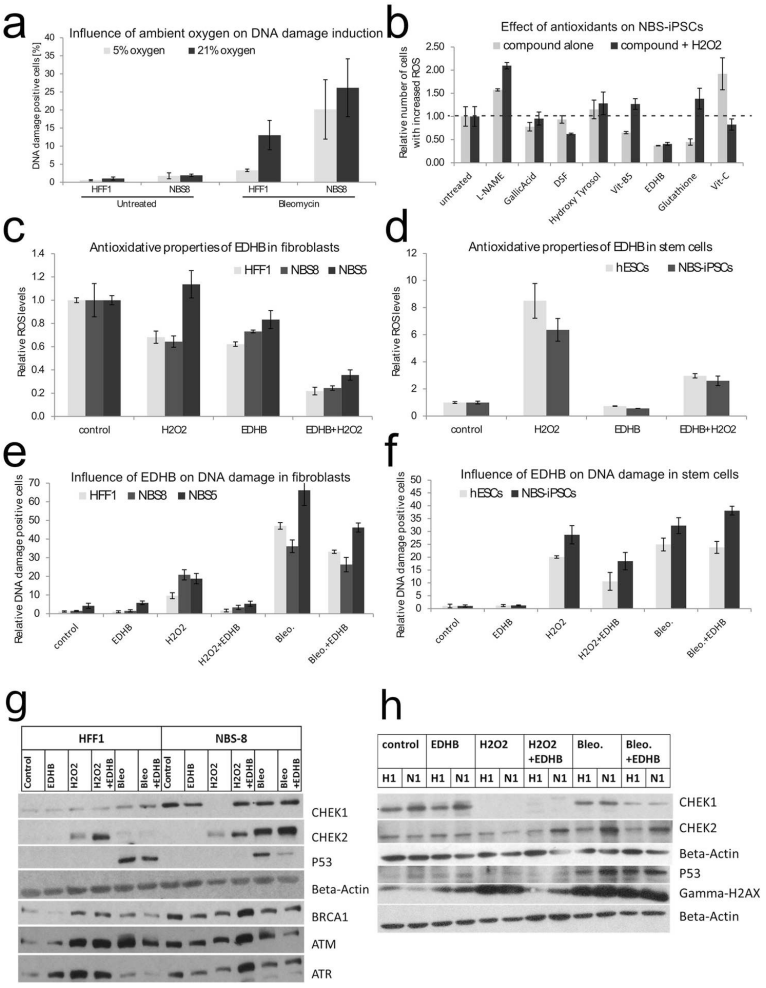


Figure 4. Response of NBS cells to oxidative stress and antioxidants. (**a**) The abundance of DNA damage measured by FACS-based detection of the DNA double-strand marker gamma-H2AX in HFF1 and NBS8 fibroblasts. DNA damage was induced by 30 µg/ml Bleomycin and compared under ambient (21%) and physiological (5%) oxygen concentrations. (**b**) NBS8-iPSCs were treated with either H_2O_2, several compounds known to influence DNA repair and ROS levels, or both. Internal ROS levels were then measured by FACS-based detection of the fluorescent ROS marker DCF-DA. The results were normalized to the untreated or peroxide-alone treated conditions respectively. (**c**) The influence of EDHB on internal ROS levels was tested on control (HFF1) and patient fibroblasts (NBS5, NBS8). The cells were either treated with EDHB alone or in combination with H_2O_2 to stimulate oxidative stress conditions. (**d**) Same experiment as in (**c**), but comparing control (hESCs) and patient (NBS-iPSCs) pluripotent stem cells. (**e**) The influence of EDHB on DNA damage (by detection of gamma-H2AX) was tested on control (HFF1) and patient fibroblasts (NBS5, NBS8). The cells were either treated with EDHB alone, in combination with H_2O_2 to stimulate oxidative stress conditions, or in combination with Bleomycin to stimulate mutagenic stress conditions. (**f**) Same experiment as in (**e**) but comparing control (hESCs) and patient (NBS-iPSCs) pluripotent stem cells. Bars indicate SD between independent experiments ($n = 3$). (**g**,**h**) Influence of DNA damage and EDHB on phosphorylation of DNA damage signaling proteins. Cells were treated with EDHB (antioxidant and inducer of HIFpathway), hydrogen peroxide (H_2O_2) and radiomimetic bleomycin (Bleo). (**g**) Immunofluorescent detection of phosphorylated signaling proteins in fibroblasts (HFF1, NBS-8) after SDS-PAGE. (**h**) Immunofluorescent detection of phosphorylated signaling proteins in hESCs (H1) and NBS-8-iPSCs (N1) after SDS-PAGE. Each lane of b-Actin corresponds to the lanes directly above and b-Actin is always unphosphorylated. For the sake of better readability western blots were cropped.

acts upstream as well as downstream of ATM[49]. Activated CHEK2 phosphorylates P53 at serine-20[50] CDC25A at serine-123[51] and CDC25C at serine-216 thus, contributing to the G1/S, S, and G2/M checkpoints respectively[52]. In HFF1 cells treated with H_2O_2, CHEK2 became phosphorylated and this increased after combined application of H_2O_2 and EDHB (Fig. 4g). The same effect was observed in NBS-8 fibroblasts. Bleomycin did not activate CHEK2 in HFF1 cells, but the activation was strong in NBS-8 cells. Here, a low base-level of P-CHEK2 in PSCs was observed (Fig. 4h). Upon DNA damage, P-CHEK2 activation was significantly high in NBS-iPSCs (N1) but decreased after treatment with EDHB (Fig. 4h).

Previous experiments using FACS analysis revealed that EDHB can decrease DNA damage caused by H_2O_2 in fibroblasts and iPSCs, but only moderately reduce the damage in fibroblasts caused by bleomycin (data not shown). Western Blot analysis of gamma-H2AX after treating hESCs and NBS-8-iPSCs with H_2O_2 and bleomycin confirmed these measurements (Fig. 4h). It also showed that the application of H_2O_2 can indeed result in DSBs as indicated by the increased detection of gamma-H2AX. But it is important to keep in mind that in comparison (γ-H2AX measurement, (Fig. 4h), H_2O_2 induced approx. the same level of DNA damage in hESCs as bleomycin, but 2-fold lower DNA damage in fibroblasts than in ESCs. Here, DNA damage was induced by oxidative stress in the form of H_2O_2 and by the DSB-inducer bleomycin (Fig. 4g,h).

This study shows, that in HFF1 cells, P53 S15 phosphorylation was observed after treatment with bleomycin, but not after H_2O_2 administration. In addition, the level of induction was lowered by addition of EDHB in the bleomycin treatment. In NBS-8 cells, the same effect was observed, but the P53 activation by bleomycin was lower and was almost abolished after EDHB treatment (Fig. 4g). In hESCs and NBS-8-iPSCs P53 was only phosphorylated at S15 after bleomycin treatment as well. But here, P-P53 was higher in NBS-8-iPSCs and EDHB treatment did not show a clear effect (Fig. 4h).

ATR is activated by ssDNA that result at a later stage in homologous recombination repair (HRR), or result from stalled replication forks. In NBS, repair from HR is impaired, so ATR signaling is rather activated by stalled replication forks. In NBS-8 cells, phosphorylated ATM exhibited a similar level without or after treatment with H_2O_2 or bleomycin (Fig. 4g). EDHB on the other hand increased the signal in cells treated with H_2O_2 (oxidative stress) and decreased the signal in cells treated with bleomycin (DSB inducer). Control cells, HFF1, exhibited lower basal levels of P-ATM than in NBS-8 cells but got strongly activated after H_2O_2 and bleomycin treatment. EDHB decreased the ATM activation by bleomycin as well. In contrast to ATM, ATR was not activated in HFF1 and NBS-8 cells by bleomycin. It was phosphorylated after treatment with EDHB or H_2O_2 in HFF1 cells, but not in NBS-8 cells. Here it only became activated after treatment with H_2O_2 and EDHB together. Comparison of ATM activation with activation of its target CHEK2, did not show the expected similar expression level in the western blot, neither did CHEK1 as target of ATM.

Here, BRCA1 S1524 phosphorylation appeared on a similar level of activation as ATR in NBS-8 cells, but different in HFF1 and also different in both cases in comparison to ATM (Fig. 4g). In detail, BRCA1 was slightly activated by H_2O_2 and bleomycin in HFF1 cells with no difference after addition of EDHB in any case. Again, there was a high level of phosphorylated protein in the control in NBS-8 cells, which was only further raised after treatment with H_2O_2 and EDHB. EDHB alone, H_2O_2 alone, bleomycin alone and bleomycin plus EDHB did show a similar expression of P-BRCA1, which was lower than in control.

The important finding here is the ability of EDHB to decrease the amount of DSBs caused by oxidative stress (administered by H_2O_2). EDHB showed a reduction of P-CHEK1 and P-CHEK2 in hESCs and iPSCs after bleomycin treatment, and an induction of P-CHEK1 and P-CHEK2 in fibroblasts after treatment of H_2O_2. Interestingly, the important tumor suppressor TP53 was less activated in NBS-8 fibroblasts as in control cells (HFF1) and exhibited an even lower signal after treatment with bleomycin.

The pattern in activation of P53, CHEK1 and CHEK2 is similar, but not identical in NBS-8 fibroblasts and iPSCs and differs in comparison to their healthy counterparts. The most prominent difference is the response of CHEK1 to EDHB in NBS-8 fibroblasts in comparison to NBS-8-iPSCs (activated in response to H_2O_2 only in the presence of EDHB). Furthermore, the relative activation of CHEK2 to bleomycin is a lot stronger in NBS-8 fibroblasts compared to iPSCs.

Discussion

The DNA damage sensing *NBN* is an adapter protein which can bind to a variety of other DNA signaling and repair proteins particularly ATM, which is a kinase that amplifies and transduces the DNA damage signal[53]. The 657del5 mutation in NBN results in a truncated protein where one specific functional domain (FHA-BRCT) is missing. This domain is also a common motif within other DNA repair signaling proteins[54]. DNA repair mechanisms especially those of repair pathway decisions are still not fully understood. NBS-iPSCs and their differentiated descendants could therefore serve as a good model to study DNA repair and cell fate after DNA damage. This could aid in elucidating the mechanisms underlying the disease. NBS-iPSCs can also provide a screening system for treatments which might increase the life span and quality of life of patients with NBS and similar diseases like Fanconi Anemia (FA), Ligase IV (LIG4) syndrome, Bloom syndrome, NBS-like disorder, ataxia-telangiectasia-like disorder (ATLD), Nonhomologous end-joining factor 1 (NHEJ1) syndrome and Seckel syndrome, which all derive from mutated genes in repair pathways[55,56]. In a recent publication, we modeled and characterized NBS by reprogramming[23]. Reprogrammed cells from patients with similar diseases like FA have been reported, though this can be done only after genetic correction or with the aid of antioxidants[24]. In another study, a mutation in the repair pathway gene ERCC6 did not prevent genetic reprogramming but exhibited elevated cell death rates and ROS production[25].

Our cellular NBS model was based on fibroblasts from NBS patients reprogrammed into iPSCs, using retroviral transduction of OCT4, SOX2, KLF4 and C-MYC. Further, by employing somatic cells and iPSCs of NBS, global transcriptome analysis was performed, to identify new phenotypes and changes in the signaling network of NBS cells compared to normal cells. In addition, the influence of oxidative stress, radiomimetics and antioxidants

was tested on the genomic integrity of NBS cells before and after reprogramming. Comparative transcriptome and associated pathway analyses revealed that, (a) NBS fibroblasts have a higher impact on cell cycle regulation, apoptosis and P53 signaling than normal fibroblasts (b) NBS-iPSCs and normal hESCs presented de-regulated genes and pathways associated with DNA replication, glycolysis, pyrimidine, fructose and mannose metabolism as well as DNA repair related pathways. Notably, these pathways can be connected to ROS homeostasis. Comparative tests based on sensitivity towards oxidative stress and DNA damaging agents such as hydrogen peroxide and Bleomycin, revealed that NBS-iPSCs and NBS-fibroblasts compared to normal fibroblasts were highly sensitive to DSB inducer Bleomycin and oxidative stress induced by exogenous hydrogen peroxide. Interestingly, DNA damage from hydrogen peroxide was efficiently relieved by addition of EDHB, an inducer of the hypoxia (HIF) pathway. The results indicate that NBS-iPSCs can serve as an excellent model to study NBS and screen for antioxidants *in vitro*.

NBS is a disease of premature aging resulting from the genomic instability caused by the *NBN* mutation which leads to hurdles in the reprogramming process. Activation of P53 is especially known to restrain reprogramming[28]. We observed that P53, a known tumor-suppressor gene, was activated in NBS fibroblasts, resulting in increased senescence in the NBS cell cultures and had extremely low reprogramming efficiency. Once reprogrammed, the cells maintained pluripotency and proliferated like normal hPSCs. NBS-iPSCs may protect themselves from oxidative stress and ROS-induced DNA damage by increased glycolysis which was up-regulated in comparison to hESCs. In previous studies, hESCs were found to have immature mitochondria and depended heavily on glycolysis[13, 15, 35, 57]. This bias towards glycolysis might be related to the down regulation of P53, increased stress in the NBS iPSCs and hence increased glycolytic lactate production for survival. Furthermore, it is known that P53 promotes oxidative phosphorylation[38], thus, reduced P53 results in reduced oxidative phosphorylation.

Also, PSCs are known to ensure genomic integrity through enhanced apoptosis induction and increased antioxidant defense, contributing to protection against DNA damage[58]. The finding that antioxidants, particularly EDHB, improved genomic stability of NBS-iPSCs can improve reprogramming of additional NBS fibroblasts and other diseases like NBS which derive from mutated genes in DNA repair pathways, examples include, Fanconi Anemia (FA)[16], Ligase IV (LIG4) syndrome[17], Bloom syndrome[18], NBS-like disorder[19], Ataxia-Telangiectasia-Like Disorder (ATLD)[20], Nonhomologous end-joining factor 1 (NHEJ1) syndrome[21] and Seckel Syndrome[22].

EDHB was used in another study to protect cells from hypoxia-mediated oxidative damage[59]. With EDHB known as an activator of the HIF pathway, these results point to a reduction of ROS-induced DNA damage and subsequent relief of the impaired DNA damage response as the cause for genomic stabilization.

DNA damage response mediated by the MRN complex, ATM/ATR, P53, CHEK1 and CHEK2 are crucial in early development of most types of cancer[60]. Although the role of this core network in relation to DNA damage, cancer and pluripotency has been widely investigated[31, 43], several mechanisms in early oncogenesis remain unclear.

In this study, we have demonstrated that our model of fibroblasts and iPSCs derived from NBS patients besides the study of NBS itself which is associated with microcephaly, premature aging and growth retardation provides the environment for a detailed study of oncogenic mechanisms. The NBS phenotype includes a predisposition to cancer due to impaired DNA damage repair. Furthermore, we have shown that the addition of stimuli such as oxidative stress and mutagenic factors to this model could be used as a screening platform for anti-oxidants capable of suppressing DNA damage. Transcriptome analysis of our NBS-model identified de-regulation of P53, cell cycle, oxidative phosphorylation and glycolysis. In screening for antioxidants we identified EDHB as a potent modulator of DNA damage. Interestingly we revealed that NBS fibroblasts have a higher susceptibility for induction of DNA damage compared to NBS-iPSCs. However, although additional research is needed to improve the reprogramming efficiency and thus the robustness we believe that NBS-iPSCs can serve as cellular tools for a screening platform for molecules with anti-oxidant capabilities.

Methods

Ethical approval. NBS patient dermal fibroblast cells with informed consent (Table 1) were provided by Prof. Dr. Karl Sperling (Institute of Medical and Human Genetics, Charité - Universitätsmedizin Berlin, 13353 Berlin). Approval was obtained from the Ethics Commission of the Charité—Universitätsmedizin. The methods and experimental protocols were carried out in accordance with their guidelines and regulations.

Cell culture. Neonatal foreskin fibroblasts, HFF1 and BJ were purchased from ATCC (#SCRC-1041 and #CRL-2522, respectively). All cells used were cultured at 37 °C, 5% CO_2 and either 21% (standard) or 5% oxygen in an incubator under humidified atmosphere. Somatic cells were cultured in DMEM medium (Gibco, USA) supplemented with 10% fetal bovine serum (FBS) and 1x penicillin/streptomycin until reaching 90% confluency and then split in a 1:4 ratio. The conditions for passaging human pluripotent stem cells (hPSC) were a combination of methods adapted from several published protocols[61, 62]. This was applied to the culture of the human ESC-lines H1 and H9 (WiCell Research Institute, Madison, WI, USA) and iPSCs generated from NBS and HFF1 cells. In combination with MEFs, hPSCs were usually cultivated in plates coated with 0,2% gelatin and fed with hESC medium containing KO-DMEM supplemented with 20% knockout serum replacement, non-essential amino acids, L-glutamine, penicillin/streptomycin, sodium pyruvate, 0.1 mM beta-mercaptoethanol and 4 ng/ml FGF-2, which was replaced every second day.

FACS analysis (detection of ROS and DNA damage). The FACSCalibur system (BD Biosciences, USA) and the software program CellQuestPro were used as described by the manufacturer's instructions. Programs

used for data analysis were CellQuestPro (BD Biosciences, USA), Cyflogic (Cyflo Ltd, Finland), Weasel 3.0 (WEHI, Australia) and Flowing software 2.5.1 (Finland) programs.

For ROS measurement, fibroblast cells were seeded onto 12-well-plates with a density of 5×10^4 cells per well one day prior to treatment. hESCs and iPSCs were seeded on Matrigel (Corning) in 6-well-plates and fed with hESC medium, one week prior to the treatment. To prepare cells for ROS-measurements, they were washed once with PBS and then incubated in 15 µM DCF-DA for 20 min at 37 °C. Afterwards, the solution was removed and the cells were briefly rinsed with PBS. Cells were treated with different concentrations of antioxidants (as indicated) and/or 20 µM H_2O_2 for 30 min. To analyze single cells by FACS, they were trypsinized by colorless 0.05% trypsin solution for 5 min. Trypsinization was stopped by adding 10% FBS in PBS. Cells were then centrifuged by 500 × g for 5 min and re-suspended in 300 µl PBS. The fluorescence was measured by FACS using the FITC channel.

For measurement of DNA damage, fibroblast cells were seeded onto 6-well-plates with a density of 4×10^5 cells per well one day prior to treatment. hESCs and iPSCs were seeded on Matrigel (Corning) in 6-well-plates and fed with hESC medium, one week prior to the treatment. Cells were either treated with antioxidants (as indicated) 5 min prior to the addition of 1 mM H_2O_2 or with H_2O_2 alone at a total incubation time of 4 h at 37 °C in a cell culture incubator with either 21% or 5% oxygen, as indicated. Other cells were treated with antioxidants (as indicated) 5 min prior to the addition of 30 µg/ml Bleomycin for 3 h, and released for 1 h by switching to Bleomycin-free medium. Afterwards, the cells were briefly rinsed with PBS and trypsinized to generate single cells. The cells were centrifuged at 500 × g for 5 min and the cell pellet was re-suspended in 100 µl PBS. Under constant shaking (to prevent clumping) 300 µl of 100% ice-cold ethanol was added dropwise to fix the cells and incubated at −20 °C for at least 30 min or until further use. Afterwards the solution was mixed with 1 ml PBS and centrifuged at 2200 × g for 5 min. The pellet was re-suspended in 50 ml PBS-T with 5% FBS and incubated 30 min at RT for blocking. FITC-labeled gamma-H2AX antibody (Millipore, 1:500) was added and incubated overnight at 4 °C. The next day, 300 µl PBS was added and the cells were measured by FACS using the FITC channel. In some cases, the cells were co-stained with TRA1-60 antibody (Santa Cruz Biotechnology, Inc.) to verify pluripotent cell populations.

Western Blot. The membrane was rinsed with dH_2O and then blocked with 5% milk powder or 5% BSA in PBS-T (blocking solution) by shaking for 1 h at RT. BSA blocking solution was used for phospho-specific antibodies, in all other cases blocking was performed with milk. After blocking, the membrane was incubated by shaking at 4 °C overnight with the primary antibodies dissolved in PBS-T with 5% milk powder or 5% BSA; Beta-actin (Sigma-Aldrich), phospho-histone H2A.X (Ser139), phospho-CHEK1 and phospho-CHEK2 (CST), phospho-P53 (Ser15, CST), phospho-BRCA1 (Ser1524, CST), phospho-ATM (Ser1981, CST), phospho-ATR (Ser428, CST). Afterwards, the membrane was washed 3 times for 10 min in PBS-T on the Lab shaker, exchanging buffer between each step. Then, the secondary antibody dissolved in milk or BSA blocking solution was applied by shaking for 1 h at RT. Afterwards the membrane was washed 3 times for 10 min in PBS-T. Appropriate peroxidase-conjugated secondary antibodies and luminescence was induced by ECL Plus Western Blotting Detection Reagents and captured on BioMAX XAR film.

Transcriptomics. The microarray hybridization experiments included biotin-labelling of cRNA by using 500 ng quality-checked total RNA (per sample) as input. Chip hybridizations, washing, Cy3 streptavidin staining, and scanning were performed on BeadStation 500 platform (Illumina) using reagents and protocols supplied by the manufacturer. cRNA samples were hybridized in duplicates on Illumina human-8 BeadChips (NBS-1, NBS-3, NBS-5, NBS-7, HFF1, BJ) or Illumina human-12 BeadChips (NBS-8, NBS-8 iPSCs, H1, H9 (single), HFF1-iPSCs, BJ-iPSCs). Basic expression data analysis was carried out using the manufacturer's software GenomeStudio (Illumina). Raw data was background-subtracted and normalized using the "rank invariant" algorithm. Normalized data was then filtered for significant expression (detection p-value) based on negative control beads. All genes with detection p-values below 0.01 were considered as expressed. All genes with differential p-values below 0.05 were considered as differentially expressed. Selection for differentially expressed genes was performed on the basis of arbitrary thresholds (1.5 fold changes) and statistical significance according to an Illumina custom model[63]. Different sets of gene lists were entered into the DAVID functional annotation tool[34], using the official gene symbol or ILLUMINA-IDs as input, to perform gene-annotation enrichment analysis, functional annotation clustering, KEGG pathway mapping (http://www.genome.jp/kegg/)[36], transcription factor binding site prediction and more. Multiple testing was assessed via the Benjamini-Hochberg correction in the results of the DAVID analysis.

For the calculation of the activation state of transcription factors, a list of differentially regulated genes between NBS and normal fibroblasts was used as input for Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity).

Microarray data is available at NCBI GEO under the accession number GSE94708 for the superseries and GSE94706 for the fibroblasts and GSE94707 for the iPSCs series.

References

1. Weemaes, C. M. *et al.* A new chromosomal instability disorder: the Nijmegen breakage syndrome. *Acta Paediatr. Scand.* **70**, 557–564 (1981).
2. Nijmegen breakage syndrome. The International Nijmegen Breakage Syndrome Study Group. *Arch. Dis. Child.* **82**, 400–406 (2000).
3. Varon, R. *et al.* Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* **93**, 467–476 (1998).
4. Difilippantonio, S. *et al.* Distinct domains in Nbs1 regulate irradiation-induced checkpoints and apoptosis. *J. Exp. Med.* **204**, 1003–1011 (2007).

5. Deriano, L., Stracker, T. H., Baker, A., Petrini, J. H. J. & Roth, D. B. Roles for NBS1 in alternative nonhomologous end-joining of V(D)J recombination intermediates. *Mol. Cell* **34**, 13–25 (2009).
6. Varon, R. *et al.* Clinical ascertainment of Nijmegen breakage syndrome (NBS) and prevalence of the major mutation, 657del5, in three Slav populations. *Eur. J. Hum. Genet. EJHG* **8**, 900–902 (2000).
7. Maser, R. S., Zinkel, R. & Petrini, J. H. An alternative mode of translation permits production of a variant NBS1 protein from the common Nijmegen breakage syndrome allele. *Nat. Genet.* **27**, 417–421 (2001).
8. Williams, G. J., Lees-Miller, S. P. & Tainer, J. A. Mre11-Rad50-Nbs1 conformations and the control of sensing, signaling, and effector responses at DNA double-strand breaks. *DNA Repair* **9**, 1299–1306 (2010).
9. Cilli, D. *et al.* Identification of the interactors of human nibrin (NBN) and of its 26 kDa and 70 kDa fragments arising from the NBN 657del5 founder mutation. *PloS One* **9**, e114651 (2014).
10. Turrens, J. F. Mitochondrial formation of reactive oxygen species. *J. Physiol.* **552**, 335–344 (2003).
11. Ray, P. D., Huang, B.-W. & Tsuji, Y. Reactive oxygen species (ROS) homeostasis and redox regulation in cellular signaling. *Cell. Signal.* **24**, 981–990 (2012).
12. Folmes, C. D. L., Dzeja, P. P., Nelson, T. J. & Terzic, A. Metabolic plasticity in stem cell homeostasis and differentiation. *Cell Stem Cell* **11**, 596–606 (2012).
13. Prigione, A., Fauler, B., Lurz, R., Lehrach, H. & Adjaye, J. The senescence-related mitochondrial/oxidative stress pathway is repressed in human induced pluripotent stem cells. *Stem Cells* **28**, 721–33 (2010).
14. Mathieu, J. *et al.* Hypoxia-inducible factors have distinct and stage-specific roles during reprogramming of human cells to pluripotency. *Cell Stem Cell* **14**, 592–605 (2014).
15. Prigione, A. *et al.* HIF1α modulates cell fate reprogramming through early glycolytic shift and upregulation of PDK1-3 and PKM2. *Stem Cells Dayt. Ohio* **32**, 364–376 (2014).
16. Walden, H. & Deans, A. J. The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annu. Rev. Biophys* **43**, 257–278 (2014).
17. Chistiakov, D. A. Ligase IV syndrome. *Adv. Exp. Med. Biol.* **685**, 175–185 (2010).
18. Arora, H. *et al.* Bloom syndrome. *Int. J. Dermatol.* **53**, 798–802 (2014).
19. Waltes, R. *et al.* Human RAD50 deficiency in a Nijmegen breakage syndrome-like disorder. *Am. J. Hum. Genet.* **84**, 605–616 (2009).
20. Taylor, A. M. R., Groom, A. & Byrd, P. J. Ataxia-telangiectasia-like disorder (ATLD)-its clinical presentation and molecular basis. *DNA Repair* **3**, 1219–1225 (2004).
21. Buck, D. *et al.* Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* **124**, 287–299 (2006).
22. O'Driscoll, M., Gennery, A. R., Seidel, J., Concannon, P. & Jeggo, P. A. An overview of three new disorders associated with genetic instability: LIG4 syndrome, RS-SCID and ATR-Seckel syndrome. *DNA Repair* **3**, 1227–1235 (2004).
23. Halevy, T. *et al.* Chromosomal Instability and Molecular Defects in Induced Pluripotent Stem Cells from Nijmegen Breakage Syndrome Patients. *Cell Rep* **16**, 2499–2511 (2016).
24. Müller, L. U. W. *et al.* Overcoming reprogramming resistance of Fanconi anemia cells. *Blood* **119**, 5449–5457 (2012).
25. Andrade, L. N., de, S., Nathanson, J. L., Yeo, G. W., Menck, C. F. M. & Muotri, A. R. Evidence for premature aging due to oxidative stress in iPSCs from Cockayne syndrome. *Hum. Mol. Genet* **21**, 3825–3834 (2012).
26. Bürger, S. *et al.* Radiation-induced DNA damage and repair in peripheral blood mononuclear cells from Nijmegen breakage syndrome patients and carriers assessed by the Comet assay. *Environ. Mol. Mutagen.* **47**, 260–270 (2006).
27. Tang, H. *et al.* Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth. *Cell Stem Cell* **18**, 587–590 (2016).
28. Marión, R. M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460**, 1149–1153 (2009).
29. Lombard, D. B. *et al.* DNA repair, genome stability, and aging. *Cell* **120**, 497–512 (2005).
30. Mlody, B. & Adjaye, J. Generation of iPSC lines from a Nijmegen Breakage Syndrome patient. *Stem Cell Res* **15**, 629–632 (2015).
31. Spike, B. T. & Wahl, G. M. p53, Stem Cells, and Reprogramming: Tumor Suppression beyond Guarding the Genome. *Genes Cancer* **2**, 404–419 (2011).
32. Chappell, J. & Dalton, S. Roles for MYC in the Establishment and Maintenance of Pluripotency. *Cold Spring Harb. Perspect. Med* **3**, a014381–a014381 (2013).
33. Babaie, Y. *et al.* Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. *Stem Cells Dayt. Ohio* **25**, 500–510 (2007).
34. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1–13 (2009).
35. Varum, S. *et al.* Energy Metabolism in Human Pluripotent Stem Cells and Their Differentiated Counterparts. *PLoS ONE* **6**, e20914 (2011).
36. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353–D361 (2017).
37. Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **18**, 85–98 (2008).
38. Kruiswijk, F., Labuschagne, C. F. & Vousden, K. H. p53 in survival, death and metabolic health: a lifeguard with a licence to kill. *Nat. Rev. Mol. Cell Biol.* **16**, 393–405 (2015).
39. Warburg, O. On the origin of cancer cells. *Science* **123**, 309–314 (1956).
40. Sasaki, T., Majamaa, K. & Uitto, J. Reduction of collagen production in keloid fibroblast cultures by ethyl-3,4-dihydroxybenzoate. Inhibition of prolyl hydroxylase activity as a mechanism of action. *J. Biol. Chem* **262**, 9397–9403 (1987).
41. Bruick, R. K. & McKnight, S. L. A conserved family of prolyl-4-hydroxylases that modify HIF. *Science* **294**, 1337–1340 (2001).
42. Yazdi, P. T. *et al.* SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. *Genes Dev* **16**, 571–582 (2002).
43. Marechal, A. & Zou, L. DNA Damage Sensing by the ATM and ATR Kinases. *Cold Spring Harb. Perspect. Biol* **5**, a012716–a012716 (2013).
44. Zhao, H. & Piwnica-Worms, H. ATR-mediated checkpoint pathways regulate phosphorylation and activation of human Chk1. *Mol. Cell. Biol.* **21**, 4129–4139 (2001).
45. Liu, Q. *et al.* Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. *Genes Dev.* **14**, 1448–1459 (2000).
46. Sanchez, Y. *et al.* Conservation of the Chk1 checkpoint pathway in mammals: linkage of DNA damage to Cdk regulation through Cdc25. *Science* **277**, 1497–1501 (1997).
47. Matsuoka, S., Huang, M. & Elledge, S. J. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science* **282**, 1893–1897 (1998).
48. Schiller, C. B. *et al.* Structure of Mre11–Nbs1 complex yields insights into ataxia-telangiectasia–like disease mutations and DNA damage signaling. *Nat. Struct. Mol. Biol.* **19**, 693–700 (2012).
49. Stracker, T. H. & Petrini, J. H. J. The MRE11 complex: starting from the ends. *Nat. Rev. Mol. Cell Biol.* **12**, 90–103 (2011).
50. Shieh, S. Y., Ahn, J., Tamai, K., Taya, Y. & Prives, C. The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites. *Genes Dev.* **14**, 289–300 (2000).

51. Falck, J., Mailand, N., Syljuåsen, R. G., Bartek, J. & Lukas, J. The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. *Nature* **410**, 842–847 (2001).
52. Xu, X., Tsvetkov, L. M. & Stern, D. F. Chk2 activation and phosphorylation-dependent oligomerization. *Mol. Cell. Biol.* **22**, 4419–4432 (2002).
53. Kitagawa, R. & Kastan, M. B. The ATM-dependent DNA damage signaling pathway. *Cold Spring Harb. Symp. Quant. Biol.* **70**, 99–109 (2005).
54. Lloyd, J. *et al.* A Supramodular FHA/BRCT-Repeat Architecture Mediates Nbs1 Adaptor Function in Response to DNA Damage. *Cell* **139**, 100–111 (2009).
55. Lee, P. *et al.* SMRT compounds abrogate cellular phenotypes of ataxia telangiectasia in neural derivatives of patient-specific hiPSCs. *Nat. Commun.* **4**, 1824 (2013).
56. Tilgner, K. *et al.* A human iPSC model of Ligase IV deficiency reveals an important role for NHEJ-mediated-DSB repair in the survival and genomic stability of induced pluripotent stem cells and emerging haematopoietic progenitors. *Cell Death Differ* **20**, 1089–1100 (2013).
57. Prigione, A. & Adjaye, J. Modulation of mitochondrial biogenesis and bioenergetic metabolism upon *in vitro* and *in vivo* differentiation of human ES and iPS cells. *Int. J. Dev. Biol.* **54**, 1729–1741 (2010).
58. Dannenmann, B. *et al.* High Glutathione and Glutathione Peroxidase-2 Levels Mediate Cell-Type-Specific DNA Damage Protection in Human Induced Pluripotent Stem Cells. *Stem Cell Rep* **4**, 886–898 (2015).
59. Nimker, C., Kaur, G., Revo, A., Chaudhary, P. & Bansal, A. Ethyl 3,4-dihydroxy benzoate, a unique preconditioning agent for alleviating hypoxia-mediated oxidative damage in L6 myoblasts cells. *J. Physiol. Sci. JPS* **65**, 77–87 (2015).
60. Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864–870 (2005).
61. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
62. Du, Z.-W. & Zhang, S.-C. Lentiviral Vector-Mediated Transgenesis in Human Embryonic Stem Cells. In *Lentivirus Gene Engineering Protocols: Second Edition* (ed. Federico, M.) 127–134 (Humana Press, 2010).
63. Kuhn, K. *et al.* A novel, high-performance random array platform for quantitative gene expression profiling. *Genome Res.* **14**, 2347–2356 (2004).

Acknowledgements

JA acknowledges financial support from the Max Planck Society, Medical Faculty of Heinrich-Heine Universität-Düsseldorf, Germany and the BMBF grant number 01GN1005. We thank Elsie Amedonu for proof-reading the manuscript.

Author Contributions

B.M., W.W., S.M. and J.A. wrote the manuscript. K.S. provided the NBS fibroblasts. B.M. performed the experiments. W.W. and B.M. analyzed the data. J.A., K.S. and B.M. initiated and conceived this study.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-07905-2

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2.1.6 Human pluripotent stem cell derived HLC transcriptome data enables molecular dissection of hepatogenesis.

Sci Data. 2018 Mar 13;5:180035. doi: 10.1038/sdata.2018.35.

Wruck W(1), Adjaye J(1).

Author information: (1)Medical Faculty, Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, 40225 Düsseldorf, Germany.

Induced pluripotent stem cells (iPSCs) and human embryonic stem cells (hESCs) differentiated into hepatocyte-like cells (HLCs) provide a defined and renewable source of cells for drug screening, toxicology and regenerative medicine. We previously reprogrammed human fetal foreskin fibroblast cells (HFF1) into iPSCs employing an episomal plasmid-based integration-free approach, this iPSC-line and the hESC lines H1 and H9 were used to model hepatogenesis in vitro. Biochemical characterisation confirmed glycogen storage, ICG uptake and release, urea and bile acid production, as well as CYP3A4 activity. Microarray-based transcriptome analyses was carried out using RNA isolated from the undifferentiated pluripotent stem cells and subsequent differentiation stages- definitive endoderm (DE) hepatic endoderm (HE) and HLCs. K-means identified 100 distinct clusters, for example, POU5F1/OCT4 marking the undifferentiated stage, SOX17 the DE stage, HNF4α the HE stage, and ALB specific to HLCs, fetal liver and primary human hepatocytes (PHH). This data descriptor describes these datasets which should be useful for gaining new insights into the molecular basis of hepatogenesis and associated gene regulatory networks.

DOI: 10.1038/sdata.2018.35 PMCID: PMC5848791 PMID: 29533390 [Indexed for MEDLINE]

URL: https://www.nature.com/articles/sdata201835

Contribution: Wasco Wruck performed bioinformatic analyses and wrote the manuscript.

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

Data Descriptor: Human pluripotent stem cell derived HLC transcriptome data enables molecular dissection of hepatogenesis

Received: 4 October 2017
Accepted: 16 January 2018
Published: 13 March 2018

Wasco Wruck[1] **& James Adjaye**[1]

Induced pluripotent stem cells (iPSCs) and human embryonic stem cells (hESCs) differentiated into hepatocyte-like cells (HLCs) provide a defined and renewable source of cells for drug screening, toxicology and regenerative medicine. We previously reprogrammed human fetal foreskin fibroblast cells (HFF1) into iPSCs employing an episomal plasmid-based integration-free approach, this iPSC-line and the hESC lines H1 and H9 were used to model hepatogenesis *in vitro*. Biochemical characterisation confirmed glycogen storage, ICG uptake and release, urea and bile acid production, as well as CYP3A4 activity. Microarray-based transcriptome analyses was carried out using RNA isolated from the undifferentiated pluripotent stem cells and subsequent differentiation stages- definitive endoderm (DE) hepatic endoderm (HE) and HLCs. K-means identified 100 distinct clusters, for example, *POU5F1*/*OCT4* marking the undifferentiated stage, *SOX17* the DE stage, *HNF4α* the HE stage, and *ALB* specific to HLCs, fetal liver and primary human hepatocytes (PHH). This data descriptor describes these datasets which should be useful for gaining new insights into the molecular basis of hepatogenesis and associated gene regulatory networks.

Design Type(s)	transcription profiling by array design • cell type comparison design • replicate design
Measurement Type(s)	transcription profiling assay
Technology Type(s)	DNA microarray
Factor Type(s)	biological replicate
Sample Characteristic(s)	Homo sapiens • embryonic stem cell • definitive endoderm • ESC derived cell line • liver endoderm • hepatocyte • liver • skin of prepuce of penis • induced pluripotent stem cell line cell • iPSC derived cell line

[1]Medical Faculty, Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, 40225 Düsseldorf, Germany. Correspondence and requests for materials should be addressed to J. A. (email:James.Adjaye@med.uni-duesseldorf.de).

Background & Summary

The implementation of a well-characterized renewable source of hepatocytes differentiated from iPSCs and hESCs provides a powerful *in vitro* model system for analysing the molecular mechanisms associated with hepatogenesis. Several essential initiators of hepatogenesis such as fibroblast growth factor 2 and 4 (FGF4 and FGF2)[1,2], bone morphogenic protein (BMP2)[3], hepatocyte growth factor (HGF), oncostatin M and dexamethasone[4] have already been described. These factors are sequentially supplemented into the media during the course of the differentiation process.

Besides detoxification the liver is responsible for a number of essential functions e.g. the uptake and storage of glycogen, various metabolic functions, synthesis of bile acids and production of plasma proteins. Available liver cellular models have disadvantages: (i) liver biopsy derived primary human hepatocytes (PHH) cannot be expanded for long periods *in vitro*, are often obtained from diseased individuals and are difficult to obtain in sufficient quantities[5,6], (ii) transformed, permanent cell lines, such as HepG2 and HepaRG, have cancer phenotypes which are significantly diverged from normal primary hepatocytes[7–9]. A potential alternative could be the differentiation into hepatocyte-like cells. Although hepatocyte-like cells (HLCs) derived from iPSCs are not fully mature compared to liver biopsy derived adult hepatocytes they are endowed with many advantages, for example easily generated from iPSCs, known genetic background and disease states thus optimal for disease modelling *in vitro,* toxicology studies and drug screening. iPSC-based cellular models have already been employed in several studies for drug screening, toxicology studies and disease modeling[10–14].

The liver develops in a stepwise process *in vivo*: first, competence is established in the foregut endoderm in response to signals emanating from cardiac mesoderm, thereafter liver-specific gene expression is initiated[15]. The differentiation of hiPSCs and hESCs also proceed via the intermediate step of definitive endoderm, the bipotential hepatic endoderm, then maturation into HLCs[16]. Distinct stage specific changes in the associated transcriptional regulatory networks control the different phases of hepatogenesis[17]. Wang *et al.* describe a developmental progression from unmarked chromatin to poised chromatin and then to histone H3K27 acetylation which is accompanied by specific transcription factor classes[18]. They suggest FOXA transcription factors - known as pioneer factors facilitating the unwinding of chromatin - to play a role at poised enhancers while lineage-specific factors such as PDX1 for pancreatic and HNF4α for hepatic lineage drive the poised to an active enhancer state[18].

Attaining maturation comparable to primary hepatocytes is still one of the most challenging issues associated HLC differentiation. Knowledge on HNF4α, as major transcription factor regulating hepatic differentiation and maturation has already been described[15]. Additionally Li *et al.* reported that HNF4α lies upstream of the transcription factors HNF1α and PXR suggesting it might initiate a cascade of gene regulatory networks driving hepatocyte differentiation[15]. In our publication related to the hESCs and hiPSC dataset pertinent to this data descriptor we confirmed expression of maturation markers such as *ALB, HNF4α, HNF1α* and *TTR*[16,17].

The data described consists of microarray gene expression data from hESCs and hiPSCs differentiated into HLCs via the DE and HE stages and also fetal liver and primary human hepatocyte samples as reference. Although transcription factors central to hepatogenesis have been described, the datasets described here will enable a more detailed analyses of gene regulatory networks associated with modelling hepatogenesis using pluripotent stem cells.

Methods

Human ES and iPS cells culture

Human ES cell lines H1 and H9 (WiCell Research Institute, Madison, Wisconsin) from passage 39 to 66 were maintained under sterile conditions in a humidified incubator in a 5% CO2-95% air atmosphere at 37 °C (INNOVA CO-170 Incubator, New Brunswick Scientific). In a routine culture cells were maintained on Matrigel® in conditioned media (CM)[19]. Under these culture conditions, hESCs were confirmed to stain positive for OCT4, SSEA-4, TRA-1-60, and TRA-1-81 (ES Cell Characterization Kit, Chemicon). Before initiating the differentiation cells were washed with PBS without Ca2+Mg2+ (Gibco, Invitrogen).

Cell culture of iPS cells which were derived from Human neonatal foreskin fibroblasts HFF1 is described in Matz *et al.*[17].

Differentiation into hepatocyte-like cells (HLCs)

The derivation of HLCs from the hESC lines H1 and H9[16] followed protocols described by Hay *et al.*[20] and Agarwal *et al.*[21]. RNA samples were extracted after each step of the differentiation protocol.

Differentiation of iPS cells into HLCs[17] followed in large parts the protocol described by Jozefczuk *et al.*[16].

The overall experimental design of this study is illustrated in Figure 1a. Two pluripotent stem cells lines (hESC- H1 and H9) and fetal foreskin derived iPSC were used. Both proceeded via the intermediate DE and HE stages to HLCs and then compared to commercially bought RNA from fetal liver (Stratagene, MVP Total RNA: tissue from single male donor, 18th week of gestation; positive control for the iPSC-based differentiations: Clontech, #636540) and adult liver biopsy-derived primary human hepatocytes-PHH (Ready Heps Fresh Hepatocytes; Lonza, 65-year old male of Asian origin; positive control for the iPSC-based differentiations: Clontech, #636531).

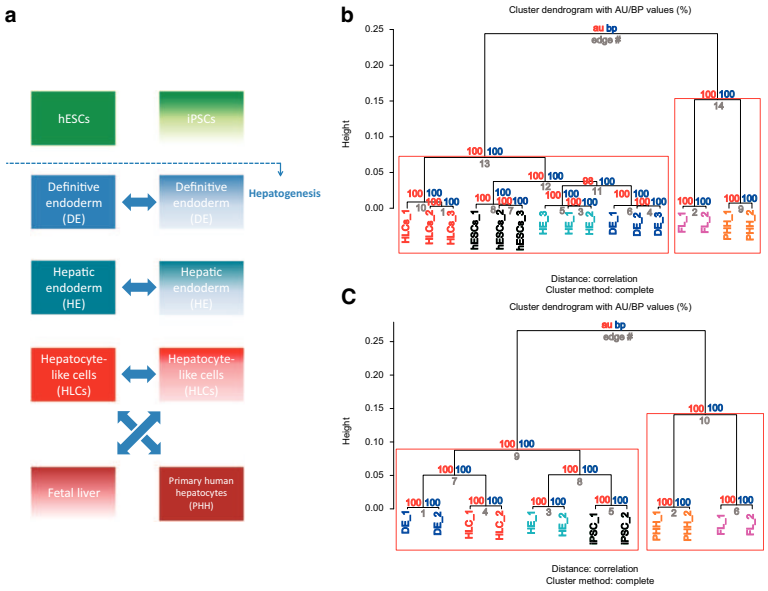


Figure 1. Comparison of hepatic differentiation of iPSCs and hESCs. (**a**) Scheme of comparative hepatic differentiation of iPSCs and hESCs. iPSCs and hESCs are differentiated into HLCs which can be compared versus each other and versus fetal liver (FL) and PHHs. Also the intermediate stages DE and HE are captured and thus can be subjected to comparative analysis. Hierarchical clustering of hESC (**b**) and iPSC (**c**) differentiation into HLCs was performed via pvclust using 1000 bootstrap samples. „au" (approximately unbiased) is computed by multiscale bootstrap resampling and „bp" (bootstrap probability) by normal bootstrapping. Red rectangles mark clusters with AU larger than 95%. Thus the dendrogram is with one exception in hESC (98%) at 100% supported by data. All replicates cluster together. HLCs cluster apart from hESCs/iPSCs, DE and HE. Fetal liver and PHH cluster together and separated from the hESC/iPSC-derived hepatic differentiation stages.

Illumina BeadChip hybridisation

Biotin-labelled cRNA was produced by means of a linear amplification kit (Ambion, Austin, TX, USA) using 500 ng of quality-checked total RNA as input. Chip hybridisations, washing, Cy3-streptavidin staining, and scanning were performed on an Illumina BeadStation 500 platform (Il-lumina, San Diego, CA, USA) using reagents and following protocols supplied by the manufacturer. cRNA samples were hybridised in biological triplicates on HumanRef-8 Expression BeadChips. The following samples were hybridized: Undifferentiated cells (H1 cell line), DE (definitive endoderm)-differentiated cells, HE (hepatic endoderm)-differentiated cells and hepatocyte-like cells (HLCs) derived with two independent protocols[20,21].

Differentiation experiments of iPSCs were hybridised on Illumina HumanHT-12 BeadChips. For details see the Methods description in Matz *et al.*[17].

The summary of bead-level data to bead-summary data was carried out using the manufacturer's software BeadStudio 3.0 (Illumina) for hESC and iPSC differentiation experiments. Table 1 provides an overview of all samples used for this study.

Data analysis and statistical methods

For further analysis, the bead-summary data saved in the BeadStudio was imported into the Bioconductor environment[22] and quantile normalized using the bioconductor package lumi[23]. Global gene expression similarities within biological replicates and between dedicated differentiation stages, pairwise Pearson correlation coefficients were calculated for all samples. Cluster analyses were performed using the R/Bioconductor environment[22] and the package pvclust[24] using n = 1000 for bootstrap sampling. k-means clustering was employed to identify clusters of genes with similar gene expression changes over the stages of the differentiation protocol using k = 100 as number of clusters. The software is available in the Supplementary Data File 1.

ID	description	replicate [#]	NCBI GEO sample	NCBI GEO accession no.
hESCs_1	human embryonic stem cells (H1)	1	GSM2683216	GSE100447
hESCs_2	human embryonic stem cells (H1)	2	GSM2683217	GSE100447
hESCs_3	human embryonic stem cells (H1)	3	GSM2683218	GSE100447
DE_hESCs_1	definite endoderm from hESCs	1	GSM2683219	GSE100447
DE_hESCs_2	definite endoderm from hESCs	2	GSM2683220	GSE100447
DE_hESCs_3	definite endoderm from hESCs	3	GSM2683221	GSE100447
HE_hESCs_1	hepatic endoderm from hESCs	1	GSM2683222	GSE100447
HE_hESCs_2	hepatic endoderm from hESCs	2	GSM2683223	GSE100447
HE_hESCs_3	hepatic endoderm from hESCs	3	GSM2683224	GSE100447
HLCs_hESCs_1	hepatocyte-like cells from hESCs	1	GSM2683225	GSE100447
HLCs_hESCs_2	hepatocyte-like cells from hESCs	2	GSM2683226	GSE100447
HLCs_hESCs_3	hepatocyte-like cells from hESCs	3	GSM2683227	GSE100447
Fetal_Liver_1	fetal liver	1	GSM2683228	GSE100447
Fetal_Liver_2	fetal liver	2	GSM2683229	GSE100447
PHH_1	primary human hepatocytes	1	GSM2683230	GSE100447
PHH_2	primary human hepatocytes	2	GSM2683231	GSE100447
iPSCs_1	induced pluripotent stem cells	1	GSM1618658	GSE66282
iPSCs_2	induced pluripotent stem cells	2	GSM1618659	GSE66282
DE_iPSCs_1	definite endoderm from iPSCs	1	GSM1618660	GSE66282
DE_iPSCs_2	definite endoderm from iPSCs	2	GSM1618661	GSE66282
HE_iPSCs_1	hepatic endoderm from iPSCs	1	GSM1618662	GSE66282
HE_iPSCs_2	hepatic endoderm from iPSCs	2	GSM1618663	GSE66282
HLCs_iPSCs_1	hepatocyte-like cells from iPSCs	1	GSM1618664	GSE66282
HLCs_iPSCs_2	hepatocyte-like cells from iPSCs	2	GSM1618665	GSE66282
fetal_liver_1	fetal liver	1	GSM1618666	GSE66282
fetal_liver_2	fetal liver	2	GSM1618667	GSE66282
PHH_1	primary human hepatocytes	1	GSM1618668	GSE66282
PHH_2	primary human hepatocytes	2	GSM1618669	GSE66282

Table 1. Samples related to data sets in repositories.

Data Records

Data Record 1

The iPSC-related microarray experiments related to this publication have been performed on the Illumina BeadStation 500 platform (Illumina, San Diego, CA, USA) using the Illumina HumanHT-12 BeadChip. The data were uploaded to NCBI GEO and are accessible under (Data Citation 1). The dataset (Data Citation 1) was first released to the public with the publication Matz *et al.*[17].

Data Record 2

The hESC-related microarray experiments related to this publication have been performed on the Illumina BeadStation 500 platform (Illumina, San Diego, CA, USA) using the HumanRef-8 Expression BeadChip. The data were uploaded to NCBI GEO and are accessible under (Data Citation 2). The dataset (Data Citation 2) is being released for the first time with the publication of this Data Descriptor.

Technical Validation

Transcriptome data

Microarray data were quality controlled via the proprietary Illumina quality control mechanisms. Tables of Pearson correlation coefficients of all samples vs. each other were generated validating the absence of outliers (Tables 2 and 3). Several samples were investigated in triplicates, all others in duplicates. Figure 1b and Figure 1c show that the replicates cluster together as well in the hESC as in the iPSC differentiation experiments as one would expect. Both bootstrap sampling methods implemented in the *pvclust* clustering software confirmed that all clusters within the dendrogram are with one exception (98%) at 100% supported by data. This demonstrates the validity of experiments on the level of whole-genome gene expression.

sample	hESCs_1	hESCs_2	hESCs_3	DE_1	DE_2	DE_3	HE_1	HE_2	HE_3	HLCs_1	HLCs_2	HLCs_3	Fetal_Liver_1	Fetal_Liver_2	PHH_1	PHH_2
hESCs_1	1.0000	0.9927	0.9913	0.9489	0.9455	0.9463	0.9534	0.9539	0.9564	0.9117	0.9057	0.9081	0.6798	0.6787	0.6062	0.6079
hESCs_2	0.9927	1.0000	0.9964	0.9483	0.9458	0.9463	0.9530	0.9538	0.9567	0.9094	0.9081	0.9100	0.6910	0.6919	0.6103	0.6113
hESCs_3	0.9913	0.9964	1.0000	0.9477	0.9463	0.9460	0.9539	0.9544	0.9563	0.9177	0.9182	0.9206	0.6903	0.6907	0.6114	0.6124
DE_1	0.9489	0.9483	0.9477	1.0000	0.9963	0.9968	0.9631	0.9631	0.9640	0.9308	0.9270	0.9273	0.6743	0.6725	0.6133	0.6149
DE_2	0.9455	0.9458	0.9463	0.9963	1.0000	0.9978	0.9620	0.9613	0.9618	0.9255	0.9255	0.9281	0.6663	0.6645	0.6097	0.6115
DE_3	0.9463	0.9463	0.9460	0.9968	0.9978	1.0000	0.9597	0.9586	0.9601	0.9213	0.9200	0.9220	0.6662	0.6643	0.6113	0.6135
HE_1	0.9534	0.9530	0.9539	0.9631	0.9620	0.9597	1.0000	0.9981	0.9974	0.9453	0.9422	0.9447	0.6805	0.6799	0.6224	0.6234
HE_2	0.9539	0.9538	0.9544	0.9631	0.9613	0.9586	0.9981	1.0000	0.9975	0.9458	0.9432	0.9451	0.6831	0.6825	0.6211	0.6223
HE_3	0.9564	0.9567	0.9563	0.9640	0.9618	0.9601	0.9974	0.9975	1.0000	0.9411	0.9379	0.9398	0.6812	0.6809	0.6211	0.6226
HLCs_1	0.9117	0.9094	0.9177	0.9308	0.9255	0.9213	0.9453	0.9458	0.9411	1.0000	0.9871	0.9834	0.6912	0.6889	0.6063	0.6086
HLCs_2	0.9057	0.9081	0.9182	0.9270	0.9255	0.9200	0.9422	0.9432	0.9379	0.9871	1.0000	0.9972	0.6931	0.6923	0.6121	0.6149
HLCs_3	0.9081	0.9100	0.9206	0.9273	0.9281	0.9220	0.9447	0.9451	0.9398	0.9834	0.9972	1.0000	0.6854	0.6849	0.6110	0.6139
Fetal_Liver_1	0.6798	0.6910	0.6903	0.6743	0.6663	0.6662	0.6805	0.6831	0.6812	0.6912	0.6931	0.6854	1.0000	0.9979	0.7384	0.7364
Fetal_Liver_2	0.6787	0.6919	0.6907	0.6725	0.6645	0.6643	0.6799	0.6825	0.6809	0.6889	0.6923	0.6849	0.9979	1.0000	0.7422	0.7403
PHH_1	0.6062	0.6103	0.6114	0.6133	0.6097	0.6113	0.6224	0.6211	0.6211	0.6063	0.6121	0.6110	0.7384	0.7422	1.0000	0.9973
PHH_2	0.6079	0.6113	0.6124	0.6149	0.6115	0.6135	0.6234	0.6223	0.6226	0.6086	0.6149	0.6139	0.7364	0.7403	0.9973	1.0000

Table 2. **Pearson correlation coefficients of hESC-derived transcriptome data of all samples vs. each other.**

sample	iPSC_B1_1	iPSC_B1_2	DE_1	DE_2	HE_1	HE_2	HLC_1	HLC_2	fetal_liver_1	fetal_liver_2	PHH_1	PHH_2
iPSC_B1_1	1.0000	0.9948	0.9356	0.9387	0.9502	0.9490	0.9155	0.9214	0.8150	0.8244	0.7413	0.7336
iPSC_B1_2	0.9948	1.0000	0.9419	0.9448	0.9541	0.9527	0.9228	0.9274	0.8236	0.8331	0.7512	0.7431
DE_1	0.9356	0.9419	1.0000	0.9980	0.9202	0.9124	0.9546	0.9496	0.8191	0.8285	0.7444	0.7363
DE_2	0.9387	0.9448	0.9980	1.0000	0.9238	0.9165	0.9548	0.9508	0.8199	0.8297	0.7453	0.7371
HE_1	0.9502	0.9541	0.9202	0.9238	1.0000	0.9966	0.9225	0.9363	0.8360	0.8379	0.7643	0.7581
HE_2	0.9490	0.9527	0.9124	0.9165	0.9966	1.0000	0.9165	0.9311	0.8328	0.8345	0.7639	0.7582
HLC_1	0.9155	0.9228	0.9546	0.9548	0.9225	0.9165	1.0000	0.9956	0.8399	0.8488	0.7714	0.7642
HLC_2	0.9214	0.9274	0.9496	0.9508	0.9363	0.9311	0.9956	1.0000	0.8440	0.8510	0.7753	0.7685
fetal_liver_1	0.8150	0.8236	0.8191	0.8199	0.8360	0.8328	0.8399	0.8440	1.0000	0.9941	0.8684	0.8624
fetal_liver_2	0.8244	0.8331	0.8285	0.8297	0.8379	0.8345	0.8488	0.8510	0.9941	1.0000	0.8662	0.8593
PHH_1	0.7413	0.7512	0.7444	0.7453	0.7643	0.7639	0.7714	0.7753	0.8684	0.8662	1.0000	0.9974
PHH_2	0.7336	0.7431	0.7363	0.7371	0.7581	0.7582	0.7642	0.7685	0.8624	0.8593	0.9974	1.0000

Table 3. **Pearson correlation coefficients of iPSC-derived transcriptome data of all samples vs. each other.**

k-means clustering to assess differentiation stages and similarity to primary hepatocytes

Normalized gene expression microarray data of the iPSC differentiation experiments were further investigated via a k-means clustering algorithm. The algorithm split the data into 100 clusters of genes with similar expression over all differentiation stages. Associations of genes with clusters are included in the publication by Matz *et al.*[17]. Several clusters were representative for distinct differentiation stages. Genes from cluster#9 were employed to make a tissue type prediction via the tool KeyGenes[25] (Figure 2a). Based on the normalized gene expression data of these genes KeyGenes predicted the tissue type "liver" for HLC, fetal liver and PHH samples. Figure 2b demonstrates that genes from cluster#9 have most abundantly peaks at the HLC stages.

Furthermore, k-means clustering provided several stage-specific clusters six of which are shown in Figure 3. These represent stages iPSCs, definite endoderm, hepatic endoderm, HLCs, fetal liver and PHHs and display a gene expression peak at the dedicated stage. They include stage-specific markers which in some cases are already known: *POU5F1/ OCT4* in the iPSC-cluster, *SOX17* in the DE-cluster, *AFP* in the fetal-liver-cluster and *ALB* in the PHH-cluster. In Supplementary Fig. S4F of our previous publication related to the iPSC dataset[17] of this data descriptor we could confirm PHH-cluster activity of the transcription factors *HNF4α* and *HNF1α* reported by Li *et al.*[15] as factors inducing hepatocyte

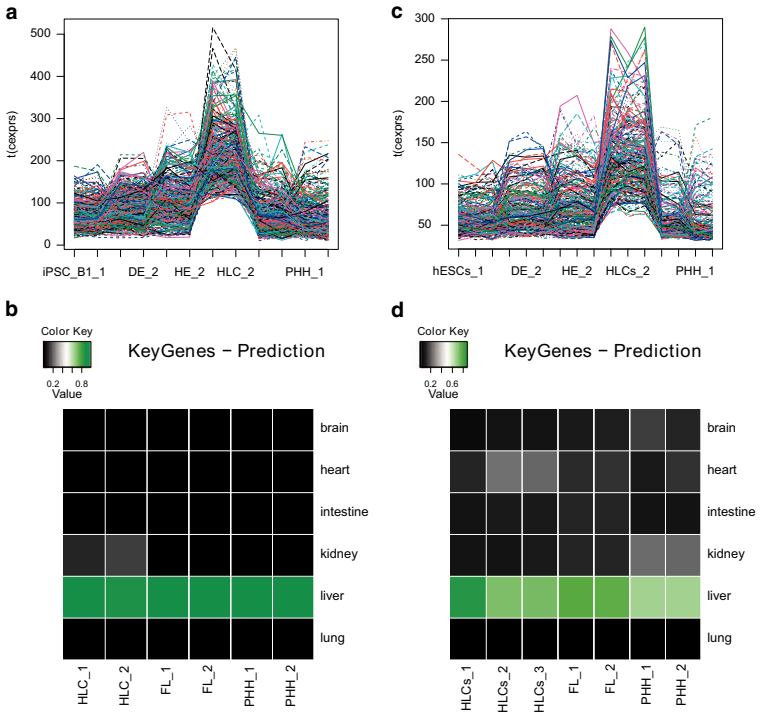


Figure 2. Characterization of hepatocyte-like cells. (**a**) Plot of 407 genes from the k-means cluster#9 over all differentiation stages derived from the iPSCs. The plot demonstrates that this cluster is representative for HLCs. (**b**) KeyGenes tissue classification of for iPSCs k-means Hepatocyte-like-cell (HLC) cluster9 (source: Matz *et al.*[17]). Based on NCBI GEO datasets for human liver, brain, intestine, kidney, lung and heart via the KeyGenes tool a training set for these Illumina microarray platform data was generated. Genes from the HLC cluster#9 resulting from k-means clustering and HLC, fetal liver (FL) and primary human hepatocyte samples were used as test set. (**c**) Plot of 263 genes from the k-means cluster#2 over all differentiation stages derived from the hESCs. The plot demonstrates that this cluster is representative for HLCs. (**d**) KeyGenes tissue-based classification for hESCs k-means Hepatocyte-like-cell (HLC) cluster#2.

differentiation and furthermore reveal the activity of *CTCF, ZFX, FOXA2, FOXA1, CEBPA*. Additionally, these datasets may provide new insights into the differences and similarities of the hepatocyte differentiation processes between hESCs and iPSCs. Figure 4 using marker genes from the representative k-means-clusters shows that the DE stage and HLCs are very similar between hESC- and iPSC-derived differentiations while the HE stage provides a pronounced peak in iPSC-derived cells and a small peak in hESC-derived cells. As a cautionary note, the iPSC and hESC differentiations into HLCs and also the microarray-based transcriptome analyses were not conducted simultaneously, hence the observed minor variations.

Usage Notes

The microarray experiments related to this publication have been performed on the Illumina BeadStation 500 platform (Illumina, San Diego, CA, USA) but on different BeadChips. The iPSC-derived differentiations were hybridized using the Illumina HumanHT-12 BeadChip while the hESC-derived differentiations were hybridized using the HumanRef-8 Expression BeadChip. The differing chip types should be taken into account when comparing transcriptomics data between hESC-derived and iPSC-derived experiments. Further points which should be considered are: (1) The fetal liver RNA was derived from liver homogenates containing all cells, but the adult liver samples were derived from isolated hepatocytes; (2) the liver disease for which the biopsy was performed could have an influence on the dataset; (3) the two different differentiation protocols used may affect the data.

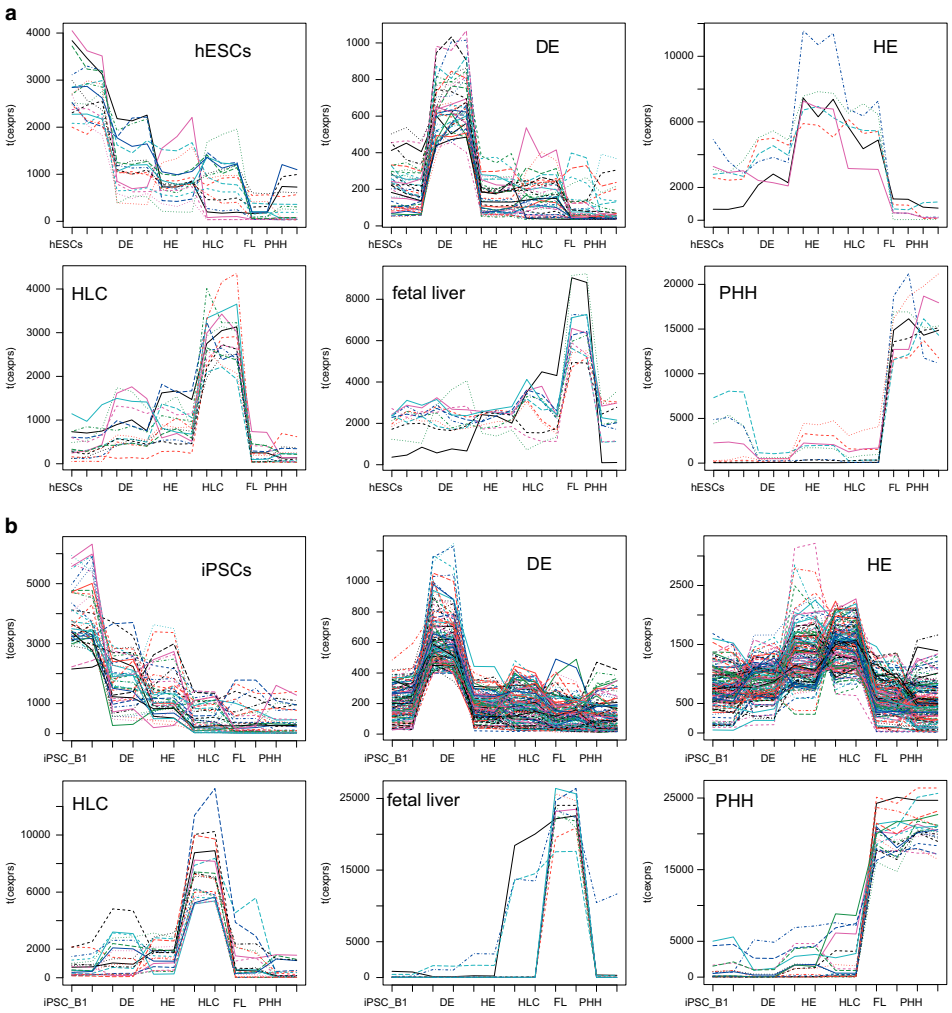


Figure 3. Clusters representative for all stages of differentiation. Via k-means clustering genes were assigned to 100 clusters having similar behaviour during the hepatic differentiation process. The plots in (**a**) show expression of genes in the hESC-differentiation experiments associated with cluster#94 representative for hESCs, cluster#23 containing *SOX17* representative for DE, cluster#64 representative for HE, cluster#77 representative for HLCs, cluster#19 containing AFP representative for fetal liver, cluster#59 containing the liver marker ALB representative for PHHs. The plots in (**b**) show expression of genes in the iPSC-differentiation experiments associated with cluster#68 representative for iPSCs, cluster#81 containing *SOX17* representative for DE, cluster#37 representative for HE, cluster#51 representative for HLCs, cluster#72 containing AFP representative for fetal liver, cluster#91 containing the liver marker *ALB* representative for PHHs.

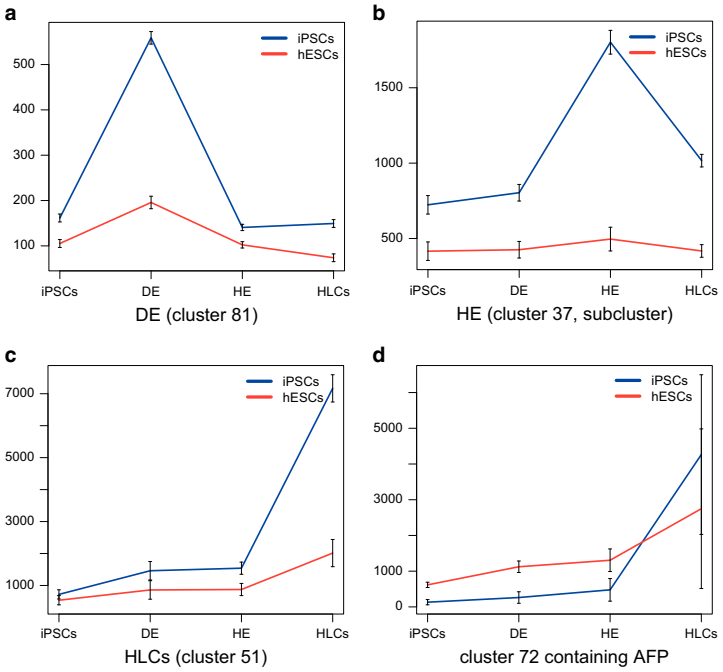


Figure 4. Comparison of clusters representative for DE, HE and HLCs between hESC- and iPSC differentiation. Genes from representative clusters for (**a**) DE (cluster#81), (**b**) HE (cluster#37,subcluster), (**c**) HLCs (cluster#51) and (**d**) cluster #72 containing AFP were compared between hESC- and iPSC-derived lines. Mean and standard error of all genes in the dedicated clusters are shown. The DE, AFP and HLC clusters show maxima at the associated stages in both differentiations. The HE cluster provides a pronounced peak in iPSC-derived cells and a small peak in hESC-derived cells.

References

1. Jung, J., Zheng, M., Goldfarb, M. & Zaret, K. S. Initiation of mammalian liver development from endoderm by fibroblast growth factors. *Science* **284,** 1998–2003 (1999).
2. Schwartz, R. E. *et al.* Defined conditions for development of functional hepatic cells from human embryonic stem cells. *Stem Cells Dev.* **14,** 643–655 (2005).
3. Rossi, J. M., Dunn, N. R., Hogan, B. L. & Zaret, K. S. Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. *Genes Dev.* **15,** 1998–2009 (2001).
4. Kamiya, A., Kinoshita, T. & Miyajima, A. Oncostatin M and hepatocyte growth factor induce hepatic maturation via distinct signaling pathways. *FEBS Lett.* **492,** 90–94 (2001).
5. Elaut, G. *et al.* Molecular mechanisms underlying the dedifferentiation process of isolated hepatocytes and their cultures. *Curr. Drug Metab.* **7,** 629–660 (2006).
6. Fraczek, J., Bolleyn, J., Vanhaecke, T., Rogiers, V. & Vinken, M. Primary hepatocyte cultures for pharmaco-toxicological studies: at the busy crossroad of various anti-dedifferentiation strategies. *Arch. Toxicol.* **87,** 577–610 (2013).
7. Hart, S. N. *et al.* A comparison of whole genome gene expression profiles of HepaRG cells and HepG2 cells to primary human hepatocytes and human liver tissues. *Drug Metab. Dispos. Biol. Fate Chem* **38,** 988–994 (2010).
8. Gerets, H. H. J. *et al.* Characterization of primary human hepatocytes, HepG2 cells, and HepaRG cells at the mRNA level and CYP activity in response to inducers and their predictivity for the detection of human hepatotoxins. *Cell Biol. Toxicol.* **28,** 69–87 (2012).
9. Jetten, M. J. A., Kleinjans, J. C. S., Claessen, S. M., Chesné, C. & van Delft, J. H. M. Baseline and genotoxic compound induced gene expression profiles in HepG2 and HepaRG compared to primary human hepatocytes. *Toxicol. Vitro Int. J. Publ. Assoc. BIBRA* **27,** 2031–2040 (2013).
10. Rashid, S. T. *et al.* Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. *J. Clin. Invest.* **120,** 3127–3136 (2010).
11. Sullivan, G. J. *et al.* Generation of functional human hepatic endoderm from human induced pluripotent stem cells. *Hepatol. Baltim. Md* **51,** 329–335 (2010).
12. Yusa, K. *et al.* Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478,** 391–394 (2011).
13. Wruck, W., Graffmann, N., Kawala, M.-A. & Adjaye, J. Concise Review: Current Status and Future Directions on Research Related to Nonalcoholic Fatty Liver Disease. *Stem Cells Dayt. Ohio* **35,** 89–96 (2017).

14. Graffmann, N. *et al.* Modeling Nonalcoholic Fatty Liver Disease with Human Pluripotent Stem Cell-Derived Immature Hepatocyte-Like Cells Reveals Activation of PLIN2 and Confirms Regulatory Functions of Peroxisome Proliferator-Activated Receptor Alpha. *Stem Cells Dev.* **25,** 1119–1133 (2016).
15. Li, J., Ning, G. & Duncan, S. A. Mammalian hepatocyte differentiation requires the transcription factor HNF-4alpha. *Genes Dev.* **14,** 464–474 (2000).
16. Jozefczuk, J., Prigione, A., Chavez, L. & Adjaye, J. Comparative Analysis of Human Embryonic Stem Cell and Induced Pluripotent Stem Cell-Derived Hepatocyte-Like Cells Reveals Current Drawbacks and Possible Strategies for Improved Differentiation. *Stem Cells Dev.* **20,** 1259–1275 (2010).
17. Matz, P. *et al.* Footprint-free human fetal foreskin derived iPSCs: A tool for modeling hepatogenesis associated gene regulatory networks. *Sci. Rep.* **7,** 6294 (2017).
18. Wang, A. *et al.* Epigenetic Priming of Enhancers Predicts Developmental Competence of hESC-Derived Endodermal Lineage Intermediates. *Cell Stem Cell* **16,** 386–399 (2015).
19. Jozefczuk, J., Drews, K. & Adjaye, J. Preparation of mouse embryonic fibroblast cells suitable for culturing human embryonic and induced pluripotent stem cells. *J. Vis. Exp. JoVE*, doi:10.3791/3854 (2012).
20. Hay, D. C. *et al.* Efficient differentiation of hepatocytes from human embryonic stem cells exhibiting markers recapitulating liver development in vivo. *Stem Cells Dayt. Ohio* **26,** 894–902 (2008).
21. Agarwal, S., Holton, K. L. & Lanza, R. Efficient differentiation of functional hepatocytes from human embryonic stem cells. *Stem Cells Dayt. Ohio* **26,** 1117–1127 (2008).
22. Gentleman, R. C *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5,** R80 (2004).
23. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24,** 1547–1548 (2008).
24. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22,** 1540–1542 (2006).
25. Roost, M. S. *et al.* KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas. *Stem Cell Rep* **4,** 1112–1124 (2015).

Data Citations

1. *NCBI Gene Expression Omnibus* GSE66282 (2017).
2. *NCBI Gene Expression Omnibus* GSE100447 (2017).

Acknowledgements

J.A. acknowledges support from the Medical faculty of Heinrich Heine University, Düsseldorf, Germany.

Author Contributions

W.W. performed bioinformatic analyses. W.W. and J.A. wrote the manuscript. J.A. conceived the concept and coordinated the work.

Additional Information

Competing interests: The authors declare no competing financial interests.

How to cite this article: Wruck, W. & Adjaye, J. Human pluripotent stem cell derived HLC transcriptome data enables molecular dissection of hepatogenesis. *Sci. Data* 5:180035 doi:10.1038/sdata.2018.35 (2018).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2.1.7 The FGF, TGFβ and WNT axis Modulate Self-renewal of Human SIX2+ Urine Derived Renal Progenitor Cells.

Sci Rep. 2020 Jan 20;10(1):739. doi: 10.1038/s41598-020-57723-2.

Rahman MS(1), Wruck W(1), Spitzhorn LS(1), Nguyen L(1), Bohndorf M(1), Martins S(1), Asar F(1), Ncube A(1), Erichsen L(1), Graffmann N(1), Adjaye J(2).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University Düsseldorf, 40225, Düsseldorf, Germany. (2)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University Düsseldorf, 40225, Düsseldorf, Germany. james.adjaye@med.uni-duesseldorf.de.

Human urine is a non-invasive source of renal stem cells with regeneration potential. Urine-derived renal progenitor cells were isolated from 10 individuals of both genders and distinct ages. These renal progenitors express pluripotency-associated proteins- TRA-1-60, TRA-1-81, SSEA4, C-KIT and CD133, as well as the renal stem cell markers -SIX2, CITED1, WT1, CD24 and CD106. The transcriptomes of all SIX2+ renal progenitors clustered together, and distinct from the human kidney biopsy-derived epithelial proximal cells (hREPCs). Stimulation of the urine-derived renal progenitor cells (UdRPCs) with the GSK3β-inhibitor (CHIR99021) induced differentiation. Transcriptome and KEGG pathway analysis revealed upregulation of WNT-associated genes- AXIN2, JUN and NKD1. Protein interaction network identified JUN- a downstream target of the WNT pathway in association with STAT3, ATF2 and MAPK1 as a putative negative regulator of self-renewal. Furthermore, like pluripotent stem cells, self-renewal is maintained by FGF2-driven TGFβ-SMAD2/3 pathway. The urine-derived renal progenitor cells and the data presented should lay the foundation for studying nephrogenesis in human.

DOI: 10.1038/s41598-020-57723-2 PMCID: PMC6970988 PMID: 31959818

URL: https://www.nature.com/articles/s41598-020-57723-2

The data presented in this publication has been used also for the PhD thesis of Dr. Md. Shaifur Rahman and Dr. Soraia Martins at the Heinrich Heine University Düsseldorf.

Contribution of Wasco Wruck: Bioinformatics; Writing of manuscript. These authors contributed equally: Md Shaifur Rahman and Wasco Wruck.

www.nature.com/scientificreports

SCIENTIFIC REPORTS
nature research

OPEN

The FGF, TGFβ and WNT axis Modulate Self-renewal of Human SIX2$^+$ Urine Derived Renal Progenitor Cells

Md Shaifur Rahman[1,2], Wasco Wruck[1,2], Lucas-Sebastian Spitzhorn[1], Lisa Nguyen[1], Martina Bohndorf[1], Soraia Martins[1], Fatima Asar[1], Audrey Ncube[1], Lars Erichsen[1], Nina Graffmann[1] & James Adjaye[1*]

Human urine is a non-invasive source of renal stem cells with regeneration potential. Urine-derived renal progenitor cells were isolated from 10 individuals of both genders and distinct ages. These renal progenitors express pluripotency-associated proteins- TRA-1-60, TRA-1-81, SSEA4, C-KIT and CD133, as well as the renal stem cell markers -SIX2, CITED1, WT1, CD24 and CD106. The transcriptomes of all SIX2$^+$ renal progenitors clustered together, and distinct from the human kidney biopsy-derived epithelial proximal cells (hREPCs). Stimulation of the urine-derived renal progenitor cells (UdRPCs) with the GSK3β-inhibitor (CHIR99021) induced differentiation. Transcriptome and KEGG pathway analysis revealed upregulation of WNT-associated genes- *AXIN2*, *JUN* and *NKD1*. Protein interaction network identified JUN- a downstream target of the WNT pathway in association with STAT3, ATF2 and MAPK1 as a putative negative regulator of self-renewal. Furthermore, like pluripotent stem cells, self-renewal is maintained by FGF2-driven TGFβ-SMAD2/3 pathway. The urine-derived renal progenitor cells and the data presented should lay the foundation for studying nephrogenesis in human.

According to the International Society of Nephrology, more than 850 million people worldwide are afflicted with kidney diseases[1], which raises the quest for alternative therapies to overcome the limitations associated with current treatments including transplantation and dialysis. One of the most promising options is the utilization of renal stem cells for treating of kidney diseases, disease modelling, and drug development[2,3]. Renal stem/progenitor cells are self-renewing, multipotent cells with the ability to generate various cell types of the kidney to maintain renal function[4]. These progenitors are in abundance during fetal kidney development in which the renal progenitor surface marker CD24 and stem cell self-renewal marker CD133 cells are required for primordial nephrogenesis[5,6]. However, in adults, CD24, CD133 (Prominin-1) and vascular cell adhesion molecule 1 (CD106)-positive renal progenitors are present in renal tubules and capsules[7]. Two progenitor cell populations can be distinguished based on the expression of CD106. For instance, CD24$^+$CD133$^+$CD106$^-$ progenitors are present in proximal tubules whereas CD24$^+$CD133$^+$CD106$^+$ cells are localized in the Bowman's capsule. The latter can differentiate into a variety of cell types of renal tissue such as podocytes and tubular epithelial cells[4–7].

Several groups have identified urine as a non-invasive and repetitive source of renal progenitor cells[8,9]. It has been estimated that each day approximately 2,000 to 7,000 cells composed of differentiated epithelial cells, bi-potential epithelial cells (transitional cells), multipotent mesenchymal stem cells, and glomerular parietal cells are flushed out from the renal tubular network and the upper urinary tract into urine[10–12]. A subpopulation of these urine-derived cells are renal stem/progenitor cells which express master renal markers such as Sine Oculis Homeobox Homolog 2 (SIX2), Cbp/P300 Interacting Transactivator With Glu/Asp Rich Carboxy-Terminal Domain 1 (CITED1) and Wilms' Tumor 1 (WT1)[13–15] and CD24 and CD106[16]. Interestingly, these cells exhibit stem cell properties, i.e. expression of pluripotency-associated markers such as TRA-1-60, TRA-1-81, C-KIT (CD117), CD133 and SSEA4 and possess high proliferation capacity as they show telomerase activity. Further, they endow multi-differentiation potential and like bone marrow derived mesenchymal stem cells express

[1]Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University Düsseldorf, 40225, Düsseldorf, Germany. [2]These authors contributed equally: Md Shaifur Rahman and Wasco Wruck. *email: james.adjaye@med.uni-duesseldorf.de

Vimentin, CD105, CD90, CD73 and not the hematopoietic stem cell markers- CD14, CD31, CD34 and CD45[17,18]. Although, research interest on urine derived renal stem cells is gradually increasing but the mechanistic role of genetic factors in these cells *in vitro* regarding progenitor/differentiated status maintenance is not clear.

Studies in mice have shown that Odd-skipped related 1 (Osr1), Six2, Wnt, Cited1 and Wt1 are required to maintain renal progenitor cells during kidney organogenesis[19–25]. Additionally, signalling pathways such as Fgf, Tgfβ and Notch play major roles in renal stem cell maintenance and differentiation[26–29]. The transcription factor Osr1 is an early marker specific for the intermediate mesenchyme (IM); *Osr1* knockout mice lack renal structures due to the failure to form the IM[30]. The homeodomain transcriptional regulator Six2 is expressed in the cap mesenchyme (CM) originating from metanephric mesenchyme. Six2 positive populations can generate all cell types of the main body of the nephron[31]. Inactivation of Six2 results in premature and ectopic renal vesicles, leading to a reduced number of nephrons and to renal hypoplasia[32]. Mechanistically, Osr1 plays a crucial role in Six2-dependent maintenance of mouse nephron progenitors by antagonizing Wnt-directed differentiation, whereas Wt1 maintains self-renewal by modulating Fgf signals[22,23]. Cited1 has been reported to be co-expressed with a fraction of $Six2^{+}$ cells undergoing self-renewal and these can be differentiated in response to activated WNT signaling during kidney development[25]. Furthermore, it has been demonstrated in mice that Bmp7 promotes proliferation of nephron progenitor cells via a Jnk-dependent mechanism involving phosphorylation of Jun and Atf2[33].

To date, research related to transcriptional regulatory control of mammalian nephrogenesis has been limited to the mouse[19,26] or to transcriptome "snapshots" in human[13]. A recent study demonstrated conserved and divergent genes associated with human and mouse kidney organogenesis[34], thus further highlighting the need for primary human renal stem cell models to better dissect nephrogenesis at the molecular level. Furthermore, species differences need to be considered, for example, mammalian nephrons arise from a limited nephron progenitor pool through a reiterative inductive process extending over days (mouse) or weeks (human) of kidney development[35]. Human kidney development initiates around 4 weeks of gestation and ends around 34–37 weeks of gestation. At the anatomical level, human and mouse kidney development differ in timing, scale, and global features such as lobe formation and progenitor niche organization[34–36]. These are all further evidence in support of the need of a reliable and robust human renal cell culture model.

Expression of pluripotency-associated proteins has enabled rapid reprogramming of urine derived mesenchymal and epithelial cells into induced pluripotent stem cells (iPSCs)[37–41]. Differentiation protocols for generating kidney-associated cell types from human pluripotent stem cells have mimicked normal kidney development[28,42–44]. For example, WNT activation using a GSK3β inhibitor (CHIR99021), FGF9, Activin A, Retinoic acid (RA) and BMP7 as instructive signals have been employed to derive functional podocytes, proximal renal tubules, and glomeruli[29,45–49]. Despite these efforts and achievements, there will always be variabilities between differentiation protocols, the maturation state of the differentiated renal cells and genes associated with temporal maturation during human kidney organoids formation from human iPSCs[50,51]. We propose that using native renal stem cells isolated directly from urine will circumvent most of the shortfalls and deficiencies associated with human pluripotent stem cell-based models.

Here we provide for the first time the full characterisation of renal progenitors at the transcriptome, secretome and cellular level, which has led to the identification of a gene regulatory network and associated signalling pathways that maintain their self-renewal. We anticipate that our data will enhance our meagre understanding of the properties of urine-derived renal stem cells, and enable the generation of renal disease models *in vitro* and eventually kidney-associated regenerative therapies.

Results

Urine-derived renal progenitors express a subset of pluripotent stem cell-associated markers and possess features typical of bone marrow-derived MSC. Urine samples were collected from 10 healthy adult donors (4 males-UM and 6 females-UF) with ages ranging from 21 to 61 years, and of mixed ethnicity (3 Africans and 7 Caucasians) (Supplemental Table S1). Attached cells emerged from processed urine as isolated clusters after 7 days, thereafter these acquired a "rice grain" fibroblast-like morphology resembling MSCs (Fig. 1A, Supplemental Fig. S1A). A selection of distinct urine-derived renal stem cells populations (n = 4) were used to assay cell proliferation and growth. After 3 days in culture, the cells exited the lag phase and growth began in an exponential phase. Cells attained stationary phase at day 7 of subculture (Fig. 1B). All four populations-UM27, UM16, UM51 and UF45 showed similar proliferation and growth patterns.

Flow cytometry analysis revealed that approximately 98.9% of the cells express SSEA4, TRA-1–60 (11.3%) and TRA-1-81 (16.5%) (Fig. 1C). These data were confirmed by immunofluorescent-based staining of SSEA4 which also express the proliferation-associated stem cell markers- C-KIT and CD133 (Fig. 1D). In order to reveal the detailed methylation pattern of the 5′-regulatory region of the OCT4 gene in the UM51, we employed standard bisulfite sequencing. In total 330 Cytosine-phosphatidyl-Guanine-dinucleotides (CpG) upstream of the transcription-starting site (TSS) of the OCT4 gene were analysed. Within this 469 bp long region, a dense methylation pattern was observed in the UM51 cells, with 92.4% (305) of the CpG dinucleotides identified were methylated (Fig. 1E). In contrast, iPSCs derived from UM51 had 72.12% (207) of analysed CpGs were unmethylated (Supplemental Fig. S1B).

Urine-derived renal progenitors express the mesenchymal marker- Vimentin and not the epithelial marker-E-Cadherin (Fig. 1D, Supplemental Fig. S1C). Flow cytometry analysis of critical MSC cell surface markers were negative for the hematopoietic markers CD14, CD20, CD34, and CD45 and positive for CD73, CD90 and CD105 albeit at variable levels (Supplemental Fig. S1D). Typical of MSCs, urine-derived renal progenitor cells can also differentiate into osteocytes, chondrocytes, and adipocytes when cultured in the respective differentiation medium for 3 weeks (Fig. 1F, Supplemental Fig. S1E). Furthermore, employing a cytokine array (n = 2), a plethora of trophic factors such as IL8, GDF-15, SERPINE-1, Angiogenin, VEGF, and Thrombospondin-1 were detected,

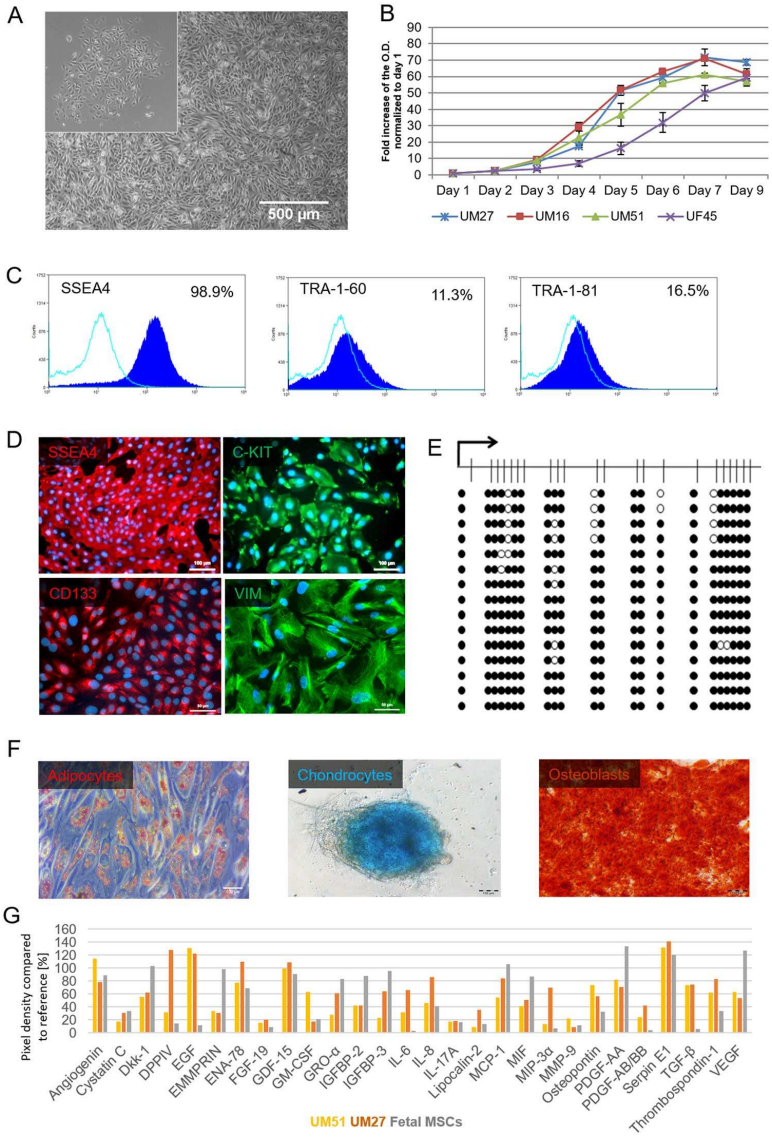


Figure 1. Propagation and characterisation of urine-derived renal progenitors. **(A)** Representative pictures of the "rice grain"-like appearance of the cells from the initial attachment to an elongated MSC-like morphology. **(B)** Growth curve analysis of selected urine-derived renal progenitors carried out using the Resazurin metabolic assay. Data are presented as means ± SEMs. **(C)** Immune-phenotyping for SSEA4, TRA-1-81 and TRA-1-60; and **(D)** immunofluorescence-based detection of the expression of pluripotency-associated stem cell- proteins SSEA4 (red), C-KIT (green), CD133 (red) and the mesenchymal-associated protein Vimentin (green); cell nuclei were stained using Hoechst/DAPI (scale bars: 100 µm and 50 µm). **(E)** Bisulfite sequencing of CpG island methylation patterns within the 5′- regulatory region of the OCT4 gene in UM51. Filled circles stand for methylated CpG dinucleotides. White circles stand for unmethylated CpGs. Arrows indicate the transcription start site. **(F)** *In vitro* Osteoblast, Chondrocyte and Adipocyte differentiation potential of urine-derived renal progenitors. **(G)** Cytokines secreted by urine-derived renal progenitors in culture media. Lists of significant GOs and KEGG pathways associated with the genes encoding the secreted cytokines are shown in Supplemental Fig. S1G.

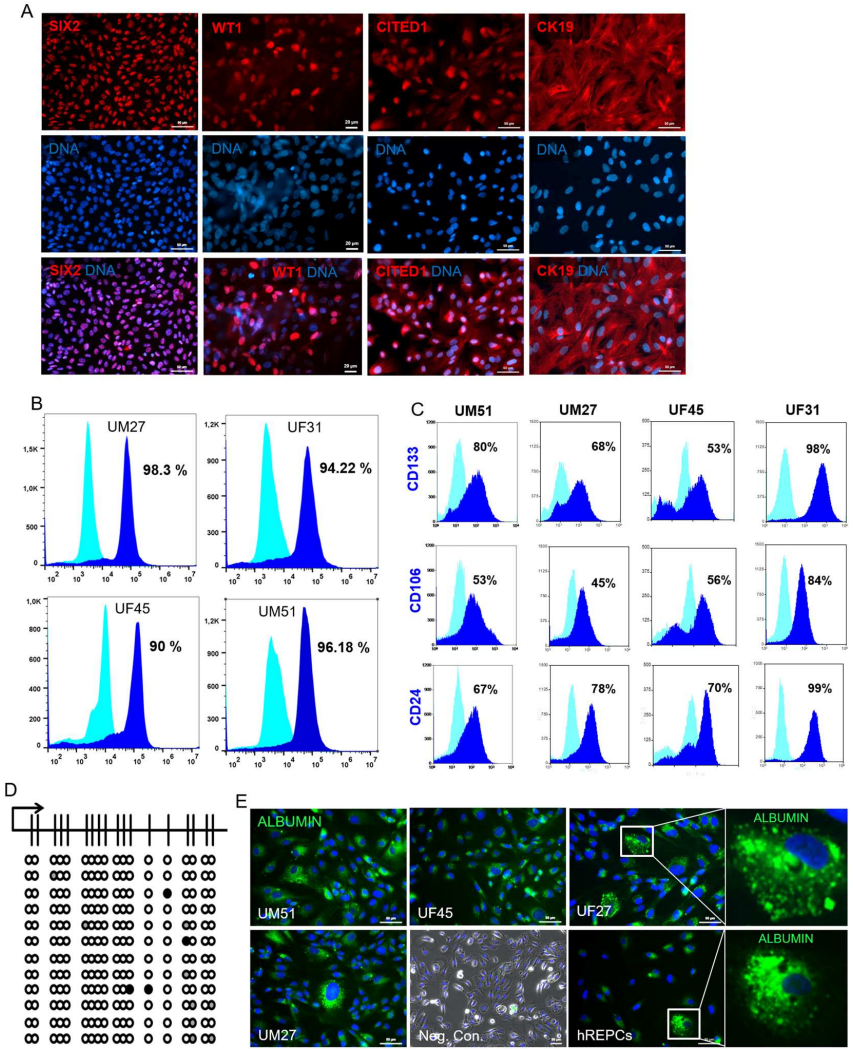


Figure 2. Expression of kidney-associated proteins in urine-derived renal progenitors and Albumin transport. **(A)** Urine-derived renal progenitors express the renal stem cell markers- SIX2, CITED1, WT1, and CK19. Renal markers (red) and cell nuclei were stained using DAPI/Hoechst (blue). **(B)** Flow cytometry analysis for the key renal stem cell transcription factor SIX2 and **(C)** Renal stem cell surface markers CD24, CD133, and CD106 of UM27, UF31, UF45 and UM51. **(D)** Detailed CpG methylation profiles of the SIX2 5′-regulatory region are documented as revealed by bisulfite sequencing. Filled circles represent methylated CpG dinucleotides and white circles unmethylated CpGs. Arrows indicate the transcription start site. 1.9% of CpG dinucleotides were found to be methylated. **(E)** Urine-derived renal progenitors (n = 4) like the human kidney biopsy-derived hREPCs also transport Albumin. Albumin was coupled to Alexa Fluor 488 (green) and cell nuclei stained with DAPI (blue). Scale bars indicate 50 µm.

and further analysis of their associated GOs and KEGG pathways revealed immune system related terms (Fig. 1G, Supplemental Fig. S1F-G).

Urine-derived renal progenitors express key renal progenitor cell markers and are able to endocytose Albumin. Immunofluorescence-based staining revealed expression of the key renal stem cell proteins such as CK19 and the transcription factors- SIX2, CITED1, WT1, as shown by representative images (Fig. 2A).

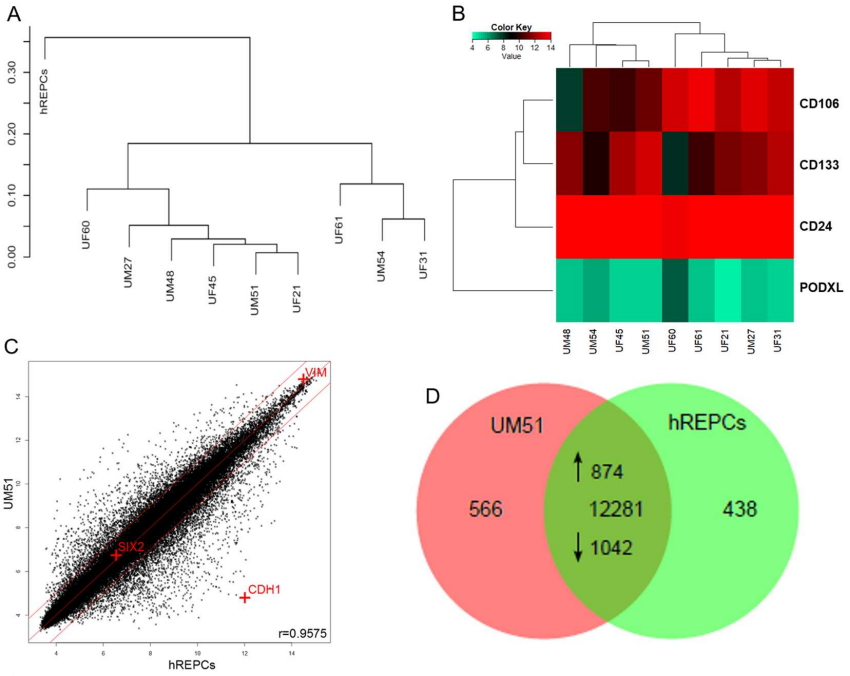
To determine the variability of SIX2$^+$cells between the progenitor cell preparations- UM27, UF31, UM51 and UF45 (n = 4) a flow cytometry analysis was performed. We observed approximately 95% SIX2$^+$ cells in UM27, UF31 and UM51 whereas UF45 had 90% SIX2$^+$ cells (Fig. 2B). In addition, to confirm the renal stemness status of the urine-derived progenitor cells a flow cytometry analysis was performed to evaluate expression of the renal progenitor markers CD24, CD106 and the self-renewal marker CD133 in the cell preparations- UM27, UF31, UM51 and UF45. CD24, CD133 and CD106 were variably expressed in the aforementioned cell preparations. For instance, 98% of the UF31 cell population was CD133$^+$, 99% of the cells were positive for CD24 and 84% of the cells were CD106$^+$. On the other hand, the UF45 sample displayed a different pattern for CD133 (68%), CD24 (70%) and CD106 (45%) positive cells, respectively (Fig. 2C). Bisulfite sequencing of a portion of the 5′-regulatory region of the SIX2 gene revealed methylation of only 1.9% of CpG dinucleotides (Fig. 2D). As, presence of albumin in urine is a mark of kidney cell functionality, and by the endo/exocytosis of albumin, kidney maintain the colloid osmotic pressure and transport biomolecules. We performed endocytosis assay and could show that urine-derived cells can transport Albumin (Fig. 2E). Furthermore, the CYP2D6 genotypes investigated were distinct between groups of individuals, thus reflecting potential diverse drug metabolizing activities. UM51 for example expresses the CYP2D6 *4/*17 genotype which confers an intermediate metabolizing activity whereas UF31 bears the CYP2D6*1/*41 genotype with an ultra-rapid metabolizing activity. The other three individuals (UF21, UF45 and UM27) are endowed with normal drug metabolizing activity (Supplemental Table S1).

Comparative transcriptome analysis of urine-derived renal progenitors and kidney-biopsy derived renal epithelial proximal cells (hREPCs). A hierarchical clustering analysis comparing the transcriptomes of urine derived renal progenitors with the kidney biopsy-derived renal epithelial proximal cells (hREPCs) revealed that all urine derived renal progenitors samples clustered together as a common cell type distinct from hREPCs (Fig. 3A). Additionally, expression of renal progenitor surface markers *CD24*, *CD106* and *CD133* were detected in urine-derived renal progenitors whilst *PODXL* was not expressed (Fig. 3B). These renal progenitors are of mesenchymal origin expressing *VIM*, however a scatter plot comparison between UM51 with hREPCs shows similarity with a high Pearson correlation of 0.9575 (Fig. 3C). The epithelial character of hREPCs is reflected by *CDH1* expression. The comparison of expressed genes (det-p < 0.05) in renal progenitors (UM51) and hREPCs in a venn diagram revealed a common 12281 gene-set, whereas 566 are expressed exclusively in UM51 and 438 exclusively in hREPCs (Fig. 3D). The 10 most over-represented GO BP terms (biological processes) in the UM51 exclusive gene-set include triglyceride homeostasis, kidney development and urogenital system development, whereas the hREPCs exclusive gene set includes chloride transmembrane transport, anion transport and response to lipopolysaccharides (Fig. 3E). The common gene set consists of 874 up-regulated genes (ratio > 2) in UM51 (e.g. renal tubule development, urogenital system development and anterior/posterior pattern specification) and 1042 down-regulated genes (ratio < 0.5) in UM51 (e.g. cell division and cholesterol biosynthetic process) (Fig. 3F).

Comparative gene expression analysis of urine-derived and kidney biopsies-derived renal progenitor cells. Gene expression of urine-derived renal progenitors was compared to public available datasets GSE23911 in which nephron progenitor cells were derived from adult human renal cortical tissue[52]. Additionally, the comparison was extended by two further datasets GSE74450 and GSE75949 which contain data from fetal kidney biopsy derived nephron progenitor cells[53,54]. We could show that urine-derived renal progenitors have a high level of similarity to other human nephron progenitors at the transcriptome level. The resulting number of expressed genes were comparable: 12112 genes in urine-derived renal progenitors, 8446 genes in GSE23911, 10597 genes in GSE74450 and 13895 genes in GSE75949. In the Venn diagram analysis most genes were found in the intersection of all genesets (4411), followed by the intersection of urine-derived renal progenitors with the fetal kidney genesets from GSE74450 and GSE75949. Among the intersection with single genesets urine-derived renal progenitors had the highest overlap with the GSE75949 pointing at the highest similarity with this gene-set (Supplemental Fig. S2). A subset of genes expressed in common between urine-derived renal progenitors, GSE74450 and GSE75949, are associated with renal system development related GO's (BP) terms, thus confirming renal progenitor cell identity (Supplemental Table S5).

Confirmation of the renal origin of urine-derived progenitor cells and retention of renal-associated genes in urine-derived progenitors-iPSCs. A venn diagram-based comparison of gene expression (det-p < 0.05) in urine-derived renal progenitors and human foreskin fibroblasts (HFF) was carried out (Fig. 4A) in order to dissect common and distinct gene expression patterns. The majority of genes (11649) are expressed in common, 463 exclusively in urine-derived renal progenitors and 891 in fibroblasts. The 463 genes were further analysed for over-represented GOs and summarized as a GO network (Fig. 4B) with the tools REVIGO, and Cytoscape was used for the GO terms of the category BP. In addition to several developmental terms such as organ induction, regulation of embryonic development (high number of edges referring to similarity to many terms), specific renal-related terms including urogenital system development, mesenchymal cell proliferation involved in ureteric bud development and positive regulation of nephron tubule epithelial cell differentiation (marked with blue ellipse, intense red indicating higher significance) were identified. Interestingly, the non-canonical WNT signalling pathway, which plays a major role in kidney development, is also over-represented (orange ring-top left).

The dendrogram based on the global transcriptome analysis revealed a clear separation of urine-derived renal progenitors lines (n = 9) from the differentiated urine-derived renal progenitors (CHIR 99021 treated urine- derived renal progenitor cells, n = 3), urine derived renal progenitors-iPSCs (n = 4) and embryonic stem cells (H1 and H9) (Fig. 4C). Characterization of the derived urine derived renal progenitors-iPSCs is depicted in Supplemental Fig. S3. In the Venn diagram (Fig. 4D) we compared expressed genes (det-p < 0.05) in urine



E

Term	Pvalue
triglyceride homeostasis	9,605E-06
kidney development	0,00042078
urogenital system development	0,00046036
collagen fibril organization	0,0004605
telencephalon development	0,000505
embryonic retina morphogenesis in camera-type eye	0,00051672
sensory organ development	0,00117891
neuronal stem cell population maintenance	0,00140151
cell morphogenesis involved in differentiation	0,00181875
generation of ovulation cycle rhythm	0,00187149

Term	Pvalue
chloride transmembrane transport	0,00014
prostate epithelial cord arborization involved in prostate glandular acinus morphogenesis	0,00031
regulation of branching involved in prostate gland morphogenesis	0,00031
response to lipopolysaccharide	0,00042
inorganic anion transport	0,00053
anion transport	0,00056
regulation of icosanoid secretion	0,00061
negative regulation of secretion	0,00072
prostate glandular acinus development	0,00102
positive regulation of cytosolic calcium ion concentration	0,00119

F

UM51 up (differentially expressed genes)	
Term	Pvalue
circulatory system development	4,102E-12
positive regulation of RNA biosynthetic process	4,4896E-09
urogenital system development	2,3935E-08
anterior/posterior pattern specification	2,6834E-08
positive regulation of nucleobase-containing compound metabolic process	4,7051E-08
positive regulation of biosynthetic process	2,6254E-07
muscle tissue development	9,4143E-07
renal tubule development	1,0127E-06
positive regulation of macromolecule metabolic process	1,5071E-06
tissue morphogenesis	2,9227E-06

UM51 down (differentially expressed genes)	
Term	Pvalue
localization of cell	5,3154E-15
organelle fission	8,1634E-14
cholesterol biosynthetic process	7,2679E-13
small molecule biosynthetic process	8,9168E-12
cell division	1,2264E-11
regulation of phosphorylation	5,8579E-11
regulation of phosphorus metabolic process	8,1264E-11
organic hydroxy compound biosynthetic process	8,8558E-10
negative regulation of molecular function	1,045E-09
secondary alcohol metabolic process	1,4932E-09

Figure 3. Transcriptome analysis of urine-derived renal progenitors in comparison to kidney biopsy-derived renal epithelial proximal cells- hREPCs. **(A)** A hierarchical cluster dendrogram based on transcriptomes of urine-derived renal progenitors with the kidney biopsy-derived renal epithelial proximal cells (hREPCs). **(B)** The heatmap of renal progenitor cell surface markers (*CD24, CD133, and CD106*) expressed in urine-derived renal progenitors. **(C)** Comparison of gene expression values of urine-derived renal progenitors (UM51) with hREPCs in a scatter plot confirms the mesenchymal phenotype of urine-derived renal progenitors, i.e. expression of Vimentin (*VIM*) and expression of E-cadherin (*CDH1*) in hREPCs. **(D)** Expressed genes (det-$p < 0.05$) in urine-derived renal progenitors (sample UM51) and hREPCs are compared in the Venn diagram. **(E)** The 10 most over-represented GO BP-terms in 566 UM51 genes include triglyceride homeostasis and kidney development and in 438 hREPCs genes include chloride transmembrane transport. **(F)** The 10 most over-represented GO BP-terms in the up- and down-regulated genes in UM51 in comparison to hREPCs are shown. The complete dataset is presented in Supplemental Table S4.

derived renal progenitors-iPSCs with ESCs and HFF-iPSCs. Most genes (12092) are expressed in common in all cell types while 150 genes are expressed exclusively in urine derived renal progenitors-iPSCs. The genes expressed exclusively in one cell type were further analysed for over-representation of GO terms. The treemap summarizing

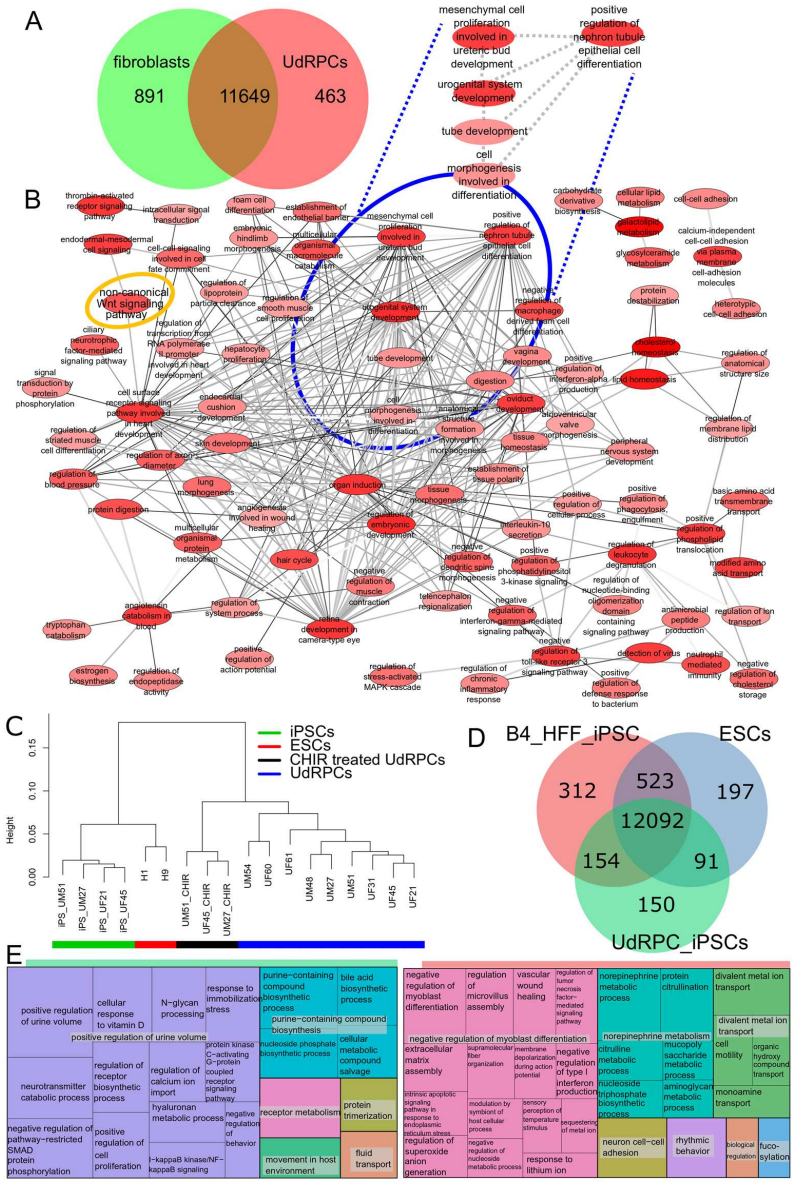


Figure 4. In-depth bioinformatic analysis of urine-derived renal progenitors and urine-derived renal progenitors derived-iPSCs. **(A)** Expressed genes (det-p < 0.05) in urine-derived renal progenitors (UdRPCs) and fibroblasts are compared in a venn diagram. Most genes are expressed in common (11649), 463 genes are expressed exclusively in urine-derived renal progenitors and 891 in fibroblasts. The subsets and urine-derived renal progenitors GOs are presented in supplemental_table_S4. **(B)** The gene ontology network was generated with the tools REVIGO and Cytoscape and summarizes the GO terms of category Biological Process (BP) over-represented in the 463 genes expressed exclusively in urine-derived renal progenitors. Several general developmental terms emerged, e.g. "organ induction". Specific renal-related terms including "urogenital system development" are marked with a blue ellipse. GOs are represented by the network nodes with the intensity of red indicating the significance of over-representation of a GO term. The edges refer to similarities between the GO terms. **(C)** The dendrogram shows a clear separation of urine-derived renal progenitors, differentiated urine-derived renal progenitors (CHIR treated UdRPCs of UM51, UM27, and UF45) (black bar), ESCs (H1 and

H9, red bar) and urine-derived renal progenitors-iPSCs (green bar). **(D)** Venn diagram of HFF-iPSCs, urine-derived renal progenitors-iPSCs (UdRPCs_iPScs) and ESCs. **(E)** GO terms of 150 genes expressed exclusively in urine-derived renal progenitors-iPSCs indicate that these iPSCs retain the memory of renal origin. In the treemap for the HFF-iPSCs the GO-BP terms of the 312 over-represented genes of the exclusive gene set are summarized. The most significant group is associated with negative regulation of myoblast differentiation including genes *DDIT3*, *MBNL3*, *TGFB1*, *ZFHX3* pointing at the fibroblast origin of these iPSCs. The entire dataset is presented in Supplemental Table S7.

the GO terms of category BP over-represented in the 150 genes expressed exclusively in urine derived renal progenitors-iPSCs (Fig. 4E) indicates that these iPSCs retain a memory of their kidney origin. In addition to the largest most significant group- positive regulation of urine volume, it consists of other renal-related GO terms (e.g. calcium transport, vitamin D). Stem-cell-related and developmental terms such as positive regulation of cell proliferation are due to their pluripotent nature. Within the treemap summarizing the GO-BP terms over-represented in the 312 genes expressed exclusively in HFF-iPSCs, the largest most significant group is associated with negative regulation of myoblast differentiation, thus pointing at the fibroblast origin of these iPSCs (Fig. 4E). Furthermore, within the treemap summarizing the GO-BP terms over-represented in the 197 genes expressed exclusively in ESCs, the largest most significant group is associated with negative regulation of astrocyte differentiation- hinting at their known propensity to differentiate into the ectodermal lineage (Supplemental Fig. S4).

WNT pathway activation by GSK3β inhibition induces differentiation of urine-derived renal progenitors into renal epithelial proximal tubular cells. To differentiate three independent renal progenitors preparations, the cells were treated with 10 μM CHIR99021 (WNT pathway activation by GSK3β inhibition) for 2 days and morphological changes from fibroblastic to elongated tubular shape were observed (Fig. 5A). In the Venn diagram, expressed genes (det-$p < 0.05$) in untreated urine-derived renal progenitors are compared to renal progenitors treated with CHIR99021. Genes expressed in common amounts to 11790, of these 2491 are upregulated in the CHIR99021 treatment ($p < 0.05$, ratio > 1.33) and 2043 are down-regulated ($p < 0.05$, ratio < 0.75) (Fig. 5B, Supplemental Table S8). Among the upregulated genes, 27 are considered "novel" (gene symbol starting with "LOC"), 21 among the down-regulated genes and 98 among the non-regulated genes (Supplemental Table S8). The heatmap based on the top 20 regulated genes shows a clear separation between untreated and treated cells (Fig. 5C). Amongst the up-regulated genes, the associated KEGG pathways include WNT-signaling (*AXIN2*, *JUN*, *NKD1*) (Supplemental Fig. S5). Over-representation analysis of the up-regulated genes and their associated KEGG pathways identified protein processing in endoplasmic reticulum as highly significant and several signalling pathways such as mTOR, Insulin, p53, AMPK and TNF. Over-representation analysis of the down-regulated genes and associated KEGG pathways revealed cell cycle, cellular senescence, focal adhesion, FoxO, ErbB and thyroid hormone signalling. Interestingly Hippo pathway was regulated in both undifferentiated and differentiated renal cells (Fig. 5D).

Regulation of self-renewal and differentiation in urine-derived renal progenitor cells. Further to the transcriptome analyses, a real-time PCR revealed downregulation of the stem cell self-renewal associated gene *CD133* and activated expression of the nephrogenesis-associated gene *BMP7* after CHIR99021 stimulation (Fig. 6A). Since FGF signaling is also crucial for maintaining self-renewal, we compared the transcriptome of differentiated cells (CHIR99021 treated) and progenitor cells to investigate the effect of CHIR99021 stimulation on FGF-signaling with respect to the genes from FGF and FGFR family; BMP7 and BMP4 from the BMP family. We detected an upregulation of FGF2 and FGF7 in undifferentiated renal progenitors (Fig. 6B). To validate this, we disrupted FGF signaling using fibroblast growth factor receptor (FGFR) inhibitor SU5402, and observed morphological changes (Supplemental Fig. S6). Interestingly, downregulation of the key renal transcription factor *SIX2* was detected in both the CHIR99021 and SU5402 treated cells (Fig. 6C). Furthermore, to identify the self-renewal regulators and pathways in urine-derived renal progenitor cells, a protein-protein-interaction network was generated. The network of the 40 proteins, encoded by the 20 most significantly up- and down-regulated genes between CHIR99021 treated and untreated urine-derived renal progeitors (Fig. 5C) indentified JUN as a major hub – in terms of having most connections to other proteins in the network. However, in the WNT-signaling pathway JUN is at the end of a downstream cascade from GSK3β, including further downstream targets- AXIN2 and CTNNB1. The genes encoding these proteins were differentially regulated by the CHIR99021 treatment (green nodes) (Fig. 6D). Several communities with more interactions within the community than to other communities can be detected in the network via community clustering of the network via edge-betweenness includes JUN (red), GSK3β / AXIN2 / CTNNB1 (green), LATS2 (yellow), EGFR (pink) (Fig. 6E). To analyze the effect of WNT activation on the TGFβ-SMAD pathway, Western blot analysis was performed to detect phosphorylation levels of SMAD 2/3 and SMAD 1/5/8 in UF45, UM51 and UM27. In the differentiated cells (urine-derived renal progenitors after CHIR treatment) a decreased level of phosphorylated SMAD 2/3 and increased levels of phosphorylated SMAD 1/5/8 were observed (Fig. 6F).

Discussion

Although relatively few renal cells are shed under healthy conditions compared to dysfunctional conditions[10,55], we were able to isolate, culture and expand urine cells from healthy donors. Here we describe urine as a reliable, non-invasive, robust and cheap source of renal stem cells, in contrast to amniotic fluid or kidney biopsies[56,57]. Urine derived stem cells can be expand from a single clone with high proliferation potency[37,58]. We

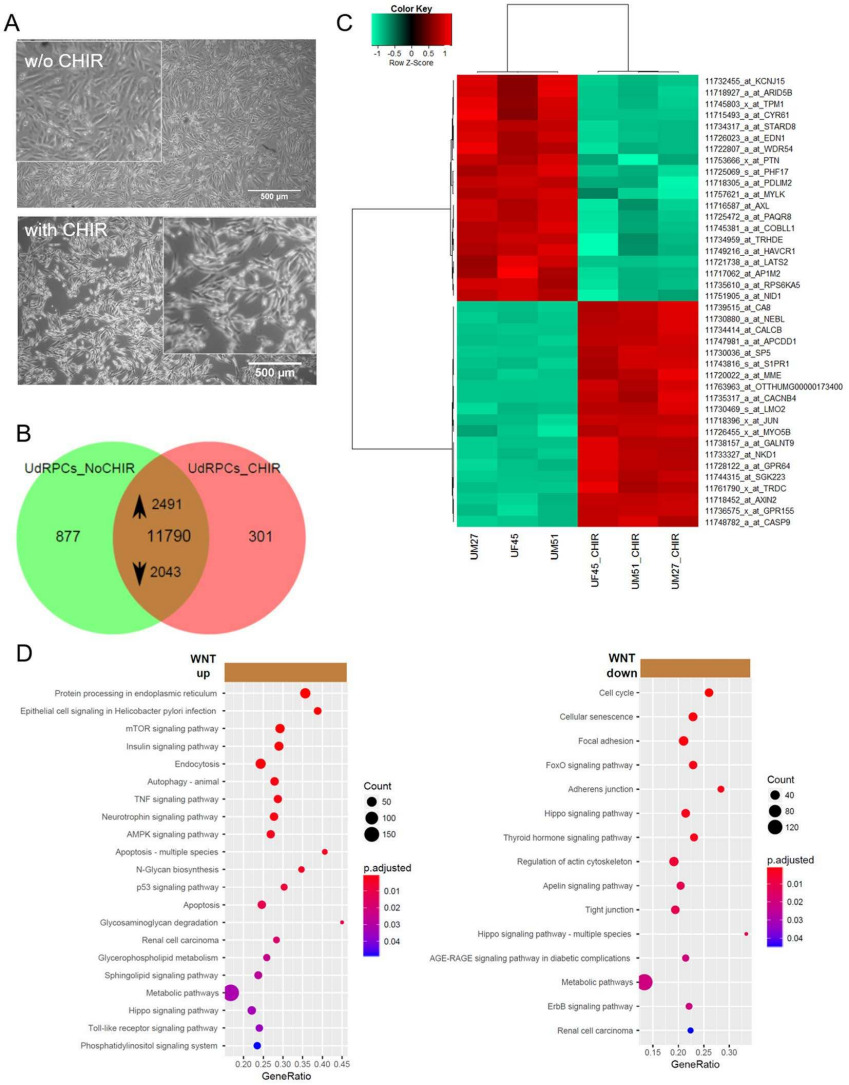


Figure 5. Supplementation of urine-derived renal progenitors with the GSK-3β inhibitor. **(A)** Activation of WNT signalling by supplementation with GSK-3β-inhibitor CHIR99021 led to differentiation into renal epithelial proximal tubular cells. **(B)** In the Venn diagram, expressed genes (det-$p < 0.05$) in untreated urine-derived renal progenitors (UdRPCs_NoCHIR) are compared to urine-derived renal progenitors treated with the GSK-3β-inhibitor CHIR99021 (UdRPCs_CHIR). Among the 11790 genes expressed in both conditions, 2491 are up-regulated in the CHIR99021 treatment ($p < 0.05$, ratio > 1.33) and 2043 down-regulated ($p < 0.05$, ratio < 0.75). **(C)** Heatmap of 3 independent urine-derived renal progenitor cell preparations with and without CHIR treatment. **(D)** Over-representation analysis of the up-regulated genes and associated KEGG pathways revealed protein processing in endoplasmic reticulum as highly significant and several signalling and metabolic pathways including mTOR, Insulin, p53 and TNF. Over-representation analysis of the down-regulated genes in KEGG pathways identified cell cycle, cellular senescence, focal adhesion, FoxO and adherens junction as most significant. Supplemental Table S8 provides the full list of regulated genes and associated pathways.

propose naming these cells as urine-derived renal progenitor cells, because they can be kept in culture for almost 12 passages whilst maintaining expression of the self-renewal associated proteins- SIX2, CITED1, CD133, C-KIT, TRA-1-60, TRA-1-81 and SSEA4 as has been shown by others[37,56]. Despite the expression of a subset of

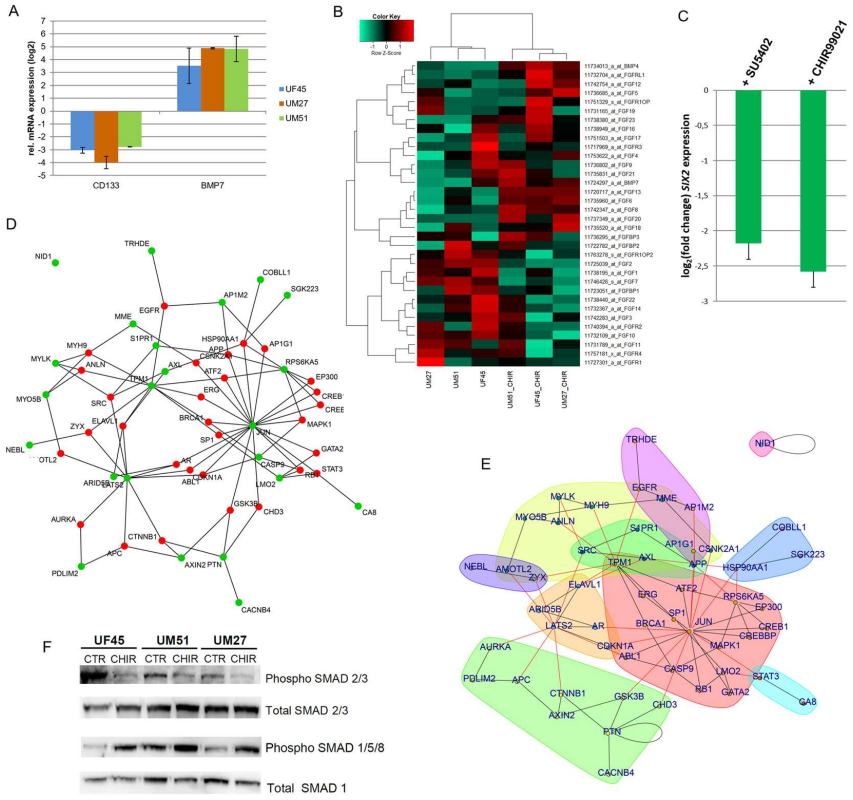


Figure 6. Regulation of self-renewal and differentiation in urine-derived renal progenitors. **(A)** Real-time PCR-based confirmation of down-regulation of *CD133* and activated expression of *BMP7* after CHIR stimulation. **(B)** Effect of CHIR99021 stimulation on FGF-signaling and BMP (BMP7 and BMP4) signaling. The heatmap depicts FGF signaling associated genes up and down regulated upon CHIR treatment of the urine-derived renal progenitors. **(C)** Downregulation of *SIX2* expression in differentiated urine-derived renal progenitors upon WNT stimulation using the GSK-3β-inhibitor CHIR99021 and blocking of FGF signaling using the FGF receptor inhibitor SU5402. **(D)** JUN is a major hub of protein interaction networks of urine-derived renal progenitors treated with CHIR. Based on the Biogrid database protein interaction networks were constructed from the set of the most highly regulated 40 genes either up- or down in the urine-derived renal progenitors treated with CHIR. The selected genes used to connect to the network with interactions from the Biogrid database are marked in green, genes added as Biogrid interactions are marked in red. Induction of WNT leading to GSK3B inhibition is reflected by the connection of GSK3B to JUN and to AXIN2 which is connected to CTNNB1 (β-catenin) – these all downstream targets of GSK3B in the WNT-signaling pathway. **(E)** Community clustering of the network identified several communities: JUN (red), GSK3B/AXIN2/CTNNB1 (green), LATS2 (yellow), EGFR (pink). Black lines refer to edges within a community, red lines to edges between different communities. **(F)** Western blot analysis of the phosphorylated levels of SMAD 2/3 and SMAD 1/5/8 in undifferentiated and differentiated UF45, UM51 and UM27.

pluripotency-associated factors, these renal progenitor cells do not express OCT4, SOX2 and NANOG- which are key pluripotency-regulating transcription factors[59,60]. Further evidence in support of the lack of OCT4 expression is our observed fully methylated CpG dinucleotides within the OCT4 promoter in the UM51 cells. Urine-derived renal progenitors are in fact bon-fide MSCs- i.e. they express VIM and not CDH1, adhere to plastic surfaces, express CD73, CD90 and CD105 and not the hematopoietic markers CD14, CD20, CD34, and CD45. Typical of MSCs, urine-derived renal progenitors can be differentiated into osteoblasts, chondrocytes and adipocytes[56,57,61]. They also secrete a plethora of cytokines and growth factors- such as EGF, GDF, PDGF and Serpin E1[62]. The multipotent features of urine-derived renal progenitors make these cells promising for studying nephrogenesis and in the future regenerative therapy of kidney-associated diseases.

Urine-derived renal progenitor cells express key renal progenitor-regulatory proteins SIX2, CITED1 and WT1 indicating they originate from the kidney as described from others[13,14,27,57,63]. Unmethylated CpG islands within

the 5′- regulatory region of the SIX2 gene confirm their progenitor status. Nuclear-localized SIX2 expression is critical for maintaining self-renewal of renal stem cell populations and has been described to co-localize with CITED1[25]. We observed CITED1 expression in both the nucleus and cytoplasm, this is supported by a subcellular fractionation study that demonstrated an abundant portion of CITED1 localized in the cytoplasm whereas only 5% were expressed in the nucleus[64]. As CITED1 is a cell cycle-dependent transcriptional co-factor and contains a nuclear export signal domain, the subcellular localization might be dependent on its phosphorylation status. We previously, showed CITED1 and WT1 expression in the nucleus and cytoplasm of amniotic fluid cells of renal origin[56]. Here, we observed that subpopulations of adult urine-derived progenitor cells express WT1 in the nucleus and in some cases, both nuclear and cytoplasmic localization was observed. Wt1 has nucleocytoplasmic shuttling activity, however, the shuttling of Wt1 between the nucleus and cytoplasm might regulate the activity as a transcription factor as a result of interaction with the cargo protein importins α1 and β[65,66].

In line with our study, urine-derived renal progenitor cells have been described to express the surface marker CD24, CD106, and CD133[16]. However, we observed variabilities in the numbers of cells expressing these markers between preparations. This variance might be due to the origin of the urine- shed cells in the adult[4,6,7,67]. For example, CD24 and CD133 positive cells have been found in renal tubules and the renal capsule, but CD106^{+} cells are only present in the renal capsule.

Furthermore, urine-derived renal progenitors transport albumin[56,68]. The albumin filtration pathway partly takes place in the kidney and the presence of albumin in urine is used as a marker for cell functionality as described by the endo/exocytosis of albumin in kidney[69]. The GOs derived from the exclusively expressed genes in urine-derived renal progenitors (compared to HFF1) unveiled renal system development- related terms. To overcome the lack of reference human kidney biopsy-derived renal progenitors, we performed a meta-analysis comparing our data to nephron progenitor cell transcriptome datasets downloaded from NCBI GEO. The analyses revealed that our urine-derived renal progenitors share a high level of similarity with other human nephron progenitors at the transcriptome level[52–54]. Moreover, the GOs from the urine cell derived-iPSC exclusive genes-set, in contrast to pluripotent stem cells, identified terms related to renal function therefore implying the preservation of their kidney origin. As the conservation of tissue of origin in iPSCs might be linked to epigenetic memory[17,70], urine-derived renal progenitors as well as corresponding-iPSCs, especially with known CYP2D6 status, might be advantageous for differentiation into renal cells, modelling kidney-related diseases, nephrotoxicity studies and regenerative medicine[55].

Dissecting the gene regulatory mechanisms that drive human renal progenitor growth and differentiation *in vitro* represents the key step for translation but remains a challenge due to the absence of well-characterised primary urine derived stem cells. Here we have shown that urine-derived renal cells are a self-renewing stem cell population unlike the kidney biopsy-derived hREPCs which are differentiated renal epithelial cells. To demonstrate that urine-derived renal progenitors can maintain self-renewal when cultured under undifferentiation conditions but yet retain the potential for epithelial differentiation and nephrogenesis, we induced active WNT signalling, by treatment with the GSK-3β inhibitor- CHIR99021. The differentiated cells adopted an elongated tubular morphology and reduced proliferation as also shown for human ESC and iPSC derived renal epithelial cells[71–73]. Although WNT pathway activation induced an epithelial phenotype, we did not see a dramatic increase in *CDH1* expression at the time point and dose used but rather activation of *CDH-3* expression (8.86 fold). *Cdh-3*, a gene encoding a member of the cadherin superfamily, functions in epithelial cell morphogenesis in *Caenorhabditis elegans*[74] an event which is poorly understood in human nephrogenesis. Furthermore, the correlation co-efficient (0.941) of the WNT-induced differentiated UM51 with hREPCs is further evidence in support of the cellular identify of the UM51 differentiated cells.

In line with our previously published observations in amniotic fluidic-derived renal cells, the down-regulated expression of *SIX2*, *WT1*, *CD133* and upregulated expression of *BMP7* induced the loss of self-renewal[56]. Global transcriptome analyses also revealed the down-regulation of 2043 genes some of which are associated with pathways such as cell cycle, FoxO, Hippo and ErbB signalling. The Hippo pathway which is composed of WNT target genes such as *LATS2, AXIN2* and *CTNNB1* have been reported to regulate epithelialization of nephron progenitors[75,76].

We detected differential expression 40 genes in which 20 most significantly up- and down-regulated between WNT-induced differentiated and self-renewing urine-derived renal progenitors. Amongst the genes up regulated in the CHIR99021 treated cells are the WNT targets- *AXIN2*, *JUN* and *NKD1* known to be associated with WNT signalling. Interestingly, a protein interaction network identified JUN as a major hub connected to GSK3β and interlinked with ATF2, STAT3, GATA2 and MAPK1. In a mouse model, it has been reported that Bmp7 phosphorylates Jun and Atf2 via Jnk signalling which promote the proliferation of mouse nephron progenitors[33]. This indeed might be contradictory to our observed elevated expression of *BMP7* upon WNT induced differentiation of urine-derived renal progenitor cells- i.e. suppression of *BMP7* expression is needed to maintain self-renewal in urine-derived renal progenitor cells. However, in line with our results, during *in vitro* differentiation (mesenchymal to epithelial transition) of human renal cell line TK173, BMP7 is required for the activation of E-Cadherin and WNT4 expression[77]. Since, SMADs are a target of MAPK particularly of JNK, both BMPs and TGFβ can activate the SMAD circuit[78,79]. Both activation of the WNT pathway and inhibition of FGF signalling led to the down-regulation of the key renal progenitor self-renewal associated transcription factor SIX2 and the up-regulated expression of BMP7. This is in line with the reported interactions of BMP and FGF signalling during nephrogenesis[80–82]. Fibroblast growth factor signaling is essential for *in vivo* renal development as well as *in vitro* cultivation and maintenance of nephron progenitor cells as demonstrated by mouse model experiments where blocking of FGF receptors led to aberrant nephrogenesis[25,81].

Based on the present study and our previously published data in human amniotic fluid-derived renal cells[56], we propose that similar to self-renewal in human pluripotent stem cells[60,83], urine-derived renal progenitor cells maintain self-renewal by active FGF signalling leading to phosphorylated TGFβ- SMAD2/3. In contrast,

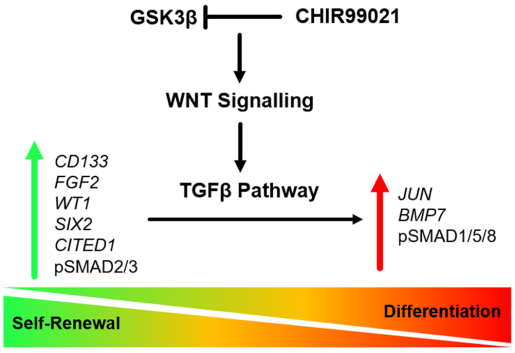


Figure 7. WNT/β-catenin and TGFβ pathway-mediated cell fate decisions in urine-derived renal progenitors. Self-renewal (inactive WNT/β-catenin signalling and active TGF-β/SMAD2/3 signalling) is maintained by elevated expression of the renal progenitor markers *SIX2*, *WT1*, *CITED1*, *CD133*, in addition to phospho-SMAD2/3 and FGF2 resulting in and down regulated expression of *BMP7*. In contrast, activation of WNT/β-catenin signalling induces upregulated expression of JUN and *BMP7* leading to activation of phospho-SMAD1/5/8, downregulated expression of *WT1, SIX2, CITED1, FGF2, CD133* and ultimately exit of self-renewal.

activation of WNT/β-catenin signalling leads to an upregulation of *JUN* and *BMP7* leading to activation of SMAD1/5/8 signalling and exit of self-renewal by downregulation of *WT1, SIX2*, *CITED1*, and *CD133* expression. To surmise, we derived a hypothetic scheme of the WNTβ catenin and TGFβ pathway-mediated cell fate decisions in urine-derived renal progenitor cells. This simplistic model is depicted in Fig. 7.

Comparing self-renewal of renal progenitor cells in both human (urine-derived renal progenitors) and mouse, it is clear that an intricate balance is needed between SIX2, WT1, CITED1 expression and Wnt/β-catenin activity in order to determine the cell fate of nephron progenitor cells[24,31,34,56]. Furthermore, it remains to be determined if indeed there exist subtle human and mouse differences in the gene regulatory network needed to maintain a self-renewing renal progenitor pool in both species and we believe that human urine-derived renal progenitor cells as described here will facilitate these studies.

Materials and methods

Ethics statement. In this study, urine samples were collected with the informed consent of the donors and the written approval (Ethical approval Number: 5704) of the ethical review board of the medical faculty of Heinrich Heine University, Düsseldorf, Germany. All methods were carried out in accordance with the approved guidelines. Medical faculty of Heinrich Heine University approved all experimental protocols.

Isolation, culture, and differentiation of urine-derived renal progenitor cells. Urine samples were collected from 10 healthy donors with diverse age, gender and ethnicity (Supplemental Table S1). Isolation and expansion of the urine-derived renal progenitors followed the previously established protocols[37,41]. For differentiation of the urine derived renal progenitors, 10 µM CHIR99021 was added to the cell culture medium for 2 days. Adult kidney biopsy derived primary human renal epithelial cells (hREPCs) (C-12665, Promo Cell, Germany) were used as control. To inhibit FGF signaling in urine derived renal progenitors, 15 µM SU5402 was added to the cell culture medium for 2 days.

Immunofluorescence staining. Immunofluorescence study was performed as described previously[84]. To analyse expression of specific markers, at 80% confluence, attached urine-derived renal progenitor cells of four individuals (Passages 4-5) were fixed with 4% PFA (Polysciences Inc., USA) for 15 min at room temperature (RT) and washed three times in PBS and permeability was increased using 1% Triton X-100 for 5 min. Next, for blocking we used: 10% normal goat serum (NGS; Sigma-Aldrich Chemie GmbH, Germany), 0.5% Triton X-100, 1% BSA (Sigma-Aldrich Chemie GmbH, Germany) and 0.05% Tween 20 (Sigma-Aldrich Chemie GmbH, Germany) in PBS for 2 h. The cells were incubated with primary antibodies (Supplemental Table S2) for 1 h at RT followed by three washes with PBS. Thereafter, the corresponding secondary Cy3-labeled, Alexa Fluor-555 or Alexa Fluor 488-labeled antibodies (Thermo Fisher Scientific, USA) and Hoechst 33,258 dye (Sigma-Aldrich Chemie GmbH, Germany) or DAPI (Southern Biotech, USA) were added. A fluorescence microscope (LSM700; Carl Zeiss Microscopy GmbH, Germany) was used for taking the pictures. All pictures were processed with the ZenBlue 2012 Software Version 1.1.2.0. (Carl Zeiss Microscopy GmbH, Germany).

CYP2D6 genotyping/phenotyping and Albumin endocytosis assay. CYP2D6 genotyping and phenotyping of five individuals (randomly selected) were carried out by CeGat GmbH Germany using genomic DNA. The CYP2D6 variant assay reveals the pharmacogenetics (PGx) profile of an individual's genotype and

phenotype based on tested pharmacogenetics markers. The assay identifies and discriminates individuals with poor, normal, intermediate and ultra-rapid metabolizing activity[85]. Albumin endocytosis assay was performed as described before[56]. For detailed description, see supplemental materials and methods.

Immunophenotyping by flow cytometry. At 90% confluence, adherent cells at passage 3–5 from UM27, UF31, UF45 and UM51 were detached from 6-well plates by incubation in TrypLE (Thermo Fisher Scientific, USA) at 37 °C. Then the cell samples were subjected to fluorescence-activated cell sorting (FACS) in order to specifically select for MSC cell surface markers, renal stem cell transcription factor $SIX2^+$ cells, and renal stem cell surface markers. Unstained cells and IgG isotype served as control for each cell sample. Dead cells and debris were gated on a two physical parameter dot plot followed by the exclusion of doublets by using pulse processing. Finally, the experimental cells positive subpopulation was gated. Sorting was done using CytoFLEX cell sorter (Beckman Coulter, USA), BD FACSCanto (BD Biosciences, Germany) and CyAn ADP (Beckman Coulter, USA). Histograms were generated using the Summit 4.3.02 software.

The analysis of MSC-associated cell surface marker expression of urine-derived renal progenitors was performed using MSC Phenotyping Kit (Miltenyi Biotec GmbH, Germany) according to the manufacturer's instructions and as described before[84]. For the pluripotency-associated markers, TRA-1-60, TRA-1-81, and SSEA4 dye-coupled antibodies were used (anti-TRA-1-60-PE, human (clone: REA157), number 130-100-347; anti-TRA-1-81-PE, human (clone: REA246), number 130-101-410, and anti-SSEA-4-PE, human (clone: REA101), number 130-098-369; Miltenyi Biotec GmbH, Germany). For $SIX2^+$ cell sorting, after blocking with Human TruStain blocking solution (Biolegend, USA) (5 µL each) for 10 min at RT, the cells (10^4 cells/condition) were stained with anti-m-SIX2 (Abnova, Taiwan) primary antibody overnight at 4 °C. After 3 times washing with the Permwash buffer (Invitrogen, Germany), mouse Alexa-Fluor 488 was conjugated by incubating 30 min at RT in the dark. For the renal stem cell surface markers anti-CD24-FITC (Sigma-Aldrich Chemie GmbH, Germany), VCAM-1/CD106-PE (R&D systems, USA) and CD133-APC (R&D systems, USA) were used according to manufacturer instructions for flow cytometry analysis. Briefly, after blocking and washing, cells were centrifuged at 300 x g for 10 min. 5 µl of antibody solution (1:50 dilution) were added to the cell suspension and the samples were incubated in the dark at 4 °C for 10 min. Cells were washed afterwards, and stored in 4% PFA at 4 °C until analysis.

Differentiation into adipocytes, chondrocytes and osteoblasts. Differentiation of urine-derived renal progenitors into adipocytes, chondrocytes and osteoblasts were tested using the StemPro Adipogenesis, Chondrogenesis, and Osteogenesis differentiation Kits (Gibco, Life Technologies, USA) as described before[56,84]. After the differentiation periods, cells were fixed using 4% PFA for 20 min at RT and stained with Oil Red-O for detecting adipocytes, Alcian Blue for chondrocytes, and Alizarin Red S for osteoblasts. A light microscope was used for imaging.

Western blot analysis. For protein extraction, cells were harvested and lysed in RIPA buffer (Sigma-Aldrich Chemie GmbH, Germany) supplemented with complete protease and phosphatase inhibitors cocktail (Roche, Switzerland). The lysates were separated on a 4–20% Bis-Tris gel and blotted onto a 0.45 µm nitrocellulose membrane (GE Healthcare Life Sciences, Germany). The membranes were then blocked with 5% skimmed milk in Tris-Buffered Saline Tween (TBS-T) and incubated overnight with the respective primary antibodies: Total Smad 1 (1:1000, TBS-T 5% BSA; CST, USA), phospho Smad 1/5/8 (1:1000, TBS-T 5% milk; CST, USA), Total Smad 2/3 (CST, 1:1000, TBS-T 5% BSA), and phospho Smad 2/3 (1:1000, TBS-T 5% milk; CST, USA). After incubation with the appropriate secondary antibodies, signals were acquired with a Fusion-FX7 imaging system.

Bisulfite genomic sequencing. Bisulfite sequencing was performed following bisulfite conversion with the EpiTec Kit (Qiagen, Germany). Primers were designed after excluding pseudogenes or other closely related genomic sequences which could interfere with specific amplification by amplicon and primer sequences comparison in BLAT sequence database (https://genome.ucsc.edu/FAQ/FAQblat.html). See supplemental materials and methods for full description.

Generation of iPSCs from urine-derived renal progenitors. Four urine-derived renal progenitor cell samples were reprogrammed into iPSCs (four lines) using an integration-free episomal based transfection system without pathway inhibition. Briefly, urine-derived renal progenitor cells were nucleofected with two plasmids pEP4 E02S ET2K (Addgene plasmid #20927) and pEP4 E02S CK2M EN2L (Addgene plasmid #20924) expressing a combination of pluripotency factors including OCT4, SOX2, LIN28, c-MYC, KLF4, and NANOG using the Amaxa 4D-Nucleofector Kit (Lonza, Swiss) according to the manufacturer's guidelines and as described previously[41]. Please see supplemental materials and methods for full description.

Quantitative RT-PCR analysis. RNA was isolated using the Direct-zol RNA MiniPrep Kit (Zymo Research, USA) according to provider guidelines. After checking the quality of mRNA, 500 ng of RNA were used for complementary DNA synthesized with the TaqMan Reverse Transcription Kit (Applied Biosystems, USA). Real-time quantitative PCR was performed in technical triplicates with Power SYBR Green Master Mix (Life Technologies, USA), 12.5 ng cDNA per sample and 0.6 µM primers on a VIIA7 (Life Technologies, USA) machine. Mean values were normalized to levels of the housekeeping gene ribosomal protein L37A calculated by the $2-\Delta\Delta Ct$ method. Primers used were purchased from MWG (Supplementary Table S3).

Microarray data analyses. Total RNA (1 µg) preparations were hybridized on the PrimeView Human Gene Expression Array (Affymetrix, Thermo Fisher Scientific, USA) at the core facility Biomedizinisches Forschungszentrum (BMFZ) of the Heinrich Heine University Düsseldorf. The raw data was imported into the

R/Bioconductor environment[86] and further processed with the package affy[87] using background-correction, logarithmic (base 2) transformation and normalization with the Robust Multi-array Average (RMA) method. For full details, please see supplemental materials and methods.

KEGG pathway, GO and network analysis. Gene ontology (GOs) terms were analysed within the Bioconductor environment employing the package GOstats[88]. GOs of category Biological Process (BP) were further summarized with the REVIGO tool[89] to generate treemaps populating the parameter for allowed similarity with tiny = 0.4. GO networks were generated from the REVIGO tool in xgmml format and imported into Cytoscape[90]. For full details, see supplemental materials and methods.

Activated WNT pathway associated protein interaction network. The network was constructed from the 20 most significantly up- and down down-regulated genes between CHIR99021 treatment and untreated controls. Genes were ranked by the limma-p-value and passed the criteria: detection p-value < 0.05 for the dedicated condition, ratio < 0.75 or ratio > 1.33, limma-p-value < 0.05. The resulting 40 genes were marked as green nodes in the network. Interacting proteins containing at least one protein coded by the 40 genes were retrieved from BioGrid version 3.4.161[91]. The plot of the interactions network was drawn employing the R package network[92]. See supplemental materials and methods for full description.

Additional materials and methods. For the materials and methods of the culture supernatant analysis, analysis of cell proliferation, meta-analysis for comparison of urine-derived renal progenitors to public nephron progenitor data sets and cell lines used in this study and culture conditions, please see supplemental materials and methods.

Statistics. All data are presented as arithmetic means ± standard error of mean. At least 3 independent experiments were used for the calculation of mean values. P values of < 0.05 were considered significant.

Data availability

All raw and processed data used in this study have been archived in NCBI gene expression omnibus under GEO accession number GSE128281. (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128281).

Received: 24 June 2019; Accepted: 31 December 2019;
Published online: 20 January 2020

References

1. https://ncdalliance.org/the-international-society-of-nephrology-isn (accessed: 23th October, 2019)
2. Rota, C., Morigi, M. & Imberti, B. Stem Cell Therapies in Kidney Diseases: Progress and Challenges. *Int. J. Mol. Sci.* **20**, 2790 (2019).
3. Marcheque, J., Bussolati, B., Csete, M. & Perin, L. Concise Reviews: Stem Cells and Kidney Regeneration: An Update. *Stem Cell Transl. Med.* **8**, 82–92 (2019).
4. Huling, J. & Yoo, J. J. Comparing adult renal stem cell identification, characterization and applications. *Journal of Biomedical Science*. **24**, Article number: 32 (2017).
5. Lazzeri, E. *et al.* Regenerative potential of embryonic renal multipotent progenitors in acute renal failure. *J. Am. Soc. Nephrol.* **18**, 3128–38 (2007).
6. Sagrinati, C. *et al.* Isolation and characterization of multipotent progenitor cells from the Bowman's capsule of adult human kidneys. *J. Am. Soc. Nephrol.* **17**, 2443–56 (2006).
7. Angelotti, M. L. *et al.* Characterization of Renal Progenitors Committed Toward Tubular Lineage and Their Regenerative Potential in Renal Tubular Injury. *Stem Cells.* **30**, 1714–1725 (2012).
8. Chen, C. Y. *et al.* Extracellular vesicles from human urine-derived stem cells prevent osteoporosis by transferring CTHRC1 and OPG. *Bone Res.* **7**, 18 (2019).
9. Choi, J. Y. *et al.* Potency of Human Urine-Derived Stem Cells for Renal Lineage Differentiation. *Tissue Eng. Regen. Med.* **14**, 775–785 (2017).
10. Ingelfinger, J. R. Nephrogenic adenomas as renal tubular outposts. *N. Engl. J. Med.* **29**, 684–686 (2002).
11. Zhang, D. *et al.* Urine-derived stem cells: A novel and versatile progenitor source for cell-based therapy and regenerative medicine. *Genes. Dis.* **1**, 8–17 (2014).
12. Ringsrud, K. M. Casts in the Urine Sediment. *Laboratory Med.* **32**, 191–193 (2001).
13. Da Sacco, S. *et al.* Direct Isolation and Characterization of Human Nephron Progenitors. *Stem Cell Transl. Med.* **6**, 419–433 (2017).
14. Harari-Steinberg, O. *et al.* Identification of human nephron progenitors capable of generation of kidney structures and functional repair of chronic renal disease. *EMBO Mol. Med.* **5**, 1556–1568 (2013).
15. Kreidberg, J. A. WT1 and kidney progenitor cells. *Organogenesis* **6**, 61–70 (2010).
16. Lazzeri, E. *et al.* Human Urine-Derived Renal Progenitors for Personalized Modeling of Genetic Kidney Disorders. *J. Am. Soc. Nephrol.* **26**, 1961–1974 (2015).
17. Manaph, N. P. A., Al-Hawwas, M., Bobrovskaya, L., Coates, P. T. & Zhou, X. F. Urine-derived cells for human cell therapy. *Stem Cell Res. Ther.* **9**, 189 (2018).
18. Bharadwaj, S. *et al.* Multipotential differentiation of human urine-derived stem cells: Potential for therapeutic applications in urology. *Stem Cell* **31**, 1840–1856 (2013).
19. O'Brien, L. L. *et al.* Transcriptional regulatory control of mammalian nephron progenitors revealed by multi-factor cistromic analysis and genetic studies. *PLoS Genet.* **14**, e1007181 (2018).
20. Ramalingam, H. *et al.* Disparate levels of beta-catenin activity determine nephron progenitor cell fate. *Dev. Biol.* **440**, 13–21 (2018).
21. Xu, J., Liu, H., Park, J. S., Lan, Y. & Jiang, R. Osr1 acts downstream of and interacts synergistically with Six2 to maintain nephron progenitor cells during kidney organogenesis. *Development* **141**, 1442–1452 (2014).
22. Kann, M. *et al.* WT1 targets Gas1 to maintain nephron progenitor cells by modulating FGF signals. *Development.* **142**, 1254–1266 (2015).
23. Motamedi, F. J. *et al.* WT1 controls antagonistic FGF and BMP-pSMAD pathways in early renal progenitors. *Nat. Commun.* **5**, 01–14 (2014).
24. Park, J. S. *et al.* Six2 and Wnt regulate self-renewal and commitment of nephron progenitors through shared gene regulatory networks. *Dev. Cell* **23**, 637–651 (2012).

25. Mari, C. & Winyard, P. Concise Review: Understanding the Renal Progenitor Cell Niche *In Vivo* to Recapitulate Nephrogenesis *In Vitro*. *Stem Cell Trans. Med.* **4**, 1463–1471 (2015).
26. Lawlor, K. T. *et al.* Nephron progenitor commitment is a stochastic process influenced by cell migration. *Elife.* **8pii**, e41156 (2019).
27. Murphy, A. J. *et al.* SIX2 and CITED1, markers of nephronic progenitor self-renewal, remain active in primitive elements of Wilms' tumor. *J. Pediatr. Surg.* **47**, 1239–1249 (2012).
28. Kobayashi, A. *et al.* Identification of a multipotent self-renewing stromal progenitor population during mammalian kidney organogenesis. *Stem Cell Reports.* **3**, 650–662 (2014).
29. Little, M. H. & Kairath, P. Does Renal Repair Recapitulate Kidney Development? *J. Am. Soc. Nephrol.* **28**, 34–46 (2017).
30. James, R. G., Kamei, C. N., Wang, Q., Jiang, R. & Schultheiss, T. M. Odd-skipped related 1 is required for development of the metanephric kidney and regulates formation and differentiation of kidney precursor cells. *Development.* **133**, 2995–3004 (2006).
31. Kobayashi, A. *et al.* Six2 defines and regulates a multipotent self-renewing nephron progenitor population throughout mammalian kidney development. *Cell Stem Cell.* **3**, 169–181 (2008).
32. Self, M. *et al.* Six2 is required for suppression of nephrogenesis and progenitor renewal in the developing kidney. *EMBO J.* **25**, 5214–5228 (2006).
33. Blank, U. *et al.* BMP7 promotes proliferation of nephron progenitor cells via a JNK-dependent mechanism. *Development.* **136**, 3557–3566 (2009).
34. Lindström, N. O. *et al.* Conserved and Divergent Features of Human and Mouse Kidney Organogenesis. *J. Am. Soc. Nephrol.* **3**, 785–805 (2018a).
35. Lindström, N. O. *et al.* Progressive Recruitment of Mes.enchymal Progenitors Reveals a Time-Dependent Process of Cell Fate Acquisition in Mouse and Human Nephrogenesis. *Dev. Cell.* **45**, 651–660 (2018b).
36. Thiagarajan, R. D. *et al.* Identification of anchor genes during kidney development defines ontological relationships, molecular sub-compartments and regulatory pathways. *PLoS One.* **6**, e17286 (2011).
37. Zhou, T. *et al.* Generation of induced pluripotent stem cells from urine. *J. Am. Soc. Nephrol.* **22**, 1221–1228 (2011).
38. Zhou, T. *et al.* Generation of human induced pluripotent stem cells from urine samples. *Nat. Protoc.* **7**, 2080–2089 (2012).
39. Uhm, K. O. *et al.* Generation of human induced pluripotent stem cells from urinary cells of a healthy donor using a non-integration system. *Stem Cell Res.* **21**, 44–46 (2017).
40. Sauer, V. *et al.* Human Urinary Epithelial Cells as a Source of Engraftable Hepatocyte-Like Cells Using Stem Cell Technology. *Cell Transplant.* **25**, 2221–2243 (2016).
41. Bohndorf, M. *et al.* Derivation and characterization of integration-free iPSC line ISRM-UM51 derived from SIX2-positive renal cells isolated from urine of an African male expressing the CYP2D6 *4/*17 variant which confers intermediate drug metabolizing activity. *Stem Cell Res.* **25**, 18–21 (2017).
42. Little, M. H., Combes, A. N. & Takasato, M. Understanding kidney morphogenesis to guide renal tissue regeneration. *Nat. Rev. Nephrol.* **12**, 624–635 (2016).
43. Chuah, J. K. C. & Zink, D. Stem cell-derived kidney cells and organoids: Recent breakthroughs and emerging applications. *Biotechnol. Adv.* **35**, 150–167 (2017).
44. van den Berg, C. W. *et al.* Renal Subcapsular Transplantation of PSC-Derived Kidney Organoids Induces Neo-vasculogenesis and Significant Glomerular and Tubular Maturation *In Vivo*. *Stem Cell Reports.* **10**, 751–765 (2018).
45. Rauch, C. *et al.* Differentiation of human iPSCs into functional podocytes. *PLoS One.* **13**, e0203869 (2018).
46. Tanigawa, S., Taguchi, A., Sharma, N., Perantoni, A. O. & Nishinakamura, R. Selective *In Vitro* Propagation of Nephron Progenitors Derived from Embryos and Pluripotent Stem Cells. *Cell Rep.* **15**, 801–813 (2016).
47. Morizane, R. *et al.* Nephron organoids derived from human pluripotent stem cells model kidney development and injury. *Nat. Biotechnol.* **33**, 1193–1200 (2015).
48. Morizane, R. & Bonventre, J. V. Generation of nephron progenitor cells and kidney organoids from human pluripotent stem cells. *Nat. Protoc.* **12**, 195–207 (2017).
49. Hariharan, K. *et al.* Parallel generation of easily selectable multiple nephronal cell types from human pluripotent stem cells. *Cell Mol. Life. Sci.* **76**, 179–192 (2019).
50. Phipson, B. *et al.* Evaluation of variability in human kidney organoids. *Nat. Methods* **16**, 79–87 (2019).
51. Wu, H. *et al.* Comparative Analysis and Refinement of Human PSC-Derived Kidney Organoid Differentiation with Single-Cell Transcriptomics. *Cell Stem Cell.* **23**, 869–881 (2018).
52. Lindgren, D. *et al.* Isolation and characterization of progenitor-like cells from human renal proximal tubules. *Am. J. Pathol.* **178**(2), 828–37 (2011).
53. Da Sacco, S. *et al.* Direct Isolation and Characterization of Human Nephron Progenitors. *Stem Cell Transl. Med.* **6**(2), 419–433 (2017).
54. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE75949, (accessed on 23th October 2019)
55. Mulder, J. *et al.* Generation of infant- and pediatric-derived urinary induced pluripotent stem cells competent to form kidney organoids. *Pediatr Res.* (2019).
56. Rahman, M. S. *et al.* The presence of human mesenchymal stem cells of renal origin in amniotic fluid increases with gestational time. *Stem Cell Res. Ther.* **9**, 113 (2018).
57. Arcolino, F. O. *et al.* Urine of Preterm Neonates as a Novel Source of Kidney Progenitor Cells. *J. Am. Soc. Nephrol.* **27**, 2762–2770 (2016).
58. Romagnani, P. *et al.* Next generation sequencing and functional analysis of patient urine renal progenitor-derived podocytes to unravel the diagnosis underlying refractory lupus nephritis. *Nephrol. Dial. Transplant.* **31**, 1541–1545 (2016).
59. Babaie, Y. *et al.* Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. *Stem Cells.* **25**, 500–510 (2007).
60. Greber, B., Lehrach, H. & Adjaye, J. Silencing of core transcription factors in human EC cells highlights the importance of autocrine FGF signaling for self-renewal. *BMC Dev. Biol.* **7**, 46 (2007a).
61. Dominici, M. *et al.* Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy.* **8**, 315–317 (2006).
62. Zhang, D., Wei, G., Li, P., Zhou, X. & Zhang, Y. Urine-derived stem cells: A novel and versatile progenitor source for cell-based therapy and regenerative medicine. *Genes. Dis.* **1**, 08–17 (2014).
63. Combes, A. N. *et al.* Haploinsufficiency for the Six2 gene increases nephron progenitor proliferation promoting branching and nephron number. *Kidney Int.* **93**, 589–598 (2018).
64. Shi, G. *et al.* The Transcriptional Activity of CITED1 Is Regulated by Phosphorylation in a Cell Cycle-dependent Manner. *J. Biol. Chem.* **281**, 27426 (2006).
65. Vajjhala, P.R. *et al.* The Wilms' tumour suppressor protein, WT1, undergoes CRM1-independent nucleocytoplasmic shuttling. *FEBS Letters*. **554** (2003).
66. Depping, R. *et al.* Nuclear Transport of Wilms' Tumour Protein Wt1 Involves Importins α and β. *Cell Physiol. Biochem.* **29**, 223–232 (2012).
67. Shrestha, S. *et al.* Human renal tubular cells contain CD24/CD133 progenitor cell populations: Implications for tubular regeneration after toxicant induced damage using cadmium as a model. *Toxicol. Appl. Pharmacol.* **331**, 116–129 (2017).
68. Tojo, A. & Kinugasa, S. Mechanisms of glomerular albumin filtration and tubular reabsorption. *Int. J. Nephrol.* **2012**, 481–520 (2012).
69. Birn, H. & Christensen, E. I. Renal albumin absorption in physiology and pathology. *Kidney Int.* **69**, 440–9 (2006).
70. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature.* **467**, 285–290 (2010).

71. Takasato, M. *et al*. Directing human embryonic stem cell differentiation towards a renal lineage generates a self-organizing kidney. *Nat. Cell Biol.* **16**, 118–126 (2014).
72. Takasato, M. *et al*. Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis. *Nature.* **526**, 564–568 (2015).
73. Brown, A. C., Muthukrishnan, S. D. & Oxburgh, L. A synthetic niche for nephron progenitor cells. *Dev. Cell.* **34**, 229–241 (2015).
74. Pettitt, J., Wood, W. B. & Plasterk, R. H. cdh-3, a gene encoding a member of the cadherin superfamily, functions in epithelial cell morphogenesis in Caenorhabditis elegans. *Development.* **122**, 4149–4157 (1996).
75. Reginensi, A. *et al*. A critical role for NF2 and the Hippo pathway in branching morphogenesis. *Nat. Commun.* **7**, 12309 (2016).
76. McNeill, H. & Reginensi, A. Lats1/2 Regulate Yap/Taz to Control Nephron Progenitor Epithelialization and Inhibit Myofibroblast Formation. *J. Am. Soc. Nephrol.* **28**, 852–861 (2017).
77. Zeisberg, M., Shah, A. A. & Kalluri, R. Bone morphogenic protein-7 induces mesenchymal to epithelial transition in adult renal fibroblasts and facilitates regeneration of injured kidney. *J. Biol. Chem.* **280**, 8094–100 (2005).
78. Derynck, R. & Zhang, Y. E. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature* **425**, 577–584 (2003).
79. Yamaguchi, K. *et al*. Identification of a member of the MAPKKK family as a potential mediator of TGF-beta signal transduction. *Science.* **270**, 2008–2011 (1995).
80. Brown, A. C. *et al*. FGF/EGF signaling regulates the renewal of early nephron progenitors during embryonic development. *Development.* **138**, 5099–112 (2011).
81. Godin, R. E., Robertson, E. J. & Dudley, A. T. Role of BMP family members during kidney development. *Int. J. Dev. Biol.* **43**, 405–411 (1999).
82. Dudley, A. T., Godin, R. E. & Robertson, E. J. Interaction between FGF and BMP signaling pathways regulates development of metanephric mesenchyme. *Genes. Dev.* **13**, 1601–1613 (1999).
83. Greber, B., Lehrach, H. & Adjaye, J. Fibroblast growth factor 2 modulates transforming growth factor beta signaling in mouse embryonic fibroblasts and human ESCs (hESCs) to support hESC self-renewal. *Stem Cells.* **25**, 455–464 (2007b).
84. Spitzhorn, L. S. *et al*. Isolation and Molecular Characterization of Amniotic Fluid-Derived Mesenchymal Stem Cells Obtained from Caesarean Sections. *Stem Cell Int.* **2017**, 5932706 (2017).
85. Pinto, N. & Dolan, M. E. Clinically relevant genetic variations in drug metabolizing enzymes. *Curr. Drug. Metab.* **12**, 487–97 (2011).
86. Gentleman, R. C. *et al*. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
87. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics.* **20**, 307–315 (2004).
88. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinforma. Oxf. Engl.* **23**, 257–258 (2007).
89. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE.* **6**, e21800 (2011).
90. Shannon, P. *et al*. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research.* **13**, 2498–2504 (2003).
91. Chatr-Aryamontri, A. *et al*. The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* **45**, D369–D379 (2017).
92. Butts, C. network: A Package for Managing Relational Data in R. *J. Stat. Software.* **24**, 01–36 (2008).

Acknowledgements

James Adjaye acknowledges support from the medical faculty of the Heinrich Heine University- Düsseldorf, Germany. In addition, James Adjaye and Md Shaifur Rahman acknowledges support from the German Academic Exchange Service (DAAD-91607303). We acknowledge Dipl. Ing Katharina Raba, Institute of Transplantation Diagnostics and Cell Therapeutics, Heinrich Heine University, Germany for her kind assistance with the FACS analyses.

Author contributions

:Conception: J.A.; Cell culture experiments: M.S.R., L.S.S., F.A., M.B. and J.A.; Real Time PCR: F.A., L.E.; Bisulphite sequencing: L.E.; Bioinformatics: W.W. and J.A.; Proliferation assays and Western blot analyses: S.M.; Derivation and characterisation of iPSCs: M.B., N.G. and A.N.; Immunostaining and FACS analysis: M.B., L.S.S., A.N., M.S.R., L.N.; Writing of manuscript: M.S.R., W.W., L.S.S., L.N. and J.A.; Final edits: J.A. All authors approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-57723-2.

Correspondence and requests for materials should be addressed to J.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2.1.8 Meta-analysis of human prefrontal cortex reveals activation of GFAP and decline of synaptic transmission in the aging brain.

Acta Neuropathol Commun. 2020 Mar 5;8(1):26. doi: 10.1186/s40478-020-00907-8.

Wruck W(1), Adjaye J(2).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Moorenstr.5, 40225, Düsseldorf, Germany. (2)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Moorenstr.5, 40225, Düsseldorf, Germany. James.Adjaye@med.uni-duesseldorf.de.

Despite ongoing research efforts, mechanisms of brain aging are still enigmatic and need to be elucidated for a better understanding of age-associated cognitive decline. The aim of this study is to investigate aging in the prefrontal cortex region of human brain in a meta-analysis of transcriptome datasets. We analyzed 591 gene expression datasets pertaining to female and male human prefrontal cortex biopsies of distinct ages. We used hierarchical clustering and principal component analysis (PCA) to determine the influence of sex and age on global transcriptome levels. In sex-specific analysis we identified genes correlating with age and differentially expressed between groups of young, middle-aged and aged. Pathways and gene ontologies (GOs) over-represented in the resulting gene sets were calculated. Potential causal relationships between genes and between GOs were explored employing the Granger test of gene expression time series over the range of ages. The most outstanding results were the age-related decline of synaptic transmission and activated expression of glial fibrillary acidic protein (GFAP) in both sexes. We found an antagonistic relationship between calcium/calmodulin dependent protein kinase IV (CAMK4) and GFAP which may include regulatory mechanisms involving cAMP responsive element binding protein (CREB) and mitogen-activated protein kinase (MAPK, alias ERK). Common to both sexes was a decline in synaptic transmission, neurogenesis and an increased base-level of inflammatory and immune-related processes. Furthermore, we detected differences in dendritic spine morphogenesis, catecholamine signaling and cellular responses to external stimuli, particularly to metal (Zinc and cadmium) ions which were higher in female brains.

DOI: 10.1186/s40478-020-00907-8 PMCID: PMC7059712 PMID: 32138778

URL: https://link.springer.com/article/10.1186/s40478-020-00907-8

Contribution: Wasco Wruck performed bioinformatic analyses and wrote the manuscript.

Wruck and Adjaye *Acta Neuropathologica Communications* (2020) 8:26
https://doi.org/10.1186/s40478-020-00907-8

Acta Neuropathologica Communications

RESEARCH Open Access

Meta-analysis of human prefrontal cortex reveals activation of GFAP and decline of synaptic transmission in the aging brain

Wasco Wruck and James Adjaye*

Abstract

Despite ongoing research efforts, mechanisms of brain aging are still enigmatic and need to be elucidated for a better understanding of age-associated cognitive decline. The aim of this study is to investigate aging in the prefrontal cortex region of human brain in a meta-analysis of transcriptome datasets. We analyzed 591 gene expression datasets pertaining to female and male human prefrontal cortex biopsies of distinct ages. We used hierarchical clustering and principal component analysis (PCA) to determine the influence of sex and age on global transcriptome levels. In sex-specific analysis we identified genes correlating with age and differentially expressed between groups of young, middle-aged and aged. Pathways and gene ontologies (GOs) over-represented in the resulting gene sets were calculated. Potential causal relationships between genes and between GOs were explored employing the Granger test of gene expression time series over the range of ages. The most outstanding results were the age-related decline of synaptic transmission and activated expression of glial fibrillary acidic protein (GFAP) in both sexes. We found an antagonistic relationship between calcium/calmodulin dependent protein kinase IV (*CAMK4)* and *GFAP* which may include regulatory mechanisms involving cAMP responsive element binding protein (CREB) and mitogen-activated protein kinase (MAPK, alias ERK). Common to both sexes was a decline in synaptic transmission, neurogenesis and an increased base-level of inflammatory and immune-related processes. Furthermore, we detected differences in dendritic spine morphogenesis, catecholamine signaling and cellular responses to external stimuli, particularly to metal (Zinc and cadmium) ions which were higher in female brains.

Keywords: Prefrontal cortex, Aging, Sex-specific, Meta-analysis, Transcriptome

Introduction

Mechanisms associated with time-dependent physical decline, i.e., *aging* are complex and despite its omnipresence in biological organisms our understanding of it is still not complete. Recently, López-Otín et al. proposed nine hallmarks of aging [35] into: (i) the four causative hallmarks *Genomic instability*, *Telomere attrition*, *Epigenetic alterations* and *Loss of proteostasis*, (ii) the three hallmarks as response to damage *De-regulated nutrient sensing*, *Mitochondrial dysfunction* and *Cellular senescence* and (iii) the two integrative hallmarks *Stem cell exhaustion* and *Altered intercellular communication* which as a result from the others are responsible for functional decline. Roles of oxidative stress in aging have been manifested in a large body of publications, e.g. [7] but have also been challenged recently [26]. Hekimi et al. do not consider reactive oxygen species (ROS) as the primary cause of aging but rather as a mediator of stress response to age-dependent damage. Brink et al. propose the metabolic stability theory of aging, which postulates that the aging process depends on maintaining stable concentrations of reactive oxygen species (ROS) and other critical metabolites [7].

* Correspondence: James.Adjaye@med.uni-duesseldorf.de
Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Moorenstr.5, 40225 Düsseldorf, Germany

The rate of aging varies in an organ-specific manner ([7]). The observation that adult brains do not grow further led to the notion that neurogenesis declines with age, this however remains contentious. The dogma of no postnatal neurogenesis was rejected as far back as the 1990s by studies dating back to the 1960s [2] also finding neurogenesis in adult brains [32]. However, the level of neurogenesis in the adult brain is at a low level and therefore the established conclusions that most of the cognitive tasks are controlled by synapsis dynamics still holds true. We previously described that aging is the most important factor in the etiology of Late-onset-Alzheimer's disease (LOAD) and identified gene-regulatory networks in hippocampus correlating with metabolic instability and oxidative stress [53]. The distinction between disease-associated and aging-related phenotypes is important. Whilst AD and Mild cognitive impairment (MCI) are associated with the loss of neurons, age-related cognitive impairment (ARCI) is not characterized by neuronal loss but rather by changes in the dynamics of synapses. Synapse dynamics depend on three types of dendritic spines: stubby, thin and mushroom spines [25]. Mushroom spines are considered responsible for long-term memory while thin spines are considered to arrange synapses for flexible cognitive tasks [6]. Morrison et al. reported that these thin spines were found to be reduced during aging and their density showed the highest correlation to performance on a cognitive task (DNMS: delayed nonmatching-to-sample) in non-human primates [39]. Mostany et al. reported that old mice possess the same spine density but a higher stability of spines when compared to mature mice and therefore might imply that age-related deficits in sensory perception are rather associated with alterations in the size and stability of spines and boutons than with the loss of synases [40]. Dendritic spine density can be increased by estradiol [52], thus, hormonal balance plays an important role in cognitive performance. Furthermore, age-associated decrease in hormone levels can also be considered as a reason for cognitive decline in elderly persons. In females after menopause, cognitive performance has been shown to be improved by estrogen-replacement therapies [45]. The body of literature is much smaller for males but regulation of dendritic spine density by testosterone has also been reported [20].

The role of astrocytes in healthy and diseased brain is gaining more attention due to the observation that astrocytes play major roles in synaptic transmission, information processing, energy supply and control of blood flow [46]. Analogous to inflammation, the re-activation of astrocytes in response to neural injury is indispensable, and uncontrolled reactivation can be detrimental- ultimately leading to brain disease. In this study, we investigated changes in the transcriptomes, associated pathways and gene ontologies in the brains of males and females during aging by a meta-analysis of 591 datasets from prefrontal cortex biopsies taking into account sex-specific differences and commonalities.

Materials and methods

Data analysis

Transcriptome datasets of 591 pre-frontal cortex biopsies measured on several Affymetrix microarray platforms and via rnaSeq (Illumina HiSeq) were downloaded from NCBI GEO (Supplementary Table 1). These datasets originate from studies by Narayan et al. [41], Barnes et al. [4], Lu et al. [36], Lanz et al. [34], Chen et al. [10], Hagenauer et al. [24] and Cheng et al. [11]. Table 1 shows the distribution of the datasets between female and male samples and over age groups. All data were read into R/Bioconductor [21] and normalized together employing the R package inSilicoMerging [48] parametrized to use the Combat method in order to remove batch effects. For the generation of dendrograms, genes were filtered with a coefficient of variation greater than 0.1 and afterwards subjected to hierarchical cluster analysis using complete linkage as an agglomeration method and Pearson correlation as similarity measure. Colour bars indicative of aging or sex were added to the dendrograms via the R package dendextend [19]. Genes for Principal Component Analysis (PCA) were filtered analogously as for dendrograms and afterwards the PCA of their logarithmic (base 2) gene expression was calculated using the R function *prcomp*. Based on the PCA,

Table 1 Characteristics of PFC datasets, distribution of female and male samples and in age groups

Dataset	Age < 30	Age 30–65	Age > 65	Male	Female	M/F	Total
GSE21138	6	19	4	24	5	4.80	29
GSE21935	2	5	12	10	9	1.11	19
GSE53890	8	12	21	20	21	0.95	41
GSE53987	1	17	1	10	9	1.11	19
GSE71620	52	316	52	332	88	3.77	420
GSE92538	3	37	15	35	20	1.75	55
GSE106669	1	3	4	8	8	1.00	8
Total	73	409	109	439	160	2.74	591

gene expression was predicted employing the function *predict* and the prediction for the first two components was plotted with age- or sex -specific colour schemes. The proportions of variance of the first two principal components were determined using the attribute named *importance* from the summary function of the *prcomp* object. The screeplot was generated with the *plot*() method of the *prcomp* object. Genes with most influence on the principal components were found with the function *get_pca_var*() from the R package factoextra [30] and plotted with the package corrplot [50].

Detection of age-associated gene expression

For each gene-*g*, the Pearson correlation with age r_{gxa} was calculated with the R function *cor()* using the normalized logarithmic (base 2) gene expression as x and the age of the corresponding individual as a. The corresponding p-value was determined via the R function *cor.test()*. The values were calculated separately for male and female prefrontal cortex gene expression. Plots over age were generated from the logarithmic normalized expression data with the R functions *matplot()* and *matlines()* fitting a third order polynomial model to the gene expression data for the regression curve.

Pathway and GO over-representation plots

Over-represented KEGG pathways were calculated employing the R built-in hypergeometric test. Pathway annotations were downloaded from the KEGG database in March 2018 [29]. Over-represented GOs were determined via the R package GOStats [16]. The n most significantly over-represented KEGG pathways and GOs ($n = 20$) were plotted in a special dot plot indicating p-value of the hypergeometric test, number of significant genes per pathway/GO and gene ratio (ratio of significant genes to all genes in the pathway/GO) using package ggplot2 [51].

Protein interaction networks

Human protein interactions and interactors of interactors were extracted from the Biogrid database version 3.4.161 [9] using genes significantly correlated and anti-correlated genes (Bonferroni-corrected $p < 0.05$). The resulting complex network was reduced to the shortest paths between the original set via the method get.shortest.paths from R package igraph [12] and was plotted employing community cluster networks identifying communities with more internal than external links via function *cluster-edge-betweenness*.

Time series analysis

In order to identify genes associated with *GFAP*, Pearson correlation coefficient of the expression of all genes to the expression of *GFAP* was calculated. The genes with the highest positive or negative correlation were filtered and subjected to time series analysis. As it was obviously not possible to generate the time series from multiple measurements at the same individual during aging they represent only a model of aging reconstructed from single measurements at multiple individuals. Thus, the measurements include gene expression variability between individuals. In order to smoothen the time series, a polynomial of degree three was fitted to the data. For follow-up analyses a stationary time series was needed. We used the function *ndiffs()* from the R package *forecast* [28] to check the stationarity of the time series and that no further differentiation was needed. The function was parametrized to use the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test with the null hypothesis of a stationary root. We adapted the Granger test which tests causality between time series [22] to test Granger causality between these time series reconstructed from gene expression measured in post-mortem brain biopsies from individuals comprising a full spectrum of ages at death. We test the null hypothesis that the time series g of one gene does not cause the time series h of another gene. This is tested via an auto-regression model of h to which lagged values of g are added so that the null hypothesis is equivalent to test the coefficients b_i for equality to zero:

$$h_t = \sum_{i=1}^{L} a_i h_{t-i} + \sum_{i=1}^{L} b_i g_{t-i} + \varepsilon_t \quad (1)$$

$$H_0 : b_1 = \cdots = b_L = 0 \text{ (gene } h \text{ does not Granger cause gene } g) \quad (2)$$

Here, a_i are coefficients of the auto-regression model of h and b_i coefficients for the added lagged values of g, ε_t is the error. The time series of the expression of these genes during aging compared to the *GFAP* time series were tested for Granger causality with the function grangertest from the R package lmtest [56].

Time series analysis on the GO level

The above described time series analysis was extended in order to uncover relationships between GOs and between genes and GOs. To achieve this, the means of the expression values of genes significantly correlated or anti-correlated with age and associated with a GO were calculated. The time series consisting of these mean values was considered a consensus time series for the dedicated GO. Let A be the set of ages for which data exists and G_{gu} and G_{gd} be the sets of genes significantly correlated and anti-correlated with age from the GO g:

$$G_{gu} = \{genes\ correlated\ (up)\ with\ age\ in\ GO\ g\} \quad (3)$$

$$G_{gd} = \{genes\ anticorrelated\ (down)\ with\ age\ in\ GO\ g\} \quad (4)$$

$$X_u = \{X_{aui}; a \in A; i \in G_{gu}\} \quad (5)$$

$$X_d = \{X_{adi}; a \in A; i \in G_{gd}\} \quad (6)$$

The consensus time series $\overline{X}_{gu}$ and $\overline{X}_{gd}$ for GO g are then:

$$\overline{X}_{gu} = \frac{1}{|G_{gu}|} \sum_{i \in G_{gu}} X_{aui} \quad (7)$$

$$\overline{X}_{gd} = \frac{1}{|G_{gd}|} \sum_{i \in G_{gd}} X_{adi} \quad (8)$$

Granger causality between this GO consensus time series and other significantly over-represented GO consensus time series was determined. Furthermore, Granger causality tests between genes of interest, e.g. *GFAP,* and GO consensus time series were carried out.

Results

Sex differences are more prominent than age differences in prefrontal cortex

Five hundred ninety-one prefrontal cortex (PFC) biopsies-derived transcriptome datasets (Supplementary Table 1) from control donors without diagnosed disease were downloaded from National Center for Biotechnology information (NCBI) Gene Expression Omnibus (GEO). After normalization and batch effect adjustment, the datasets were characterized via Principal component analysis (PCA). The plot of the first two components explaining the highest percentage of variance (Fig. 1a) shows a separation between female (red) and male (blue). Pooled samples containing both sexes are located in the middle between male and female. The dendrogram of male and female transcriptomes essentially confirms the sex effect showing large sex-specific contiguous regions (Fig. 1d). Trying to find reasons for this sex effect, we directly compared male and female transcriptomes and found that the most significantly differentially expressed genes were located on the sex chromosomes (Supplementary Table 2). Based on this, we performed the follow-up analyses in a sex-specific manner. Separate cluster analyses for male and female showed predominantly age-independent clusters with some sub-clusters possessing tendencies for younger or older samples in male (Fig. 1b) as well as female (Fig. 1c).

Sex-specific differential expression between young, middle-aged and old

Differentially down-regulated (ratio < 0.833, $p < 0.05$; Fig. 2a, c, e) and up-regulated (ratio > 1.2, $p < 0.05$; Fig. 2b, d, f) genes were calculated between three age groups and compared in venn diagrams between female (red circles) and male (green circles) prefrontal cortex. Sex-specific age groups contained age younger than 30 (F30, M30), age between 30 and 65 (F30_65, M30_65) and age over 65 (F65, M65). Most genes were differentially expressed between the more distant groups of age > 65 and age < 30 while in the comparisons with the middle-aged group there were fewer genes differentially expressed. This demonstrates continuous long-term changes in gene expression. In general, in the male samples fewer genes were differentially expressed than in females which may partly be due to the different sample numbers in male and female. Thus, except for the comparison of down-regulated middle-aged vs. young (Fig. 2a) more genes found in male biopsies overlapped with female genes than were exclusive in male. This overlap between male and female shows congruency between the sexes thus seeming to contradict the sex-effect found previously in the PCA plot and dendrogram (Fig. 1a, d). An explanation could be that while most genes are expressed similarly in male and female, sex-specific expression is mostly induced by genes on the sex chromosomes. As more detailed functional annotation of genes become available later in this study, analysis carried out so far revealed that *GFAP* is up-regulated with increasing age whereas *ALB1* and *CX3CR1* are down-regulated with age in both sexes. For the complete gene lists corresponding to the venn diagram analyses refer to Supplementary Table 3.

Genes down-regulated during aging are associated with synaptic processes

For each gene the Pearson correlation coefficient and corresponding p-value of its expression with the age of the associated individuals was calculated separately for male and female prefrontal cortex (Supplementary Table 4). Figure 3a shows a plot of the expression of the ten genes most significantly anti-correlated with age in female ranked by the correlation, Fig. 3c analogously in male. *CYP46A1* (F: $r = -0.57$, M: $r = -0.53$) and *RIMS1* (F: $r = -0.58$, M: $r = -0.51$) were among these in both sexes, *CX3CL1* ($r = -0.61$) was lowest in female, *EXPH5* ($r = -0.58$) was lowest in male (Supplementary Table 4). Gene ontologies (GOs) of genes which were most significantly anti-correlated with age (Bonferroni-corrected $p < 0.05$, $r < -0.1$) were analyzed separately for male and female prefrontal cortex. The 20 most significantly over-represented GO terms (GO type Biological process) are shown in dot plots indicating p-value of hypergeometric test, gene count and ratios of genes annotated with the GO term (Fig. 3b for female, Fig. 3d for male). In both sexes, GO terms related to synaptic signaling were found as most significant (F: $p = 1.2E\text{-}19$, M: $p = 8.1E\text{-}21$, Supplementary Table 5, Fig. 3b, d). Numerous neuron-

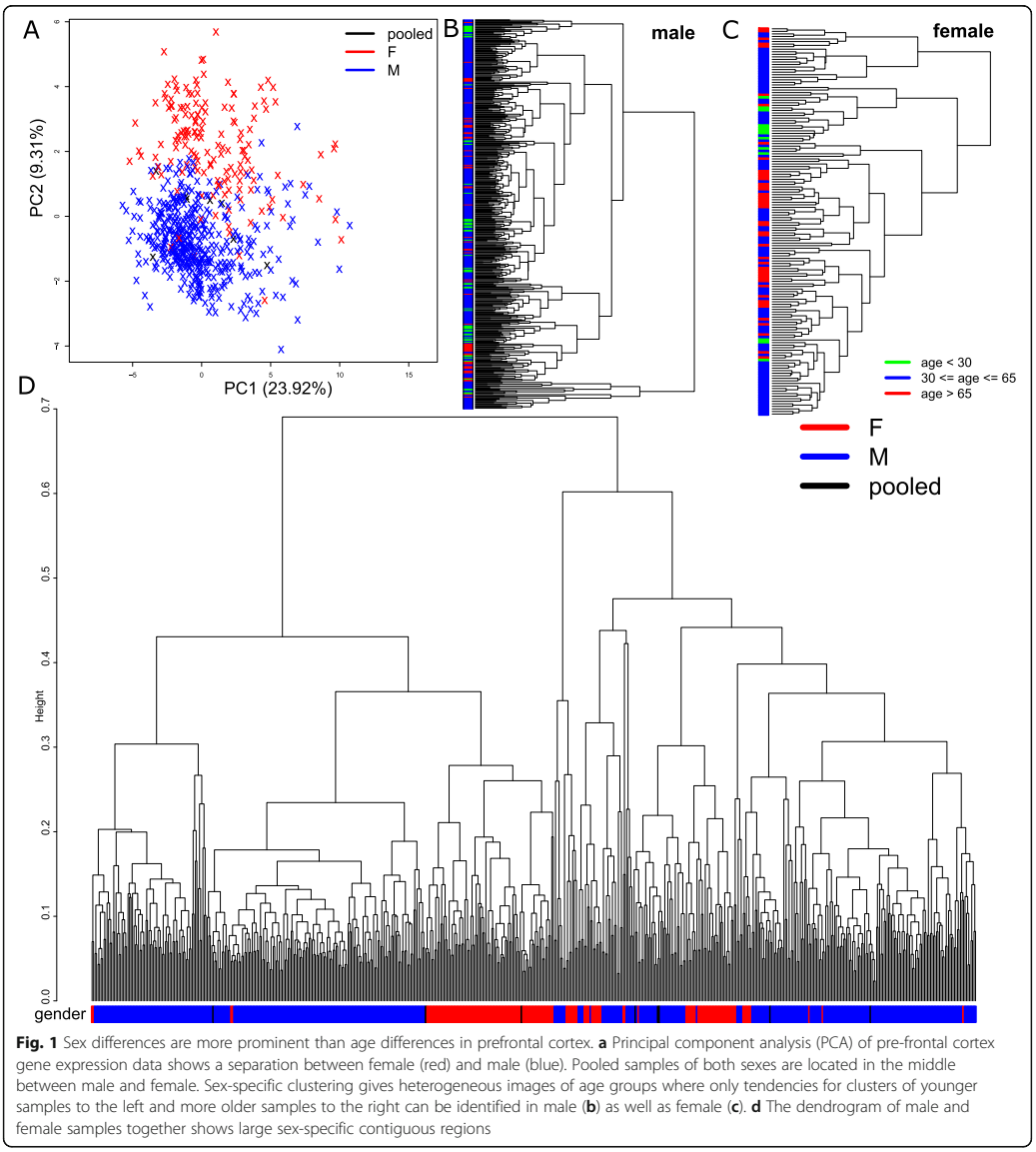


Fig. 1 Sex differences are more prominent than age differences in prefrontal cortex. **a** Principal component analysis (PCA) of pre-frontal cortex gene expression data shows a separation between female (red) and male (blue). Pooled samples of both sexes are located in the middle between male and female. Sex-specific clustering gives heterogeneous images of age groups where only tendencies for clusters of younger samples to the left and more older samples to the right can be identified in male (**b**) as well as female (**c**). **d** The dendrogram of male and female samples together shows large sex-specific contiguous regions

related GO terms were detected as down-regulated with age – amongst these are, *axon development, nervous system development, generation of neurons, glutamate receptor signaling pathway, cell morphogenesis involved in neuron differentiation*. Additionally, further functional groups including hormones, glucocorticoids, catecholamine, neurogenesis and synapse related processes such as Long-Term-Potentiation (LTP), cAMP signaling, dendritic spines, could be identified among the significant GO terms (Table 2). While most of these GO terms provide further detail for the central finding of age-related reduction of synaptic transmission hormones and dendritic spines may be causative. The expression levels of numerous hormones such as estradiol decrease with age

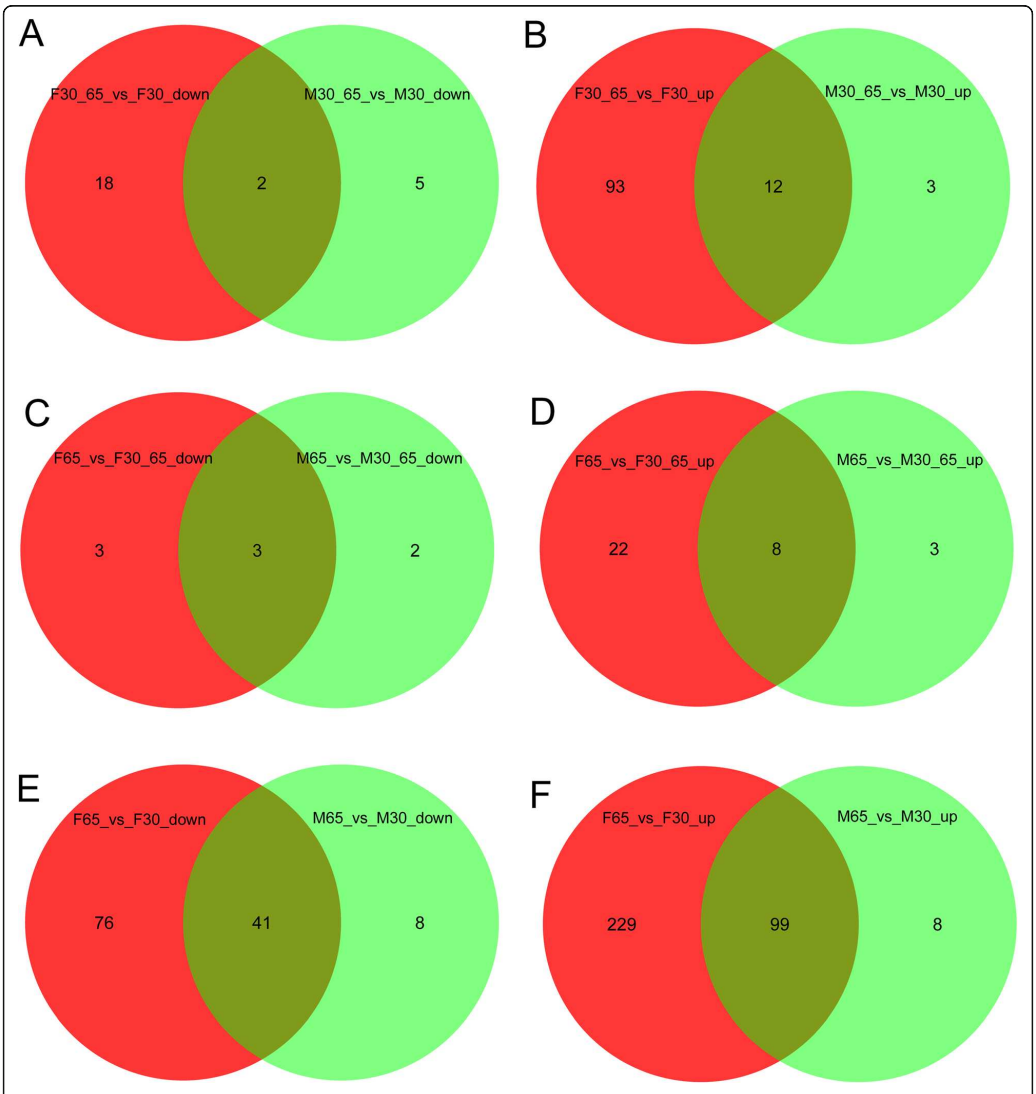


Fig. 2 Most genes were differentially expressed between groups of age > 65 and age < 30. Down-regulated (**a**, **c**, **e**) and up-regulated (**b**, **d**, **f**) genes were compared in venn diagrams between female (red circles) and male (green circles) prefrontal cortex. Age was grouped in a sex-specific way into age < 30 (F30, M30), 30 < age < 65 (F30_65, M30_65) and age > 65 (F65, M65). **a** Genes down-regulated in F30_65 vs. F30 were compared with genes down-regulated in M30_65 vs. M30. **b** Genes up-regulated in F30_65 vs. F30 were compared with genes up-regulated in M30_65 vs. M30. **c** Genes down-regulated in F65 vs. F30_65 were compared with genes down-regulated in M65 vs. M30_65. **d** Genes up-regulated in F65 vs. F30_65 were compared with genes up-regulated in M65 vs. M30_65. **e** Genes down-regulated in F65 vs. F30 were compared with genes down-regulated in M65 vs. M30. **f** Genes up-regulated in F65 vs. F30 were compared with genes up-regulated in M65 vs. M30. Most genes were differentially expressed between the more distant groups of age > 65 and age < 30 while in the comparisons between the adjacent age groups there were fewer genes differentially expressed. This demonstrates continuous long-term changes in gene expression. From the fewer genes differentially expressed in male biopsies most were in common with the female genes

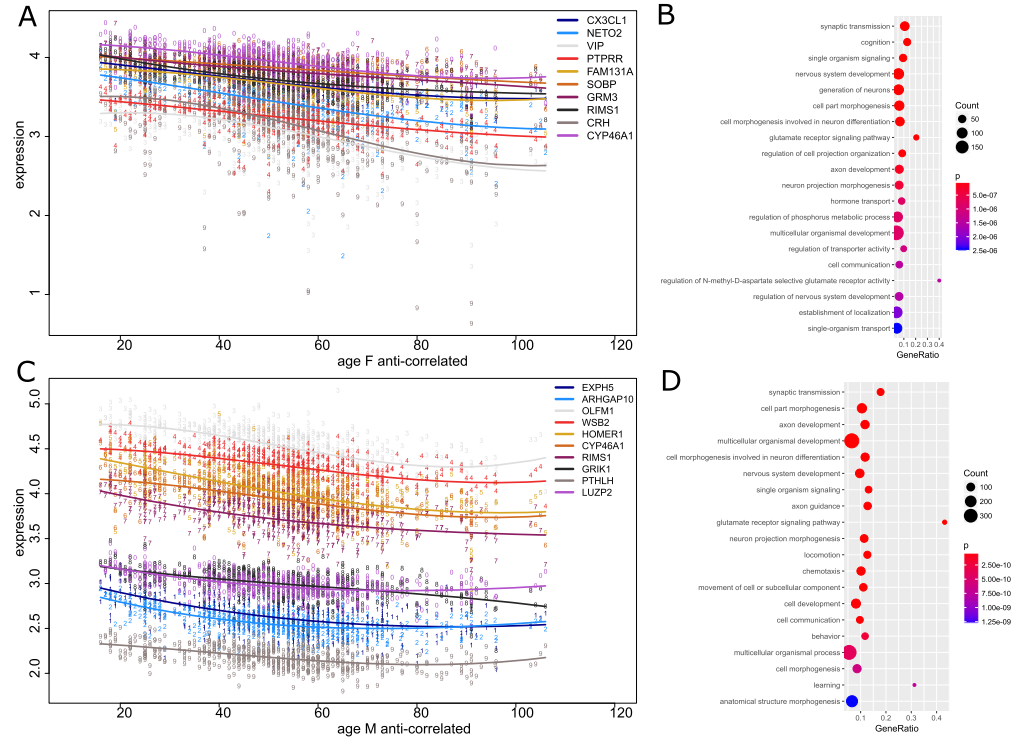


Fig. 3 Genes down-regulated during aging are associated with synaptic processes . Gene ontologies (GOs) of genes which were most significantly anti-correlated with age were analyzed separately for male and female prefrontal cortex. GO terms related to synaptic signaling were found in both sexes

and are known to influence synaptic plasticity by changing the numbers and characteristics of dendritic spines.

Genes up-regulated during aging are associated with the astrocyte marker GFAP and inflammation

Based on the Pearson correlations with age (Supplementary Table 4) the ten genes most significantly correlated with age were plotted in female (Fig. 4a) and male (Fig. 4c). *GFAP* (F: $r = 0.62$, M: $r = 0.55$), FKBP5 (F: $r = 0.62$, M: $r = 0.47$), ITGB4 (F: $r = 0.56$, M: $r = 0.51$) and *ERBB2IP* (F: $r = 0.56$, M: $r = 0.44$) were among these in both sexes, *GFAP* was highest in both female and male (Supplementary Table 4). GOs of genes which were most significantly correlated with age (Bonferroni-corrected $p < 0.05$, $r > 0.1$) were analyzed separately for male and female prefrontal cortex. The dot plots in Fig. 4b (female) and Fig. 4d (male) show the 20 most significantly over-represented GO terms (as in Fig. 3b, d). The GO terms *extracellular matrix organization* and *circulatory system development* and *positive regulation of gene expression* (probably due to selection of upregulated genes) appear in both sexes while the rest of these top 20 terms differ between sexes. Further functional GO term groups shown in Table 3 include *immunity*, *inflammation*, *ROS* and *integrin-associated terms*. The immunity and inflammation-related terms are much more abundant in females, thus implying probable sex-associated regulation of inflammation and immune response during aging.

Aging-related changes in pathways

Sex-specific pathway analysis of genes which were most significantly correlated (Bonferroni-corrected $p < 0.05$, $r > 0.1$) and anti-correlated (Bonferroni-corrected $p < 0.05$, $r < -0.1$) revealed several over-represented KEGG pathways [29]. The dot plots in Fig. 5 show the 20 most significantly over-represented KEGG pathways for each of these four analyses. The full pathway analysis results are provided in Supplementary Table 6. Down-regulation (anti-correlation) with age was associated with various types of

Table 2 Selected groups of significant GO terms overrepresented in genes anti-correlated with age in female and male

Group	Term_female	*P* value_F	Term_male	*P* value_M
Catecholamine	catecholamine uptake involved in synaptic transmission	1.47E-04	catecholamine secretion	5.33E-04
	cellular response to catecholamine stimulus	8.13E-04		
	catecholamine transport	1.47E-03		
	regulation of catecholamine secretion	1.76E-02		
	catecholamine binding	4.99E-02		
Hormone	hormone transport	8.20E-07	hormone transport	6.75E-07
	regulation of hormone secretion	2.56E-06	regulation of hormone secretion	3.64E-06
	peptide hormone secretion	4.88E-05	peptide hormone secretion	1.77E-05
	response to peptide hormone	1.01E-03	response to peptide hormone	2.62E-05
	hormone-mediated apoptotic signaling pathway	4.18E-03	cellular response to hormone stimulus	1.11E-04
	cellular response to hormone stimulus	5.97E-03	positive regulation of peptide hormone secretion	9.68E-03
	negative regulation of peptide hormone secretion	6.89E-03	hormone-mediated apoptotic signaling pathway	1.06E-02
	regulation of intracellular steroid hormone receptor signaling pathway	1.97E-02	thyroid hormone transport	1.72E-02
	neuropeptide hormone activity	2.45E-06	positive regulation of corticosteroid hormone secretion	3.41E-02
			regulation of intracellular steroid hormone receptor signaling pathway	3.90E-02
			cellular response to parathyroid hormone stimulus	4.42E-02
			neuropeptide hormone activity	6.12E-05
			peptide hormone receptor binding	7.63E-03
Corticoid	positive regulation of glucocorticoid receptor signaling pathway	1.94E-05	positive regulation of glucocorticoid receptor signaling pathway	8.12E-05
	corticosteroid receptor signaling pathway	1.17E-03	corticosteroid receptor signaling pathway	6.59E-03
			positive regulation of corticosteroid hormone secretion	3.41E-02
Neurogenesis	positive regulation of neurogenesis	2.84E-05	positive regulation of neurogenesis	1.51E-05
	negative regulation of neurogenesis	3.31E-03		
cAMP	regulation of cAMP biosynthetic process	7.42E-05	regulation of cAMP biosynthetic process	6.39E-07
	negative regulation of cAMP metabolic process	7.43E-04	negative regulation of cAMP metabolic process	5.18E-05
	positive regulation of cAMP metabolic process	1.01E-02	positive regulation of cAMP metabolic process	4.47E-04
			hippocampus development	7.40E-04
			cAMP-mediated signaling	9.50E-04
			negative regulation of cAMP-mediated signaling	3.86E-03
			cAMP catabolic process	4.08E-02
LTP	positive regulation of long-term synaptic potentiation	6.25E-04	long-term synaptic potentiation	5.54E-06
	long-term synaptic potentiation	4.07E-	positive regulation of long-term synaptic potentiation	2.49E-03

Table 2 Selected groups of significant GO terms overrepresented in genes anti-correlated with age in female and male *(Continued)*

Group	Term_female	*P* value_F	Term_male	*P* value_M
		03		
Dendritic spine	negative regulation of dendritic spine development	2.02E-03	dendritic spine morphogenesis	4.75E-05
	dendritic spine organization	3.74E-03	regulation of dendritic spine morphogenesis	2.44E-03
	regulation of dendritic spine morphogenesis	5.59E-03	negative regulation of dendritic spine development	7.75E-03
	dendritic spine development	9.70E-03	positive regulation of dendritic spine morphogenesis	1.03E-02
	positive regulation of dendritic spine morphogenesis	3.38E-02	dendritic spine	4.60E-09
	dendritic spine	5.69E-05	dendritic spine head	2.26E-03
	dendritic spine head	5.73E-04		
	dendritic spine membrane	2.17E-02		

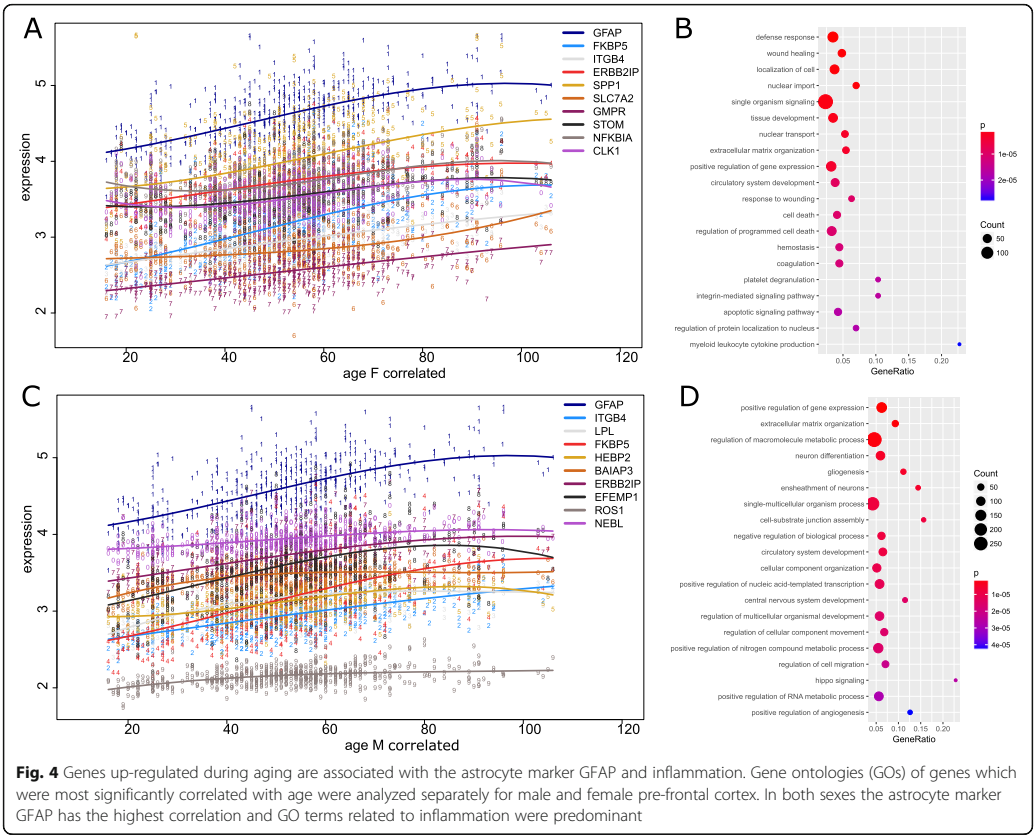


Fig. 4 Genes up-regulated during aging are associated with the astrocyte marker GFAP and inflammation. Gene ontologies (GOs) of genes which were most significantly correlated with age were analyzed separately for male and female pre-frontal cortex. In both sexes the astrocyte marker GFAP has the highest correlation and GO terms related to inflammation were predominant

Table 3 Selected groups of significant GO terms overrepresented in genes correlated with age in female and male

Group	Term_female	*P* value_F	Term_male	*P* value_M
Immunity	immune response	2.16E-04	negative regulation of immune system process	1.25E-02
	regulation of immune system process	2.58E-04		
	regulation of production of molecular mediator of immune response	7.18E-03		
	positive regulation of cytokine production involved in immune response	1.67E-02		
	leukocyte mediated immunity	2.98E-02		
	immune system process	3.53E-02		
	activation of immune response	3.67E-02		
	regulation of innate immune response	4.00E-02		
	immunoglobulin secretion	4.47E-02		
	negative regulation of immune response	4.80E-02		
Inflammation	positive regulation of inflammatory response	6.21E-03	acute inflammatory response	1.69E-02
	regulation of inflammatory response	6.90E-03		
	acute inflammatory response	1.12E-02		
ROS	regulation of reactive oxygen species biosynthetic process	5.79E-04	positive regulation of reactive oxygen species metabolic process	2.63E-03
	positive regulation of reactive oxygen species metabolic process	2.72E-03	regulation of reactive oxygen species biosynthetic process	6.00E-03
	response to oxidative stress	2.22E-02	response to oxidative stress	3.94E-02
	intrinsic apoptotic signaling pathway in response to oxidative stress	3.64E-02		
Integrin-associated terms	integrin-mediated signaling pathway	1.67E-05	integrin-mediated signaling pathway	2.72E-04
	integrin binding	7.38E-03	cell adhesion mediated by integrin	3.53E-02
			integrin binding	2.83E-04

synapses, calcium signaling and long-term-potentiation in both sexes (Fig. 5a, b). To elucidate further causes leading to decline of synaptic transmission pathways *Cortisol synthesis and secretion* (F:p = 0.02,q = 0.1; M:p = 0.0001,q = 0.001), *cAMP signaling* (F:p = 0.05, q = 0.23; M:p = 0.0001, q = 0.001) and *Estrogen signaling* (F:p = 0.03,q = 0.17;M:p = 0.005,q = 0.03) were found (Fig. 5b, Supplementary Table 6A,B). Estrogens have been reported to regulate dendritic spine density [52].

Amonsgt the genes correlated with age, we identified over-represented pathways associated with the extracellular matrix, cytoskeleton and Hippo- and PI3K-Akt –signaling (Fig. 5c, d). For the detailed pathways see (Fig. 5c, d, Supplementary Table 6C, D): *Regulation of actin cytoskeleton* (F:p = 0.001,q = 0.02; M:p = 0.0004,q = 0.09), *Proteoglycans in cancer* (F:p = 7.6E-05,q = 0.01; M:p = 0.002,q = 0.16), *ECM-receptor interaction* (F:p = 0.001,q = 0.02; M:p = 0.007, q = 0.26), Hippo signaling (F:p = 0.0004,q = 0.02; M:p = 0.01,q = 0.27), and *PI3K-Akt signaling* (F:p = 0.0009,q = 0.02; M:p = 0.02,q = 0.38). Interestingly, the *cholesterol metabolism* pathway was over-represented in male (p = 0.001, q = 0.12) but not in female (p = 0.09, q = 0.28).

Protein interaction networks

Protein interaction networks were generated based on interactions from the BioGrid database (version 3.4.161) using proteins coded by genes going down with age as filtered with the criteria of a Pearson correlation < − 0.4 and a Bonferroni adjusted p < 0.05 (Fig. 6a). G protein subunit alpha L (GNAL; r = − 0.4, p = 4E-18 in male; r = − 0.46, p = 2E-09 in female; Supplementary Table 4A) is at the center of this network accounting for the involvement of G-protein and its receptors in most physiological responses to hormones, neurotransmitters. Several clusters are arranged around GNAL which are characterized by hub proteins BABAM1 (red), GNAS (yellow), TRIM25 (petrol), SPATA2 (green), APP (violet) and ELAVL1 (blue). Analogously to the downregulated

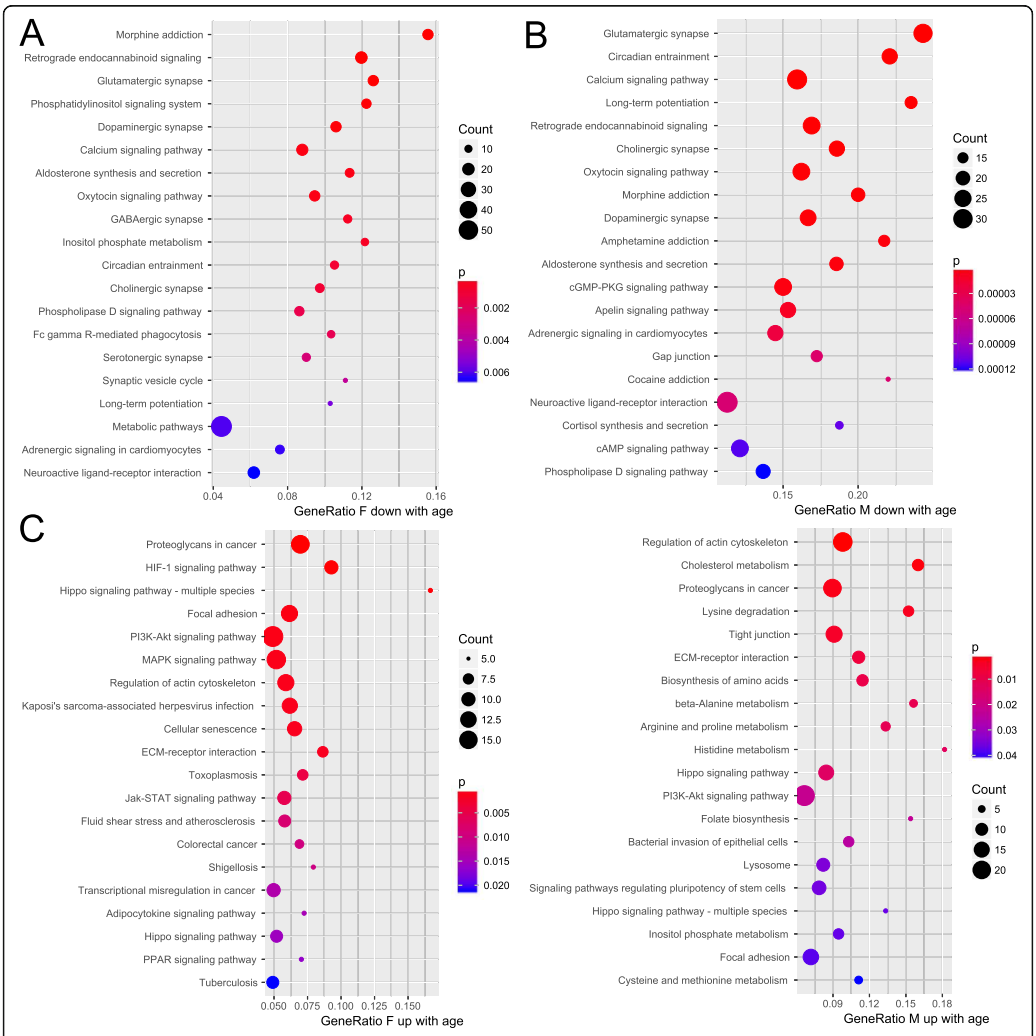


Fig. 5 Pathways down-regulated during aging are associated with various types of synapses, calcium signaling and long-term potentiation while up-regulated pathways are associated with the extracellular matrix, cytoskeleton, Hippo- and PI3K-Akt signaling. KEGG pathways of genes which were most significantly correlated and anti-correlated with age were analyzed separately for male and female prefrontal cortex. In both sexes pathways related to various types of synapses, calcium signaling and long-term potentiation were found overrepresented in the genes anti-correlated with age. In the genes correlated with age pathways associated with the extracellular matrix, cytoskeleton, Hippo- and PI3K-Akt -signaling are overrepresented

genes, the protein network of the upregulated genes was generated by filtering with the same *p*-value but with a Pearson correlation > 0.4 (Fig. 6b). The reactive astrocyte marker GFAP – coded by the gene with the highest correlation with age ($r = 0.55$ in male, $r = 0.62$ in female; Supplementary Table 4A) - has a central role in this network and is directly connected with APP.

Time series analysis of GFAP

Time series of *GFAP* gene expression with age were analyzed and compared with highly age-correlated and anti-correlated candidate genes with the aim of finding possible causal relationships. The gene *CAMK4* was found causative for the *GFAP* time series with the Granger causality test from the R package lmtest ($p = 0.015$). The

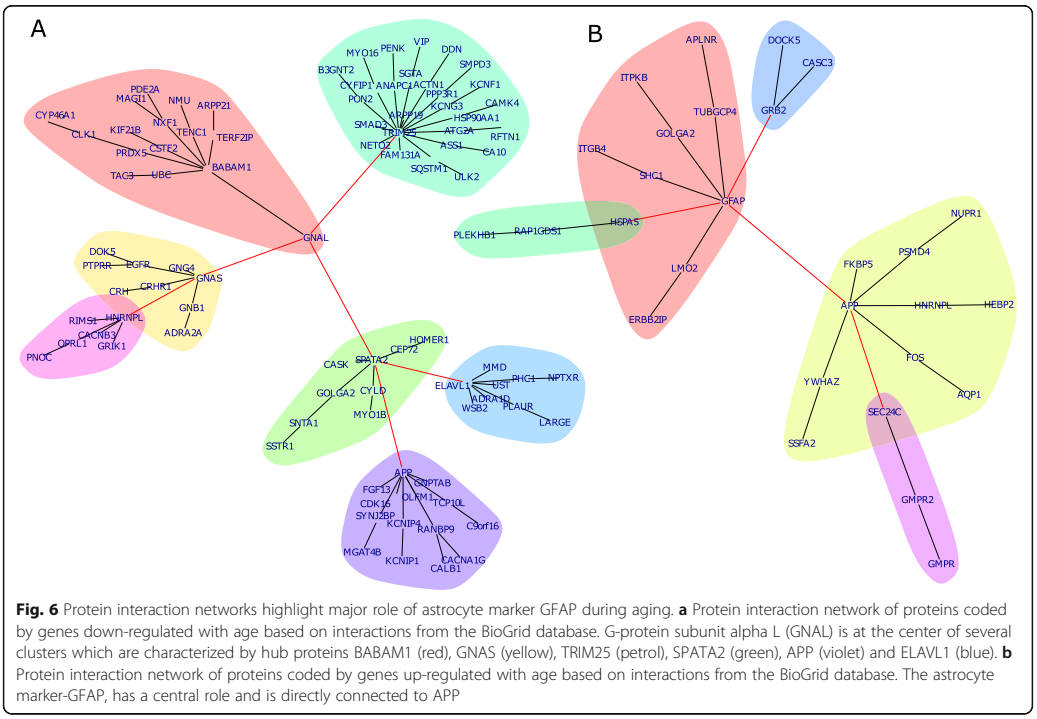


Fig. 6 Protein interaction networks highlight major role of astrocyte marker GFAP during aging. **a** Protein interaction network of proteins coded by genes down-regulated with age based on interactions from the BioGrid database. G-protein subunit alpha L (GNAL) is at the center of several clusters which are characterized by hub proteins BABAM1 (red), GNAS (yellow), TRIM25 (petrol), SPATA2 (green), APP (violet) and ELAVL1 (blue). **b** Protein interaction network of proteins coded by genes up-regulated with age based on interactions from the BioGrid database. The astrocyte marker-GFAP, has a central role and is directly connected to APP

test for causality in the opposite direction was not significant ($p = 0.52$) indicating that regulation by a third gene can be excluded. The time series of *GFAP* possessing the highest positive and *CAMK4* possessing negative correlation with age are plotted in Fig. 7a.

A simplified scheme (Fig. 7b) illustrates these findings together with results from the previous analyses: astrocytes (marker *GFAP*) react to neuronal injury and ROS thereby regulating inflammatory processes. They regulate the uptake and release of neurotransmitters responsible for synaptic transmission - as described by Sofroniew et al. [46]. Age-related decline of Calcium signaling decreases the levels of downstream *CAMK4* – as mentioned above Granger-causing - up-regulation of *GFAP*. CAMK4 has been reported as a direct activator of CREB via phosphorylation of the Ser-133 residue [5] or also indirectly via MAPK [54] . By analyzing the *GFAP* promoter region we identified binding sites for CREB - beside STAT and NF-κB (Supplementary Table 7) which are usually considered as regulators of GFAP expression [38]. Antagonistic regulation of CREB and GFAP has been reported [43]. The levels of hormones such as estrogen, which decline with age play a major role in regulating the density of dendritic spines and as a consequence, modulation of synaptic transmission.

Time series analysis of GO *synaptic transmission*

In order to elucidate which processes induce synaptic transmission, we set out to test Granger causality between significant GOs and the GO *synaptic transmission*. A consensus time series for the GO *synaptic transmission* was generated by taking the mean of all time series of genes significantly up-regulated with age in this GO (for details see Methods section). Among the over-represented GO terms we looked for causal relationships to this consensus time series of *synaptic transmission* via the Granger test. Tables 4 and 5 show the up- and down-regulated GOs found causative for *synaptic transmission* this way. Interestingly, on top of the up-regulated terms in Table 4, numerous terms related to nitric oxide appear as most significant. Nitric oxide plays important roles in the nervous system and in mitochondria and has been described to mediate mitochondrial fragmentation leading to age-related neurodegenerative diseases [31]. There was also evidence that nitric oxide elevates intracellular calcium levels and thus mediates reactive astrogliosis [47]. Furthermore, in Table 4, the term *negative regulation of monocyte chemotactic protein-1 (MCP1/CCL2) production* indicates an aging-related loss of CCL2. CCL2 has been reported to be protective against neurotoxic effects of excessive glutamate

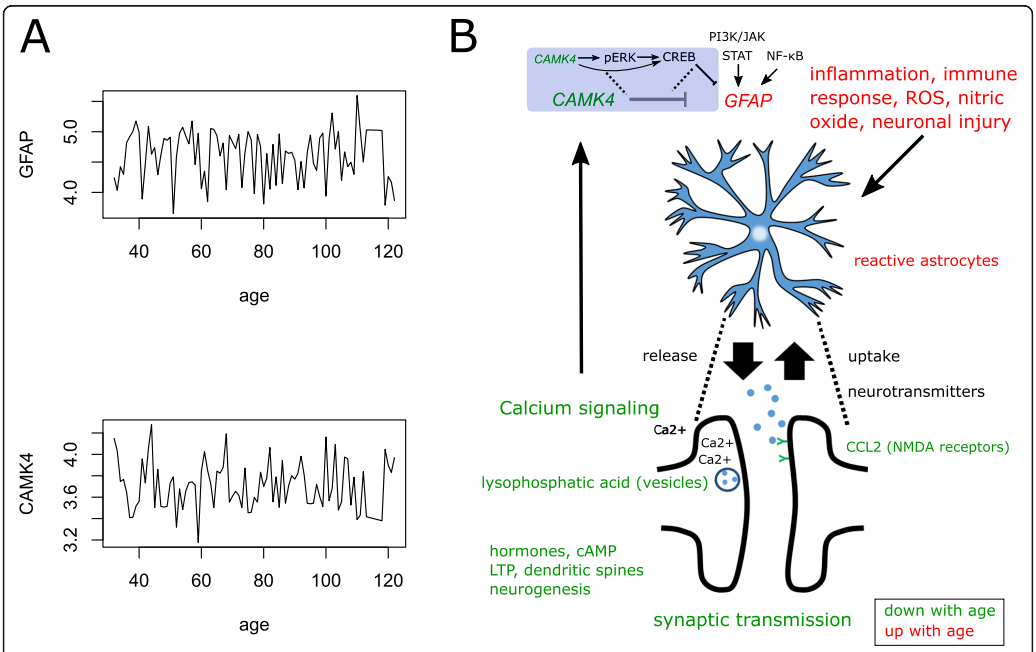


Fig. 7 Astrocyte marker GFAP has the highest correlation with prefrontal cortex aging and depends causally on CAMK4 in the time series. **a** The plots display time series of the genes GFAP possessing the highest positive and CAMK4 possessing negative correlation with age. The Wald test shows that the time series of CAMK4 is causative for GFAP time series. **b** A simplified scheme illustrates activation of astrocytes (marker GFAP) by inflammation, ROS and neuronal injury regulating uptake and release of neurotransmitters responsible for synaptic transmission. GFAP is regulated by CAMK4 – possibly via pERK and CREB (blue shading) - which is going down during aging and is downstream of Calcium signaling pathway. Down-regulation during aging is marked with green colour, up-regulation with red colour

Table 4 GOs going up with age "granger-causing" GO *synaptic transmission*

Term	ts2_c_ts1_p	ts1_c_ts2_p
Nitric oxide metabolic process	**0.0027**	0.1969
Regulation of nitric-oxide synthase biosynthetic process	**0.0034**	0.1853
Positive regulation of nitric oxide biosynthetic process	**0.0081**	0.3591
Positive regulation of myelination	**0.0107**	0.0538
Negative regulation of monocyte chemotactic protein-1 production	**0.0142**	0.5627
Regulation of cell-matrix adhesion	**0.0151**	**0.0445**
Schwann cell development	**0.0207**	0.0897
Histamine secretion	**0.0370**	0.0772
Azole transport	**0.0370**	0.0772
Positive regulation of reactive oxygen species metabolic process	**0.0372**	0.3284
Macrophage activation	**0.0386**	0.9125
Skin development	**0.0436**	0.4341
Renal absorption	**0.0451**	0.9061
Response to muscle stretch	**0.0497**	0.3531

ts2_c_ts1_p: *p*-value from Granger test between time series 2 (ts2,synaptic transmission) and ts1 (order of lags = 4)
ts1_c_ts2_p: *p*-value from Granger test between ts1 and ts2 (order of lags = 4)
Significant *p*-values < 0.05 are marked in bold

at NMDA receptors [15]. El Khoury et al. additionally described protective effects of CCL2 in Alzheimer-like disease by triggering the recruitment of astrocytes and microglia and subsequent removal of Amyloid-β [14].

In Table 5, the first term *microtubule nucleation* has a p-value below 0.05 in both directions indicating that a third factor may cause both. The term *lysophosphatic acid binding* has a low p-value of 0.0092 in the direction of "granger-causing" *synaptic transmission* and a relatively high p-value of 0.3682 in the opposite direction thus pointing to *lysophosphatic acid binding* as "granger-causing" *synaptic transmission*. Lysophosphatic acid has been reported to play a crucial role in the formation of vesicles at synapses [44]. The decline of this activity and its consequences in the exchange of neurotransmitters would be one coherent explanation for the decrease of synaptic transmission. Besides, many synapsis-related terms appear in Table 5 such as cAMP-, dendrite- and calcium-transport–related terms and also aging-related oxidative-stress-mediated apoptosis.

Discussion

In this meta-analysis of transcriptomes derived from 591 prefrontal cortex biopsies, we found a gene set with significantly increasing and another with significantly decreasing expression during aging. The most outstanding gene within these gene sets was the reactive astrocyte marker *GFAP* which showed significantly increasing expression levels in the brains of aging males and females. The biological process most significantly down-regulated with aging was synaptic transmission - as expected due to its close relation to the aging-related symptoms of reduced cognitive performance. On the other hand, there is a complex causal chain of aging-related changes eventually leading to reduced synaptic transmission. We tried to elucidate these mechanistically taking into account known aging hallmarks such as metabolic instability, increasing inflammation levels and changes in intercellular communication and could identify several functional groups. Directly related to the decline of synaptic transmission was the observation of multiple types of synapses negatively correlated with aging in the pathway analyses - for example, glutamatergic, cholinergic, dopaminergic, GABAergic and serotonergic synapses. We found expression of the reactive astrocyte marker GFAP increasing with age. Of course, this has to be confirmed experimentally but however beyond the scope of this study. Astrocytes play an important role at synapses by taking up and releasing excessive neurotransmitters and transferring lactate as energy substrate [46]. Furthermore, they influence pruning and remodeling of synapses [46]. In our previous meta-analysis of human hippocampus derived biopsies, we also observed that *GFAP* expression strongly correlated with Alzheimer's disease (AD) [53]. Thus, *GFAP* represents astroglia activation and gliosis not only in the AD-affected brain during neurodegeneration [27] but also in the disease-free aging brain.

We identified calcium signaling as decreasing with age in both sexes. Calcium has been implicated in brain aging in the Calcium dysregulation hypothesis of brain aging and AD [33]. Calcium has a 10,000 times higher concentration outside the cells and is shuffled inside through ligand-gated glutamate receptors, such as N-methyl-d-aspartate receptor (NMDAR) or various voltage-gated channels [33]. The expression levels of NMDARs decrease with age in our analysis (Fig. 3b, Supplementary Table 5A, B). We found that up-regulation of *GFAP* is connected to the decrease of *CAMK4* possibly involving gene-regulation by CREB. *CAMK4*, a member of the family of calcium/calmodulin-dependent kinases was also found oppositely regulated to *GFAP* in the neocortex of frontotemporal dementia-like mice with TDP-43 depletion [55]. Sticozzi et al. reported that nitric oxide can elevate intracellular calcium and via calcium together with the ERK/calmodulin signaling pathway can mediate reactive astrogliosis trigerred by cytokines in a specific time frame [47].

cAMP signaling decreases with age in our analysis (Table 2) and has been reported to be disrupted by aging while in the healthy brain it modulates the strength of the synapses [39]. cAMP also regulates $Ca2^{+}$ release from the endoplasmic reticulum via ryanodine receptors (RYR) to eventually release it to the cytosol [33, 42].

A further interesting functional group declining with age are hormones (Table 2). Hormones are known to decrease during aging and hormones such as estrogen have a major impact on synaptic plasticity and cognitive performance [39].

Interestingly, the KEGG pathway- insulin secretion decreases with age in both sexes (Supplementary Table 6). It has not been fully clarified if there is insulin production in the brain but there is some evidence for it and at least it has been reported for several species [23]. An explanation for our observation is more likely the considerable overlap between down-regulated genes within the pathways of Insulin secretion and cAMP signaling which definitely plays a role in brain aging but also in pancreatic islets [18].. Frölich et al. found that insulin concentration and insulin receptor densities in the brain decrease with aging [17]. The role of insulin in aging has been assessed by a body of literature stating one major finding that insulin sensitivity is associated with longevity while insulin resistance is associated with higher mortality [1]. Evidence for the involvement of insulin in brain aging is provided by the correlation between type 2 diabetes and neurodegenerative dementias [3] and it culminates in the annotation of Alzheimer's disease as

Table 5 GOs going down with age "granger-causing" GO *synaptic transmission*

Term	ts2_c_ts1_p	ts1_c_ts2_p
Microtubule nucleation	**0.0041**	**0.0296**
Nuclear lamina	**0.0057**	0.0501
Physiological muscle hypertrophy	**0.0080**	0.1201
Cell growth involved in cardiac muscle cell development	**0.0080**	0.1201
Lysophosphatidic acid binding	**0.0092**	0.3682
Positive regulation of dendrite morphogenesis	**0.0094**	0.3850
Cyclic purine nucleotide metabolic process	**0.0098**	**0.0463**
Regulation of synaptic transmission, glutamatergic	**0.0114**	**0.0279**
Dermatan sulfate biosynthetic process	**0.0124**	0.8968
Positive regulation of cAMP metabolic process	**0.0125**	0.2016
Positive regulation of cyclic nucleotide biosynthetic process	**0.0125**	0.2016
Uropod	**0.0130**	0.7732
1-phosphatidylinositol-4-phosphate 5-kinase activity	**0.0130**	0.7732
Proton-transporting V-type ATPase, V0 domain	**0.0153**	0.0617
Regulation of cAMP biosynthetic process	**0.0160**	0.0609
Regulation of cyclic nucleotide metabolic process	**0.0160**	0.0609
Synaptic vesicle docking	**0.0180**	0.2162
Cell-matrix adhesion	**0.0183**	0.0998
rRNA 3′-end processing	**0.0186**	0.0670
Asymmetric stem cell division	**0.0192**	0.1360
Rac GTPase binding	**0.0197**	0.5859
Macromolecular complex assembly	**0.0203**	**0.0231**
Golgi cis cisterna	**0.0213**	0.1023
Endomembrane system	**0.0217**	**0.0247**
Intrinsic apoptotic signaling pathway in response to oxidative stress	**0.0222**	0.1157
Positive regulation of purine nucleotide biosynthetic process	**0.0241**	0.1792
Positive regulation of nucleotide metabolic process	**0.0249**	0.1746
Muscle tissue development	**0.0249**	0.0884
Transporter activity	**0.0250**	**0.0323**
Spindle microtubule	**0.0256**	0.0620
Striated muscle cell development	**0.0264**	0.0765
Neuromuscular junction development	**0.0269**	0.3610
Regulation of nucleotide biosynthetic process	**0.0275**	0.0665
Endoplasmic reticulum	**0.0356**	0.0755
Calcium:cation antiporter activity	**0.0385**	0.3791
Ligand-gated channel activity	**0.0400**	0.0526
Lipid modification	**0.0412**	0.3182
Phosphatidylinositol phosphorylation	**0.0429**	0.6188
Proteoglycan biosynthetic process	**0.0433**	0.7653
Regulation of purine nucleotide metabolic process	**0.0442**	0.0842
Positive regulation of nucleocytoplasmic transport	**0.0445**	0.1855
Chloride channel inhibitor activity	**0.0464**	0.2701
Regulation of synaptic vesicle transport	**0.0469**	0.0834
Glutamate secretion	**0.0480**	**0.0499**
Dendrite terminus	**0.0481**	0.7091

ts2_c_ts1_p: *p*-value from Granger test between time series 2 (ts2,synaptic transmission) and ts1 (order of lags = 4)
ts1_c_ts2_p: *p*-value from Granger test between ts1 and ts2 (order of lags = 4)
Significant *p*-values < 0.05 are marked in bold

"diabetes type 3" [13]. Anti-ageing effects have been attributed to cAMP signaling which is part of a negative feedback loop with insulin as it regulates insulin secretion in the pancreatic islets but on the other hand is itself regulated by insulin [49]. Our findings of down-regulated cAMP emphasize its role in aging because it plays a dual role in regulating insulin secretion and synapse strength.

Furthermore, the levels of reactive oxygen species (ROS) increase with age in both sexes as indicated by the significantly over-represented GOs *Regulation of ROS biosynthetic and metabolic processes* and *Response to oxidative stress* (Table 3). A large body of literature has described oxidative stress as a major player in the aging process, furthermore, Sofroniew et al. have associated increased levels of ROS as a trigger of astrogliosis [46].

We also identified down-regulation of neurogenesis with age in both sexes (Table 2). However, neurogenesis in human brain is only reported for hippocampus but not for cortex [37] and thus this finding may be rather due to similar gene expression patterns with the hippocampus or cell migration from the hippocampus. For the hippocampus, age-related decline in neurogenesis has been reported [37] what may partially contribute to diminished cognitive abilities.

Finally, we found increased inflammation and immune response predominantly in females (Table 3). These are well known aging-associated factors [7, 35] and related to reactive astrogliosis indicated by increased expression of *GFAP* [46]. Inflammation and immunity seem to be the only major functional group with sex differences. However, also in males, inflammation and immune responses are activated, thus confirming the results reported by Brink et al. [8].

This study may be limited by potential technical inaccuracies including differences between platforms that may not fully be equalized by cross-platform-normalization and gene expression changes in the post-mortem interval. Furthermore, causality tested by the Granger test refers to the ability of prior values of one time series to predict values of another time series that may not be necessarily causative. For the explanatory power of the time series one has to take into account the construction from multiple individuals.

In this sex -specific meta-analysis of PFC biopsy-derived transcriptomes, we uncovered gene sets positively and negatively correlated with age which eventually could be condensed to similar functionality in both sexes. Synaptic transmission was found to be most significantly down-regulated with age while the expression of the reactive astrocyte marker *GFAP* was the most significantly up–regulated gene with age. However, many more players are involved in the complex mechanisms of brain aging. We identified age-associated downregulated expression of *CAMK4* - potentially contributing to up-regulation of *GFAP* - and Calcium signaling, hormones, insulin secretion, cAMP, long-term potentiation, neurogenesis and dendritic spines declining with age. On the other hand, inflammation, oxidative stress and neuronal injury increased with age. In summary, we found that during aging synaptic transmission declines due to a complex interplay of increasing factors such as inflammation, oxidative stress, nitric oxide and decreasing factors such as calcium signaling, cAMP, dendritic spines, long-term potentiation, hormones and CCL2. These findings are summarized in the scheme presented in Fig. 7b.

The dataset provided here should be useful for experimentalist to test and derive novel hypothesis on brain aging using iPSC-based tools.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s40478-020-00907-8.

Additional file 1: Supplementary Figure 1: Characteristics of PCA. (A) Correlation plot of variances of genes influencing PC1 the most (B) Correlation plot of variances of genes influencing PC2 the most. (C) Scree plot of variances against the most important principal components (D) Variances of genes influencing PC1 the most. (E) Variances of genes influencing PC2 the most.

Additional file 2: Supplementary Table 1. Prefrontal cortex transcriptome datasets and their GEO accession numbers employed for the meta-analysis.

Additional file 3: Supplementary Table 2. Results of statistical tests comparing female versus male PFC biopsy-derived gene expression data. *P*-values and q-values are based on the R packages limma and qvalue, ratios are calculated by dividing mean female by mean male expression.

Additional file 4: Supplementary Table 3. Subsets of the venn diagrams comparing male and female differentially up- and down-regulated genes in young, middle-aged and old (sheets as in Fig. 2). (A) Genes down-regulated in F30_65 vs. F30 were compared with genes down-regulated in M30_65 vs. M30. (B) Genes up-regulated in F30_65 vs. F30 were compared with genes up-regulated in M30_65 vs. M30. (C) Genes down-regulated in F65 vs. F30_65 were compared with genes down-regulated in M65 vs. M30_65. (D) Genes up-regulated in F65 vs. F30_65 were compared with genes up-regulated in M65 vs. M30_65. (E) Genes down-regulated in F65 vs. F30 were compared with genes down-regulated in M65 vs. M30. (F) Genes up-regulated in F65 vs. F30 were compared with genes up-regulated in M65 vs. M30.

Additional file 5: Supplementary Table 4. Correlation of gene expression with age. (A) Pearson correlation coefficients of gene expression with age and corresponding *p*-value, q-values and Bonferroni corrected *p*-values. (B) 48 genes with highest differences in age correlation between female and male. Genes with higher age correlation in male than in female are marked red in column *cor_M-F*, with lower age correlation in green.

Additional file 6: Supplementary Table 5. Over-represented GO terms in genes anti-correlated and correlated with age. (A) anti-correlated in female, (B) anti-correlated in male, (C) correlated in female, (D) correlated in male. Genes for GO analysis were filtered with Bonferroni-corrected $p < 0.05$ and $r < -0.1$ for anti-correlated genes or $r > 0.1$ for correlated genes.

Additional file 7: Supplementary Table 6. Over-represented KEGG pathways in genes anti-correlated and correlated with age. (A) anti-

correlated in female, (B) anti-correlated in male, (C) correlated in female, (D) correlated in male. Genes for GO analysis were filtered with Bonferroni-corrected $p < 0.05$ and $r < -0.1$ for anti-correlated genes or $r > 0.1$ for correlated genes.

Additional file 8: Supplementary Table 7. Trancription factors in the GFAP 2 k base upstream region found with a p-Match search of the public Transfac database filtering with core-d-score < 0.9 and matrix-d-score < 0.9

Acknowledgements
James Adjaye acknowledges financial support from the Medical Faculty of Heinrich Heine University, Düsseldorf.

Authors' contributions
WW performed bioinformatic analyses. WW and JA wrote the manuscript. JA conceived the concept and coordinated the work. The authors read and approved the final manuscript.

Funding
James Adjaye acknowledges support from the Medical faculty of Heinrich Heine University, Düsseldorf.

Availability of data and materials
In this meta-analysis publicly available datasets from NCBI GEO were used. Generated data is submitted as supplementary material.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 23 January 2020 Accepted: 1 March 2020
Published online: 05 March 2020

References

1. Akintola AA, van Heemst D (2015) Insulin, aging, and the brain: mechanisms and implications. Front Endocrinol 6:13. https://doi.org/10.3389/fendo.2015.00013
2. Altman J, Das GD (1965) Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. J Comp Neurol 124:319–335
3. Arnold SE, Arvanitakis Z, Macauley-Rambach SL, Koenig AM, Wang H-Y, Ahima RS, Craft S, Gandy S, Buettner C, Stoeckel LE, Holtzman DM, Nathan DM (2018) Brain insulin resistance in type 2 diabetes and Alzheimer disease: concepts and conundrums. Nat Rev Neurol 14:168–181. https://doi.org/10.1038/nrneurol.2017.185
4. Barnes MR, Huxley-Jones J, Maycox PR, Lennon M, Thornber A, Kelly F, Bates S, Taylor A, Reid J, Jones N, Schroeder J, Scorer CA, Davies C, Hagan JJ, Kew JNC, Angelinetta C, Akbar T, Hirsch S, Mortimer AM, Barnes TRE, de Belleroche J (2011) Transcription and pathway analysis of the superior temporal cortex and anterior prefrontal cortex in schizophrenia. J Neurosci Res 89:1218–1227. https://doi.org/10.1002/jnr.22647
5. Bito H, Deisseroth K, Tsien RW (1996) CREB phosphorylation and dephosphorylation: a Ca(2+)- and stimulus duration-dependent switch for hippocampal gene expression. Cell 87:1203–1214. https://doi.org/10.1016/s0092-8674(00)81816-4
6. Bourne J, Harris KM (2007) Do thin spines learn to be mushroom spines that remember? Curr Opin Neurobiol 17:381–386. https://doi.org/10.1016/j.conb.2007.04.009
7. Brink TC, Demetrius L, Lehrach H, Adjaye J (2009) Age-related transcriptional changes in gene expression in different organs of mice support the metabolic stability theory of aging. Biogerontology 10:549–564. https://doi.org/10.1007/s10522-008-9197-8
8. Brink TC, Regenbrecht C, Demetrius L, Lehrach H, Adjaye J (2009) Activation of the immune response is a key feature of aging in mice. Biogerontology 10:721–734. https://doi.org/10.1007/s10522-009-9219-1
9. Chatr-Aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, O'Donnell L, Oster S, Theesfeld C, Sellam A, Stark C, Breitkreutz B-J, Dolinski K, Tyers M (2017) The BioGRID interaction database: 2017 update. Nucleic Acids Res 45:D369–D379. https://doi.org/10.1093/nar/gkw1102
10. Chen C-Y, Logan RW, Ma T, Lewis DA, Tseng GC, Sibille E, McClung CA (2016) Effects of aging on circadian patterns of gene expression in the human prefrontal cortex. Proc Natl Acad Sci U S A 113:206–211. https://doi.org/10.1073/pnas.1508249112
11. Cheng H, Xuan H, Green CD, Han Y, Sun N, Shen H, McDermott J, Bennett DA, Lan F, Han J-DJ (2018) Repression of human and mouse brain inflammaging transcriptome by broad gene-body histone hyperacetylation. Proc Natl Acad Sci U S A 115:7611–7616. https://doi.org/10.1073/pnas.1800656115
12. Csardi G, Nepusz T (2006) The igraph software package for complex network research. InterJ Complex Syst 1695:1–9
13. de la Monte SM, Wands JR (2008) Alzheimer's disease is type 3 diabetes--evidence reviewed. J Diabetes Sci Technol 2:1101–1113. https://doi.org/10.1177/193229680800200619
14. El Khoury J, Toft M, Hickman SE, Means TK, Terada K, Geula C, Luster AD (2007) Ccr2 deficiency impairs microglial accumulation and accelerates progression of Alzheimer-like disease. Nat Med 13:432–438. https://doi.org/10.1038/nm1555
15. Eugenin EA, D'Aversa TG, Lopez L, Calderon TM, Berman JW (2003) MCP-1 (CCL2) protects human neurons and astrocytes from NMDA or HIV-tat-induced apoptosis. J Neurochem 85:1299–1311. https://doi.org/10.1046/j.1471-4159.2003.01775.x
16. Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. Bioinformatics 23:257–258. https://doi.org/10.1093/bioinformatics/btl567
17. Frölich L, Blum-Degen D, Bernstein HG, Engelsberger S, Humrich J, Laufer S, Muschner D, Thalheimer A, Türk A, Hoyer S, Zöchling R, Boissl KW, Jellinger K, Riederer P (1998) Brain insulin and insulin receptors in aging and sporadic Alzheimer's disease. J Neural Transm (Vienna) 105:423–438. https://doi.org/10.1007/s007020050068
18. Furman B, Ong WK, Pyne NJ (2010) Cyclic AMP signaling in pancreatic islets. Adv Exp Med Biol 654:281–304. https://doi.org/10.1007/978-90-481-3271-3_13
19. Galili T (2015) Dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics 31:3718–3720. https://doi.org/10.1093/bioinformatics/btv428
20. Garelick T, Swann J (2014) Testosterone regulates the density of dendritic spines in the male preoptic area. Horm Behav 65:249–253. https://doi.org/10.1016/j.yhbeh.2014.01.008
21. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80. https://doi.org/10.1186/gb-2004-5-10-r80
22. Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37:424–438. https://doi.org/10.2307/1912791
23. Gray SM, Meijer RI, Barrett EJ (2014) Insulin regulates brain function, but how does it get there? Diabetes 63:3992–3997. https://doi.org/10.2337/db14-0340
24. Hagenauer MH, Schulmann A, Li JZ, Vawter MP, Walsh DM, Thompson RC, Turner CA, Bunney WE, Myers RM, Barchas JD, Schatzberg AF, Watson SJ, Akil H (2018) Inference of cell type content from human brain transcriptomic datasets illuminates the effects of age, manner of death, dissection, and psychiatric diagnosis. PLoS One 13:e0200003. https://doi.org/10.1371/journal.pone.0200003
25. Harris KM, Kater SB (1994) Dendritic spines: cellular specializations imparting both stability and flexibility to synaptic function. Annu Rev Neurosci 17:341–371. https://doi.org/10.1146/annurev.ne.17.030194.002013
26. Hekimi S, Lapointe J, Wen Y (2011) Taking a "good" look at free radicals in the aging process. Trends Cell Biol 21:569–576. https://doi.org/10.1016/j.tcb.2011.06.008
27. Heppner FL, Ransohoff RM, Becher B (2015) Immune attack: the role of inflammation in Alzheimer disease. Nat Rev Neurosci 16:358–372. https://doi.org/10.1038/nrn3880
28. Hyndman RJ, Khandakar Y (2008) Automatic time series forecasting: the forecast package for R. J Stat Softw 26:1–22

29. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353–D361. https://doi.org/10.1093/nar/gkw1092
30. Kassambara A (2017) Practical guide to cluster analysis in R: unsupervised machine learning, edition 1. STHDA, Frankreich
31. Knott AB, Bossy-Wetzel E (2010) Impact of nitric oxide on metabolism in health and age-related disease. Diabetes Obes Metab 12(Suppl 2):126–133. https://doi.org/10.1111/j.1463-1326.2010.01267.x
32. Kohman RA, Rhodes JS (2013) Neurogenesis, inflammation and behavior. Brain Behav Immun 27:22–32. https://doi.org/10.1016/j.bbi.2012.09.003
33. Kumar A, Bodhinathan K, Foster TC (2009) Susceptibility to calcium dysregulation during brain aging. Front Aging Neurosci 1:2. https://doi.org/10.3389/neuro.24.002.2009
34. Lanz TA, Joshi JJ, Reinhart V, Johnson K, Grantham LE, Volfson D (2015) STEP levels are unchanged in pre-frontal cortex and associative striatum in post-mortem human brain samples from subjects with schizophrenia, bipolar disorder and major depressive disorder. PLoS One 10:e0121744. https://doi.org/10.1371/journal.pone.0121744
35. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G (2013) The hallmarks of aging. Cell 153:1194–1217. https://doi.org/10.1016/j.cell.2013.05.039
36. Lu T, Aron L, Zullo J, Pan Y, Kim H, Chen Y, Yang T-H, Kim H-M, Drake D, Liu XS, Bennett DA, Colaiácovo MP, Yankner BA (2014) REST and stress resistance in ageing and Alzheimer's disease. Nature 507:448–454. https://doi.org/10.1038/nature13163
37. Manganas LN, Zhang X, Li Y, Hazel RD, Smith SD, Wagshul ME, Henn F, Benveniste H, Djuric PM, Enikolopov G, Maletic-Savatic M (2007) Magnetic resonance spectroscopy identifies neural progenitor cells in the live human brain. Science 318:980–985. https://doi.org/10.1126/science.1147851
38. Middeldorp J, Hol EM (2011) GFAP in health and disease. Prog Neurobiol 93:421–443. https://doi.org/10.1016/j.pneurobio.2011.01.005
39. Morrison JH, Baxter MG (2012) The ageing cortical synapse: hallmarks and implications for cognitive decline. Nat Rev Neurosci 13:240–250. https://doi.org/10.1038/nrn3200
40. Mostany R, Anstey JE, Crump KL, Maco B, Knott G, Portera-Cailliau C (2013) Altered synaptic dynamics during normal brain aging. J Neurosci 33:4094–4104. https://doi.org/10.1523/JNEUROSCI.4825-12.2013
41. Narayan S, Tang B, Head SR, Gilmartin TJ, Sutcliffe JG, Dean B, Thomas EA (2008) Molecular profiles of schizophrenia in the CNS at different stages of illness. Brain Res 1239:235–248. https://doi.org/10.1016/j.brainres.2008.08.023
42. Park S-J, Ahmad F, Philp A, Baar K, Williams T, Luo H, Ke H, Rehmann H, Taussig R, Brown AL, Kim MK, Beaven MA, Burgin AB, Manganiello V, Chung JH (2012) Resveratrol ameliorates aging-related metabolic phenotypes by inhibiting cAMP phosphodiesterases. Cell 148:421–433. https://doi.org/10.1016/j.cell.2012.01.017
43. Pugazhenthi S, Wang M, Pham S, Sze C-I, Eckman CB (2011) Downregulation of CREB expression in Alzheimer's brain and in Aβ-treated rat hippocampal neurons. Mol Neurodegener 6:60. https://doi.org/10.1186/1750-1326-6-60
44. Schmidt A, Wolde M, Thiele C, Fest W, Kratzin H, Podtelejnikov AV, Witke W, Huttner WB, Söling HD (1999) Endophilin I mediates synaptic vesicle formation by transfer of arachidonate to lysophosphatidic acid. Nature 401:133–141. https://doi.org/10.1038/43613
45. Smith CC, Vedder LC, McMahon LL (2009) Estradiol and the relationship between dendritic spines, NR2B containing NMDA receptors, and the magnitude of long-term potentiation at hippocampal CA3-CA1 synapses. Psychoneuroendocrinology 34(Suppl 1):S130–S142. https://doi.org/10.1016/j.psyneuen.2009.06.003
46. Sofroniew MV, Vinters HV (2010) Astrocytes: biology and pathology. Acta Neuropathol 119:7–35. https://doi.org/10.1007/s00401-009-0619-8
47. Sticozzi C, Belmonte G, Meini A, Carbotti P, Grasso G, Palmi M (2013) IL-1β induces GFAP expression in vitro and in vivo and protects neurons from traumatic injury-associated apoptosis in rat brain striatum via NFκB/Ca^{2+}-calmodulin/ERK mitogen-activated protein kinase signaling pathway. Neuroscience 252:367–383. https://doi.org/10.1016/j.neuroscience.2013.07.061
48. Taminau J, Meganck S, Lazar C, Steenhoff D, Coletta A, Molter C, Duque R, de Schaetzen V, Weiss Solís DY, Bersini H, Nowé A (2012) Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/bioconductor packages. BMC Bioinformatics 13:335. https://doi.org/10.1186/1471-2105-13-335
49. Wang Z, Zhang L, Liang Y, Zhang C, Xu Z, Zhang L, Fuji R, Mu W, Li L, Jiang J, Ju Y, Wang Z (2015) Cyclic AMP mimics the anti-ageing effects of calorie restriction by up-regulating Sirtuin. Sci Rep 5:1–10. https://doi.org/10.1038/srep12012
50. Wei T, Simko V (2017) R package "corrplot": visualization of a correlation matrix
51. Wickham H (2009) Ggplot2: elegant graphics for data analysis. Springer, New York
52. Woolley CS, McEwen BS (1994) Estradiol regulates hippocampal dendritic spine density via an N-methyl-D-aspartate receptor-dependent mechanism. J Neurosci 14:7680–7687
53. Wruck W, Schröter F, Adjaye J (2016) Meta-analysis of transcriptome data related to hippocampus biopsies and iPSC-derived neuronal cells from alzheimer's disease patients reveals an association with FOXA1 and FOXA2 gene regulatory networks. J Alzheimers Dis 50:1065–1082. https://doi.org/10.3233/JAD-150733
54. Wu GY, Deisseroth K, Tsien RW (2001) Activity-dependent CREB phosphorylation: convergence of a fast, sensitive calmodulin kinase pathway and a slow, less sensitive mitogen-activated protein kinase pathway. Proc Natl Acad Sci U S A 98:2808–2813. https://doi.org/10.1073/pnas.051634198
55. Wu L-S, Cheng W-C, Chen C-Y, Wu M-C, Wang Y-C, Tseng Y-H, Chuang T-J, Shen C-KJ (2019) Transcriptomopathies of pre- and post-symptomatic frontotemporal dementia-like mice with TDP-43 depletion in forebrain neurons. Acta Neuropathol Commun 7:50. https://doi.org/10.1186/s40478-019-0674-x
56. Zeileis A, Hothorn T (2002) Diagnostic checking in regression relationships. R News 2:7–10

Publisher's Note

2.1.9 SARS-CoV-2 receptor ACE2 is co-expressed with genes related to transmembrane serine proteases, viral entry, immunity and cellular stress.

Sci Rep. 2020 Dec 8;10(1):21415. doi: 10.1038/s41598-020-78402-2.

Wruck W(1), Adjaye J(2).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Moorenstr.5, 40225, Düsseldorf, Germany. (2)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Moorenstr.5, 40225, Düsseldorf, Germany. james.adjaye@med.uni-duesseldorf.de.

The COVID-19 pandemic resulting from the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which emerged in December 2019 in Wuhan in China has placed immense burden on national economies and global health. At present neither vaccination nor therapies are available. Here, we performed a meta-analysis of RNA-sequencing data from three studies employing human lung epithelial cells. Of these one focused on lung epithelial cells infected with SARS-CoV-2. We aimed at identifying genes co-expressed with angiotensin I converting enzyme 2 (ACE2) the human cell entry receptor of SARS-CoV-2, and unveiled several genes correlated or inversely correlated with high significance, among the most significant of these was the transmembrane serine protease 4 (TMPRSS4). Serine proteases are known to be involved in the infection process by priming the virus spike protein. Pathway analysis revealed virus infection amongst the most significantly correlated pathways. Gene Ontologies revealed regulation of viral life cycle, immune responses, pro-inflammatory responses- several interleukins such as IL6, IL1, IL20 and IL33, IFI16 regulating the interferon response to a virus, chemo-attraction of macrophages, and cellular stress resulting from activated Reactive Oxygen Species. We believe that this dataset will aid in a better understanding of the molecular mechanism(s) underlying COVID-19.

DOI: 10.1038/s41598-020-78402-2 PMCID: PMC7723043 PMID: 33293627

URL: https://www.nature.com/articles/s41598-020-78402-2

Contribution: Wasco Wruck analysed the data and wrote the manuscript.

www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

SARS-CoV-2 receptor ACE2 is co-expressed with genes related to transmembrane serine proteases, viral entry, immunity and cellular stress

Wasco Wruck & James Adjaye✉

The COVID-19 pandemic resulting from the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which emerged in December 2019 in Wuhan in China has placed immense burden on national economies and global health. At present neither vaccination nor therapies are available. Here, we performed a meta-analysis of RNA-sequencing data from three studies employing human lung epithelial cells. Of these one focused on lung epithelial cells infected with SARS-CoV-2. We aimed at identifying genes co-expressed with angiotensin I converting enzyme 2 (ACE2) the human cell entry receptor of SARS-CoV-2, and unveiled several genes correlated or inversely correlated with high significance, among the most significant of these was the transmembrane serine protease 4 (TMPRSS4). Serine proteases are known to be involved in the infection process by priming the virus spike protein. Pathway analysis revealed virus infection amongst the most significantly correlated pathways. Gene Ontologies revealed regulation of viral life cycle, immune responses, pro-inflammatory responses- several interleukins such as IL6, IL1, IL20 and IL33, IFI16 regulating the interferon response to a virus, chemo-attraction of macrophages, and cellular stress resulting from activated Reactive Oxygen Species. We believe that this dataset will aid in a better understanding of the molecular mechanism(s) underlying COVID-19.

Severe acute respiratory disease COVID-19 is a result of infections with the coronavirus SARS-CoV-2 first reported in the Chinese city Wuhan, Province Hubei, in December 2019 and has since 11 March 2020 been designated as a pandemic by WHO. The origin of the virus is most likely zoonotic[1,2] but the exact species transferring it is still under investigation as some studies suggest that it was transferred from pangolins[3] or bats[4]. Tilocca et al. analysed the SARS-CoV-2 nucleocapsid and envelop proteins and found besides highest similarities with bat and pangolin also considerable similarities with dog, cat, cattle and other species[5,6]. Based on this, they suggest that earlier contact to similar viruses hosted by other species might be responsible for protection or—in case of multiple contacts—for antibody defense enhancement[7]. At the end of April 2020, the number of globally confirmed cases of COVID-19 exceeded 3 million and recorded deaths beyond 200,000 in the real-time statistics of the Johns Hopkins University[8]. Due to many unreported and asymptomatic cases, the infection fatality rate (IFR) is difficult to determine however Verity et al. estimate approximately 0.66% (0·39–1·33) in China[9]. The age-associated IFR is approximately 7.8% for those above 80 years[9]. Drugs have been re-purposed for stabilizing COVID-19, but these are not effective therapies[10], examples include hydroxy-chloroquine (Malaria)[11,12] and nelfinavir (HIV)[13]. However, remdesivir which was designed for Ebola treatment[11,12] was at least shown to shorten the time to recovery and to reduce infection in the lower respiratory tract[14]. Another treatment option is to indirectly immunize individuals with plasma from convalescent COVID-19 patients[15]. Further approaches aim at mimicking the human virus cell entry receptor ACE2[4] with human recombinant soluble ACE2 (hrsACE2)[16]. The cell entry receptor ACE2 associates with transmembrane proteases which prime the spike protein of the virus. Hoffmann et al. assigned this task to the protein TMPRSS2[17] which they identified in the predecessor virus SARS-CoV and showed it is the same for SARS-CoV-2. The protease can be inhibited by existing compounds to interrupt further propagation of the virus in the human host. Several other publications confirmed the role

Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Moorenstr.5, 40225 Düsseldorf, Germany. ✉email: james.adjaye@med.uni-duesseldorf.de

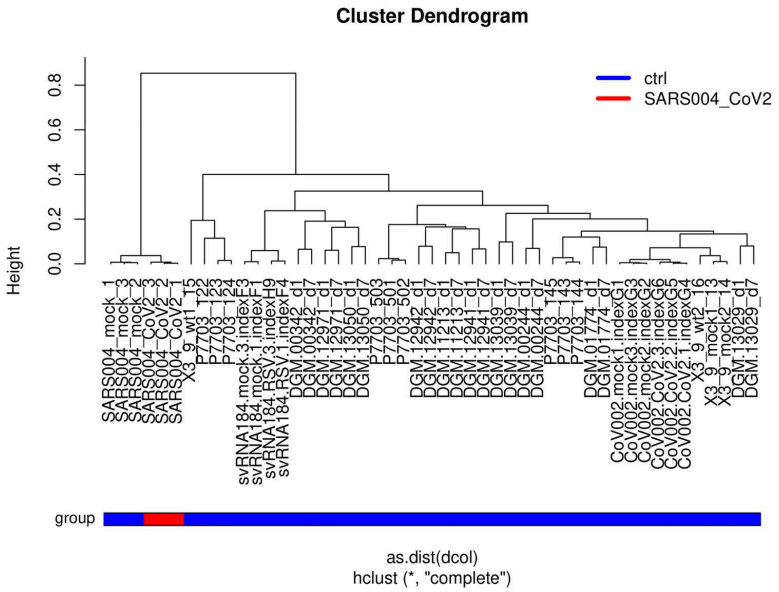


Figure 1. Lung epithelial cells (labeled "SARS004") infected with SARS-CoV-2 cluster together with mock-infected lung epithelial cells but separated from all other lung epithelial cells.

Dataset	Description	Reference
GSE147507	Primary human bronchial epithelial cells and lung adenocarcinoma infected with SARS-CoV-2 or Mock	Blanco-Melo, D. et al. Cell 181, 1036–1045.e9 (2020)
GSE146482	Human bronchial epithelium cell line BEAS-2B	Mukherjee SP et al. unpublished
GSE85121	Small airway epithelium brushing	Staudt MR et al. Respir Res 2018 May 14;19(1):78. PMID: 29,754,582

Table 1. Datasets used in this focussed meta-analysis.

of TMPRSS2 including a study by Matsuyama et al. showing enhanced SARS-CoV-2 infection in TMPRSS2 expressing VeroE6 cells[18] and a study by Sungnak et al. reporting co-expression of ACE2 and TMPRSS2 in multiple tissues[19]. Collin et al. found that SARS-CoV-2 may enter the human body through the ocular surface epithelium mediated by ACE2 and TMPRSS2[20].

Interestingly, ACE2 has been reported to be down-regulated following lung injury by Imai et al.[21] and in the previous virus SAR-CoV by Haga et al.[22] via a mechanism involving TNF-α-converting enzyme (TACE) shedding of the ectodomain of ACE2.

Here, we describe a meta-analysis focussing on the transcriptome data from human lung epithelial cells including samples infected with SARS-CoV-2 from a study described by Blanco Melo et al.[23]. We directed the exploration to co-expression with the known SARS-CoV-2 receptor ACE2. The analysis revealed a signature consisting of 72 genes significantly co-expressed with ACE2 either with positive or negative Pearson correlation. Of the transmembrane serine proteases, the most significantly co-expressed with ACE2 was TMPRSS4, suggesting it to be a putative druggable target.

Results

Cluster analysis of SARS-CoV-2 infected cells compared to other non-infected lung cells. Figure 1 shows a hierarchical cluster analysis of all samples used in this meta-analysis. Lung epithelial cells (labeled "SARS004" in Fig. 1) infected with SARS-CoV-2 cluster together with mock-infected lung epithelial cells but separated from all other lung epithelial cells and lung carcinoma cell lines which together we consider as control in this analysis. The table of Pearson correlation co-efficients (suppl. Table 2) reflects the grouping implicated by the hierarchical clustering: SARS-CoV-2 samples have highest correlation (r = 0.9884–0.9936) to the mock-infected SARS cells (Table 1).

Analysis of genes with correlated and anti-correlated expression with ACE2. Building on the knowledge about ACE2 as receptor of the SARS-CoV-2 virus we set out to indentify genes with highly correlated expression with ACE2 with the aim of elucidating the molecular mechanisms underlying COVID-19. Figure 2 shows the genes most significantly (Bonferoni-adjusted p < 1E−1) correlated (red to yellow in last column) or anti-correlated (green) with ACE2 (full table in suppl. Table 3). The transmembrane serine protease 4 (TMPRSS4) is one of the most significantly correlated (r = 0.9142, p = 4.59E−20) with ACE2 therefore implying a major role in priming the SARS-CoV-2 spike protein. CXCL17 (r = 0.9273, p = 1.1E−21), ABCA12 (r = 0.9256, p = 1-92E−21) and ATP10B (r = 0.9193, p = 1.14E−20) had marginally higher correlation with ACE2 while another transmembrane protease TMPRS11E (r = 0.9121, p = 7.91E-20) had a slightly lower correlation. The expression of CXCL17 is probably due to a reaction to the infection by chemo-attracting macrophages[24]. The role of the ATP binding cassette subfamily A member 12 (ABCA12) is not fully elucidated with respect to COVID-19 but assumed to transport lipid via lipid granules to the intracellular space and transporting specific proteases – in the case of harlequin ichtyosis associated with desquamation[25]. Current knowledge on ATP10B is scarce. However, Wilk et al. (Table 3 in their publication) report Atp10b gene expression levels as highly inversely (negative) correlated with influenza gene expression changes in infected C57BL/6 J mice[26]. In Fig. 3a a cluster analysis and gene expression heatmap of the 72 most significantly (Bonferoni-adjusted p < 1E−11) correlated and anti-correlated genes with ACE2 shows close clustering of the serine protease TMPRSS4 with ACE2. Also in this analysis of 72 genes, SARS-CoV-2 (red color bar) cluster together with mock-infected SARS lung epithelial cells and separated from the other lung cells (blue color bar indicates control). In the heatmap presented in Fig. 3b, TMPRSS family members TMPRSS11D/E and TMPRSS4 show close clustering with ACE2 but also TMPRSS2 and TMPRSS13 have similar expression in all experiments, especially in the SARS-CoV-2 infected samples.

Pathway analysis of genes co-regulated with ACE2. In order to investigate the functionality of genes interacting with *ACE2* we filtered genes correlated with ACE2 by Bonferoni-adjusted p value < 0.05 and Pearson correlation coefficient > 0.6. 1891 genes fulfilled these criteria and were subjected to over-representation analysis of KEGG pathways[27]. The most significantly over-represented pathways associated with the 1891 genes correlated with *ACE2* (Fig. 4a, suppl. Table 4) are for example, *Bacterial invasion of epithelial cells (q = 4.4E−06), Human papillomavirus infection (q = 0.0006), Transcriptional misregulation in cancer (q = 0.0006)* and *Endocytosis (q = 0.002)*. This reflects the mechanisms of virus infection via invasion of epithelial cells and endocytosis.

Pathway analysis of genes anti-correlated with ACE2. Analogously to the positively correlated genes we also examined the negatively correlated genes with ACE2 by filtering for Bonferoni-adjusted p value < 0.05 and Pearson correlation coefficient < − 0.6. 1993 genes passed these filtering criteria and were subjected to over-representation analysis of KEGG pathways[27]. The most significantly over-represented pathways in the 1993 genes negatively correlated with *ACE2* (Fig. 4b, suppl. Table 5) are *DNA replication (q = 1E−12), Metabolic pathways (q = 1.86E.08), Cell cycle (q = 1.1E−05), Fanconi anemia pathway (q = 1.24E−05), Mismatch repair (q = 9.89E−05)* and *Homologouos recombination (q = 0.0046)*. Many of these pathways are associated with DNA processing or repair. That these are over-represented in genes down-regulated upon infection with SARS-CoV-2 is in line with reports about interferon and interferon stimulated genes (ISGs) inhibiting virus replication[28]. This would be a defense against the attempts of the virus to recruit the host's DNA repair and homologous recombination mechanisms[29].

GO analysis of genes co-regulated with ACE2. We furthermore assessed the GOs over-represented in the 1891 genes positively and the 1993 genes negatively correlated with *ACE2*. Table 2 shows a selection of significant GOs from all three categories Biological Process (BP, Fig. 5a), Cellular Component (CC) and Molecular Function (MF), suppl. Table 6 provides the full table and suppl. Table 7 provides the full table for the 1993 genes negatively correlated with *ACE2*. Amongst the GO-BPs, metabolic processes are the most significant. GO-BP terms such as *Interspecies interaction between organisms, Cytokine production* and *positive regulation of viral process* reflect activated mechanisms post-viral infection. Interestingly, we found *regulation of coagulation* amongst the GO-BPs what may help elucidate reports about co-agulation in acro-ischemic COVID-19 patients[30]. In the GO-CCs, the terms *intracellular* and *membrane-bounded organelle* are most significant. In the GO-MFs, the terms *metal ion binding* and *protein binding* emerge as most significant due probably reflecting the binding of the virus proteins to the host cells. For the full gene lists associated with these terms refer to suppl. Tables 5 and 6.

Immune system associated GOs co-regulated with ACE2. Table 3 and Fig. 5b show GOs (all Biological Processes) related to the immune system over-represented in genes correlated with *ACE2*. Myeloid cells involved in the immune response (GO:0,002,275, p = 5E−07) as well as T-cells (GO:0,050,870, p = 0.02) are activated. Chemokines, in particular interleukin-1 are produced (GO:0,032,642, p = 0.0049, GO:0,032,652, p = 0.0049) also IL33 and TNF. Additionally, positive regulation of the innate immune response was prominent (GO:0,045,089, p = 0.0079). *CXCL17* was the gene with the highest correlation with *ACE2* and is involved in immune system process—as described above by chemo-attracting macrophages[24]. Among the most significantly *ACE2*-correlated genes was *IFI16* which is associated with several immune system GOs and known as regulator of the interferon response to viruses[31] and will be described in more detail in the next section about *Protein interaction networks*. Also in the protein-interaction network of the most significantly ACE2-correlated genes was the interleukin 20 receptor B (IL20RB) which appeared in several GOs listed in Table 3. With respect to viruses, there is meagre knowledge on IL20RB, however a study reported over-expression in the pneumonia causing avian influenza A H7N9 virus[32].

symbol	cor.p	cor.BH	cor.bonferoni	cor
ACE2	0.00E+00	0.00E+00	0.00E+00	1.0000
CXCL17	1.10E-21	7.53E-18	1.51E-17	0.9273
ABCA12	1.92E-21	8.77E-18	2.63E-17	0.9255
ATP10B	1.14E-20	3.92E-17	1.57E-16	0.9193
TMPRSS4	4.59E-20	1.26E-16	6.30E-16	0.9142
TMPRSS11E	7.91E-20	1.81E-16	1.09E-15	0.9121
CASP4	1.52E-19	2.98E-16	2.09E-15	0.9095
FUT3	3.20E-19	5.49E-16	4.39E-15	0.9065
GABRB3	4.99E-19	7.61E-16	6.85E-15	-0.9046
C14orf132	9.13E-19	1.20E-15	1.25E-14	-0.9020
TMPRSS11D	9.61E-19	1.20E-15	1.32E-14	0.9018
INHBB	1.27E-18	1.45E-15	1.75E-14	-0.9006
CDH13	1.89E-18	1.75E-15	2.60E-14	0.8988
MMP1	1.91E-18	1.75E-15	2.62E-14	0.8987
GNA15	1.91E-18	1.75E-15	2.63E-14	0.8987
DUOXA2	2.77E-18	2.38E-15	3.81E-14	0.8970
IL20RB	3.20E-18	2.59E-15	4.40E-14	0.8964
RPH3AL	3.70E-18	2.66E-15	5.08E-14	-0.8957
QSOX2	3.71E-18	2.66E-15	5.09E-14	-0.8957
LIMA1	3.87E-18	2.66E-15	5.32E-14	0.8955
ZNF750	4.47E-18	2.92E-15	6.14E-14	0.8948
NT5M	5.79E-18	3.62E-15	7.96E-14	-0.8936
HIP1	6.40E-18	3.82E-15	8.79E-14	-0.8931
TSPAN1	6.71E-18	3.84E-15	9.22E-14	0.8929
PIGR	8.73E-18	4.80E-15	1.20E-13	0.8916
CYP24A1	2.27E-17	1.15E-14	3.12E-13	-0.8868
ALDH3B1	2.30E-17	1.15E-14	3.16E-13	-0.8868
FOXA2	2.34E-17	1.15E-14	3.22E-13	-0.8867
HNF1A	2.78E-17	1.32E-14	3.82E-13	-0.8858
NMNAT1	3.09E-17	1.36E-14	4.25E-13	0.8852
DUOX2	3.14E-17	1.36E-14	4.31E-13	0.8852
PMP22	3.17E-17	1.36E-14	4.36E-13	-0.8851
TMEM121	3.58E-17	1.49E-14	4.92E-13	-0.8845
TUB	3.68E-17	1.49E-14	5.05E-13	-0.8843
BLNK	4.61E-17	1.81E-14	6.34E-13	0.8832
TRAF3IP2	6.20E-17	2.35E-14	8.52E-13	0.8816
PTHLH	6.32E-17	2.35E-14	8.69E-13	0.8815
WNT7B	8.74E-17	3.10E-14	1.20E-12	-0.8797
RNF39	8.80E-17	3.10E-14	1.21E-12	0.8797
ASB2	9.75E-17	3.35E-14	1.34E-12	0.8791
SPP1	1.02E-16	3.43E-14	1.41E-12	-0.8789
GPR68	1.19E-16	3.90E-14	1.64E-12	0.8780
PAQR5	1.23E-16	3.94E-14	1.69E-12	-0.8778
IGFBP4	1.33E-16	4.16E-14	1.83E-12	-0.8774
SAMD12	1.42E-16	4.32E-14	1.95E-12	0.8771
C2orf72	1.62E-16	4.64E-14	2.22E-12	-0.8763
PARP8	1.62E-16	4.64E-14	2.22E-12	0.8763
DDR1	1.63E-16	4.64E-14	2.24E-12	0.8763
NUP62CL	1.65E-16	4.64E-14	2.27E-12	-0.8762
PCDH7	1.87E-16	5.14E-14	2.57E-12	0.8755
C19orf33	1.91E-16	5.15E-14	2.63E-12	0.8754
TMEM229B	1.98E-16	5.24E-14	2.73E-12	0.8752
HNF4A	2.05E-16	5.32E-14	2.82E-12	-0.8750
MYRF	2.39E-16	6.07E-14	3.28E-12	-0.8741
SGCE	2.59E-16	6.46E-14	3.56E-12	-0.8737
IRAK3	2.63E-16	6.46E-14	3.62E-12	0.8736
CNNM3	2.80E-16	6.74E-14	3.84E-12	-0.8732
HK2	3.28E-16	7.76E-14	4.50E-12	0.8723
CRMP1	3.49E-16	8.06E-14	4.80E-12	-0.8719
IFI16	3.52E-16	8.06E-14	4.84E-12	0.8719
HLA-DOB	3.62E-16	8.15E-14	4.97E-12	0.8717
DAPP1	4.37E-16	9.57E-14	6.01E-12	0.8706
PCSK6	4.39E-16	9.57E-14	6.03E-12	-0.8706
CCDC102A	5.40E-16	1.15E-13	7.42E-12	-0.8694
S100A11	5.44E-16	1.15E-13	7.47E-12	0.8693
TRIM9	5.53E-16	1.15E-13	7.59E-12	-0.8692
SGK2	6.16E-16	1.26E-13	8.46E-12	-0.8686
EYA2	6.46E-16	1.30E-13	8.87E-12	0.8683
FANCC	6.54E-16	1.30E-13	8.98E-12	-0.8682
CHST10	6.71E-16	1.30E-13	9.22E-12	-0.8681
TMEM63C	6.76E-16	1.30E-13	9.28E-12	0.8680
ARNT2	6.83E-16	1.30E-13	9.39E-12	-0.8680
CSGALNACT1	7.14E-16	1.34E-13	9.81E-12	-0.8677

Figure 2. Most significantly (Bonferoni-adjusted p < 1E−11) correlated (red to yellow in last column) or anti-correlated (green) genes with ACE2. The transmembrane serine protease 4 (TMPRSS4) is one of the most significantly correlated (r = 0.9142, p = 4.59E−20) with ACE2 suggesting a major role in priming the SARS-CoV2 spike protein. CXCL17 (r = 0.9273, p = 1.1E−21), ABCA12 (r = 0.9256, p = 1−92E−21) and ATP10B (r = 0.9193, p = 1.14E−20) had marginally higher correlation with ACE2 while another transmembrane protease TMPRS11E (r = 0.9121, p = 7.91E−20) had slightly lower correlation.

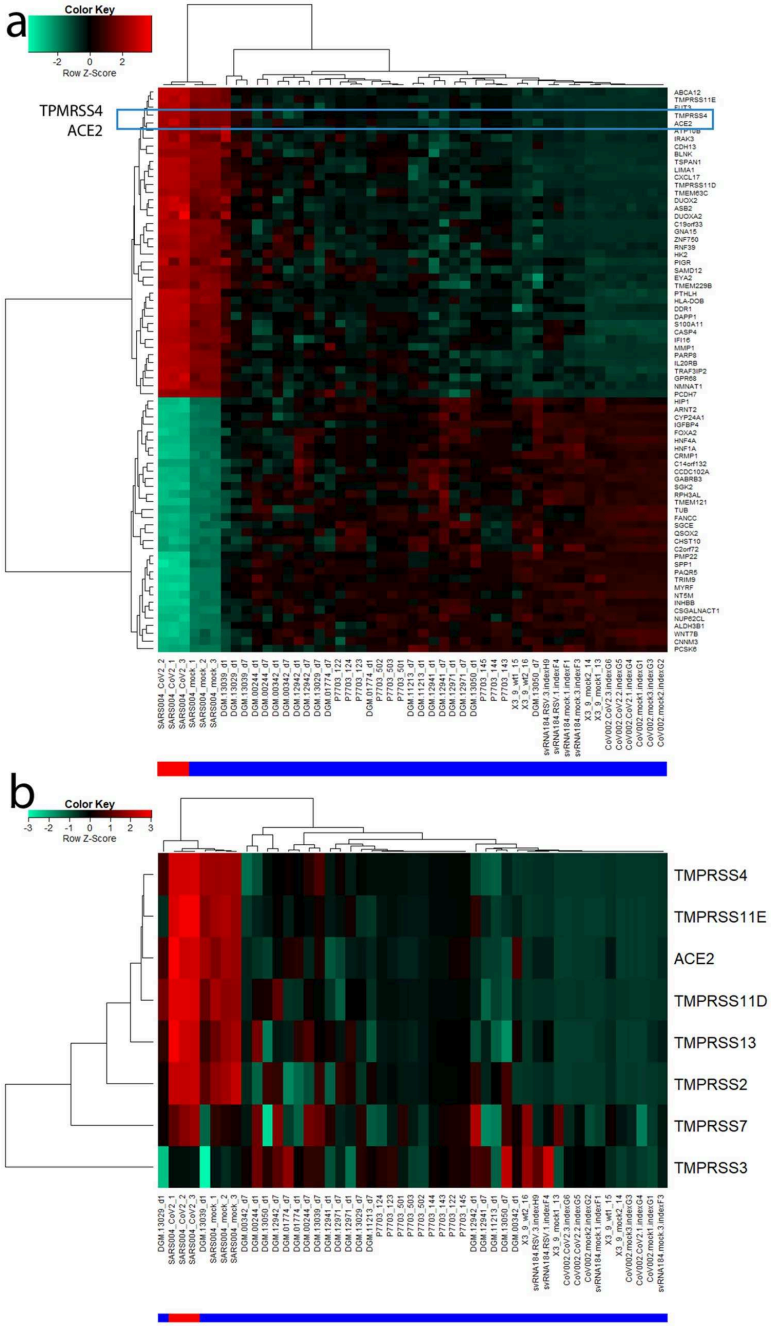


Figure 3. (**a**) Cluster analysis and gene expression heatmap of 72 most significantly (Bonferoni-adjusted p < 1E−11) correlated and anti.correlated genes with ACE2 shows close clustering of the serine protease TMPRSS4 with ACE2. (**b**) Heatmap of TMPRSS family members shows close clustering of TMPRSS11D/E and TMPRSS4 with ACE2 but also TMPRSS2 and TMPRSS13 have similar expression, particularly in SARS-CoV-2 infected samples. The red color bar indicates SARS-CoV-2, blue color bar control.

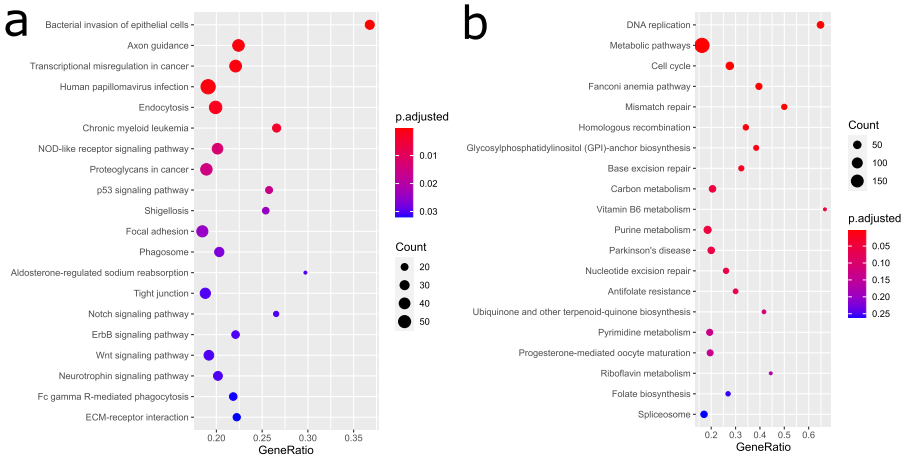


Figure 4. (**a**) The five most significantly overrepresented pathways correlated with ACE2 are *Human papillomavirus infection, Bacterial invasion of epithelial cells, Endocytosis, Axon Guidance* and *Transcriptional mis-regulation in cancer*. (**b**) The six most significantly overrepresented pathways anti-correlated with ACE2 are *DNA replication, Metabolic pathways, Cell cycle, Fanconi anemia pathway, Mismatch repair* and *Homologous recombination*. Many of these pathways are associated with DNA processing or repair. That these are down-regulated upon infection with SARS-CoV-2 is in line with reports about interferon and interferon stimulated genes (ISGs) inhibiting virus replication[16]. This would be a defense against the attempts of the virus to recruit the host's DNA repair and homologous recombination mechanisms as Gillespie et al. report[29].

GOs associated with inflammation and reactive oxygen species (ROS) amongst genes co-regulated with ACE2. Table 4 and Fig. 5c show GOs (all Biological Processes) related to inflammation and ROS over-represented in genes correlated with ACE2. The GO, *positive regulation of inflammatory response* ($p = 0.0039$) would imply that an inflammatory process is induced which finally leads to apoptosis—as demonstrated by the GO *inflammatory cell apoptotic process* ($p = 0.042$). The virus receptor ACE2 is involved in the *positive regulation of reactive oxygen species metabolic process* ($p = 0.0185$). Induction of ROS by a respiratory virus and subsequent inflammation has been reported Jamaluddin et al.[33].

Protein-interaction networks. We further restricted the set of ACE2-correlated or -anti-correlated genes by drastically filtering with Bonferoni-adjusted $p < 1E{-}11$ in order to construct a human readable protein interaction network of the most significant proteins (Fig. 6). The protein-interaction network generated from correlated and anti-correlated genes with *ACE2* shows *IFI16 ($r = 0.8719$), LIMA1 ($r = 0.8955$), CNNM3 ($r = -0.8732$), HNF4A ($r = -0.8750$), TRAF3IP2 ($r = 0.8816$), ASB2 ($r = 0.8791$)* and *FANCC* ($r = -0.8682$) as hub genes (interactors from the BioGrid database are marked in red, original *ACE2*-correlated/anti-correlated genes are marked in green, hub genes and *ACE2* are highlighted with yellow shading). Interferon plays a major role in the host response to a virus and Thompson et al. reported that IFI16 – one of the hub genes in our network—controls the interferon response to DNA and RNA viruses[31]. Lin and Richardson review that *LIMA1* (formerly EPLIN) – which is known to enhance bundling of actin filaments[34]—mediates the interaction between Cadherins and Actin in the context of adherens junctions—playing a role in measles virus infection—via trans-binding with molecular interactors on adjacent cells[35]. In line with this, LIMA1 is connected with E-cadherin (CDH1) in the network and also Cadherin 13 (CDH13) is part of it and among the most significantly correlated genes to ACE2. The connection of ACE2 to calmodulin 1 (CALM1) is based on a publication by Lambert et al.[36] in which they show that CALM1 interacts with the corona virus receptor ACE2 and inhibits shedding of its ectodomain[36]. CALM1 inhibitors in turn can reverse this process so that the ACE2 ectodomain is shed, and is partially mediated by a metalloproteinase[36]. The direct connection between CALM1 and LIMA1 was found in a large-scale interactome study[37]. The involvement of the hub gene Fanconi anemia complementation group C (FANCC), although not experimentally proven, might reflect recruitment of DNA repair and homologous recombination mechanisms from the host by the virus.

Role of TMPRSS4 and other TMPRSS gene family members. A pivotal result of this meta-analysis is the transmembrane serine protease TMPRSS4 which was one of the most significantly correlated genes with ACE2. Additionally TMPRSS11E ($r = 0.9121$, Bonferoni-corrected $p = 1.09E{-}15$) and TMPRS11D ($r = 0.9018$, Bonferoni-corrected $p = 1.32E{-}14$) from the same gene family were found more significant than TMPRSS2 which however was still significantly correlated with ACE2 ($r = 0.767$, Bonferoni-corrected $p = 1.79E{-}6$). The assignment of the SARS-CoV-2 spike protein priming functionality to TMPRSS2 was based on the assumption that it would be the same as for its predecessor SARS-CoV[17] and was confirmed by several publications[19,20].

GO_BP term	P value	GO_CC term	P value	GO_MF term	P value
Regulation of primary metabolic process	5.23E−11	Intracellular	3.35E−27	Metal ion binding	4.98E−08
Regulation of cellular metabolic process	6.04E−09	Membrane−bounded organelle	8.95E−23	Protein binding	1.84E−05
Organic substance biosynthetic process	8.56E−07	Extracellular exosome	9.79E−18	Catalytic activity, acting on a protein	6.44E−03
Positive regulation of cellular process	2.09E−06	Extracellular organelle	1.21E−17	Metallopeptidase activity	4.99E−02
Negative regulation of cellular metabolic process	2.69E−05	Cytoplasm	2.07E−14		
Gene expression	3.24E−05	Vesicle	8.85E−14		
Interspecies interaction between organisms	3.30E−05	Ion binding	5.22E−08		
Cell junction assembly	1.20E−04	Membrane raft	2.03E−04		
Cytokine production	2.21E−04	Membrane region	5.95E−04		
Regulation of nitrogen compound metabolic process	3.90E−04	Extracellular region	1.33E−03		
Cellular component organization or biogenesis	4.77E−04	Plasma membrane	5.48E−03		
Positive regulation of viral process	4.86E−04	Whole membrane	8.94E−03		
Establishment of localization	6.03E−04				
Regulation of cell junction assembly	7.11E−04				
Amyloid precursor protein metabolic process	7.30E−04				
Signaling	7.61E−04				
Regulation of symbiosis, encompassing mutualism through parasitism	9.15E−04				
Organic substance transport	9.22E−04				
Regulation of viral life cycle	1.04E−03				
Regulation of biological quality	1.31E−03				
Regulation of cellular component organization	1.69E−03				
Vesicle-mediated transport	2.18E−03				
Cellular metabolic process	2.21E−03				
Cell−cell junction organization	2.53E−03				
Positive regulation of multicellular organismal process	2.74E−03				
Viral process	2.81E−03				
Endocytosis	3.08E−03				
Cellular nitrogen compound metabolic process	3.48E−03				
Positive regulation of cellular component biogenesis	5.04E−03				
Regulation of biological process	5.23E−03				
Catabolic process	6.34E−03				
Positive regulation of biological process	1.09E−02				
Entry into host cell	1.27E−02				
Entry into other organism involved in symbiotic interaction	1.27E−02				
Proteolysis	1.72E−02				
Positive regulation of reactive oxygen species metabolic process	1.85E−02				
Regulation of defense response	2.06E−02				
Regulation of multi-organism process	2.09E−02				
Regulation of multicellular organismal process	2.12E−02				
Regulation of viral entry into host cell	2.49E−02				
Gap junction assembly	2.51E−02				
Cell proliferation	2.52E−02				
Regulation of response to stress	2.57E−02				
Receptor biosynthetic process	2.68E−02				
Organonitrogen compound catabolic process	2.79E−02				
Macromolecule metabolic process	2.91E−02				
Response to stress	3.85E−02				
Regulation of cardiac muscle contraction	3.93E−02				
Zinc ion binding	4.07E−02				
Organonitrogen compound metabolic process	4.45E−02				

Table 2. Selected over-represented GOs in genes significantly correlated with ACE2.

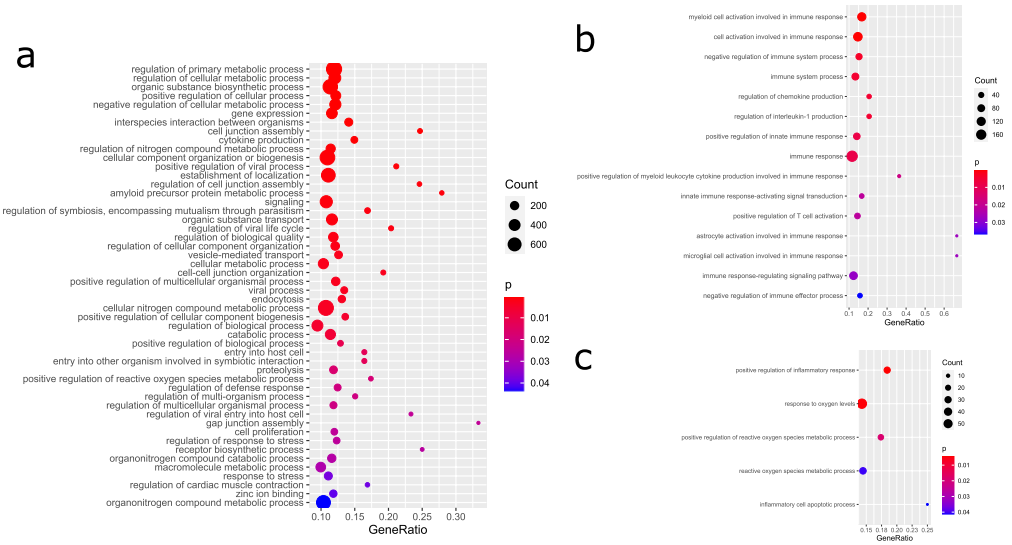


Figure 5. GO analysis reflects virus entry and immune response involving ROS and inflammation. (**a**) Selected GOs (Biological processes) shows interaction of virus and host, cell-junction organization, endocytosis, reaction involving cytokine production. (**b**) Immunity-related GOs illustrate the immune response involving activation of myeloid cells and T-cells and interleukin-1 and other chemokine production. (**c**) GOs associated with ROS and inflammation demonstrate involvement of ROS and inflammation leading to apoptosis.

Term	P value
Myeloid cell activation involved in immune response	5.3978E−07
Cell activation involved in immune response	4.3656E−05
Negative regulation of immune system process	0.00331363
Immune system process	0.0048604
Regulation of chemokine production	0.00486129
Regulation of interleukin-1 production	0.00486129
Positive regulation of innate immune response	0.0078574
Immune response	0.00855493
Positive regulation of myeloid leukocyte cytokine production involved in immune response	0.01809849
Innate immune response-activating signal transduction	0.02190672
Positive regulation of T cell activation	0.02209419
Astrocyte activation involved in immune response	0.02764374
Microglial cell activation involved in immune response	0.02764374
Immune response-regulating signaling pathway	0.0286071
Negative regulation of immune effector process	0.03757784

Table 3. GOs (all biological process) related to immune system in genes correlated with *ACE2*.

Term	P value
Positive regulation of inflammatory response	0.0039263
Response to oxygen levels	0.00438988
Positive regulation of reactive oxygen species metabolic process	0.01851891
Reactive oxygen species metabolic process	0.04088524
Inflammatory cell apoptotic process	0.04206722

Table 4. GOs (all biological process) related to inflammation and reactive oxygen species (ROS) in genes correlated with *ACE2*.

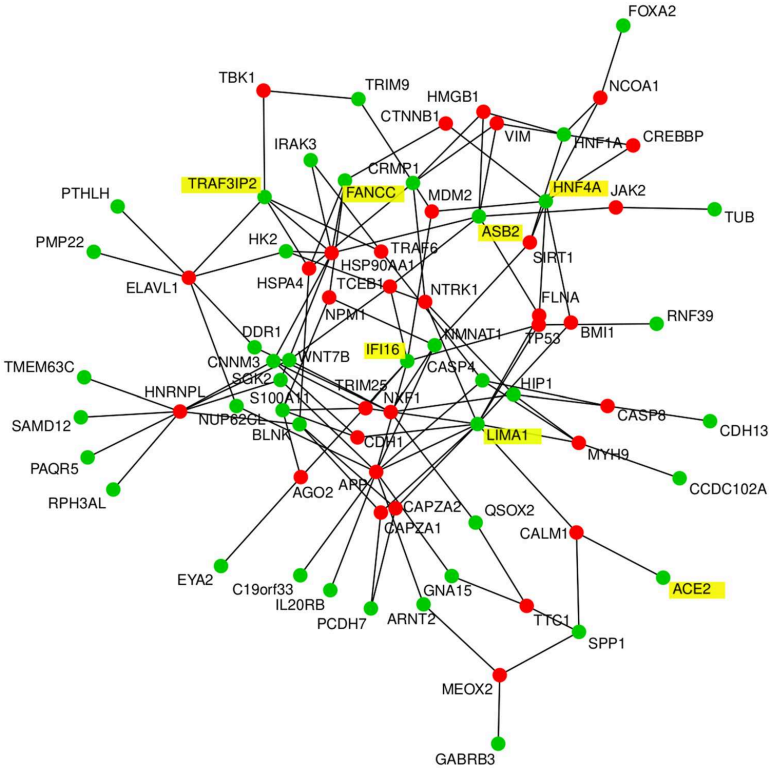


Figure 6. Protein interaction network of genes most significantly (Bonferoni-adjusted p < 1E−11) correlated and anti-correlated genes with ACE2 shows IFI16 (r = 0.8719), LIMA1 (r = 0.8955), CNNM3 (r = − 0.8732), HNF4A (r = − 0.8750), TRAF3IP2 (r = 0.8816), ASB2 (r = 0.8791) and FANCC (r = − 0.8682) as hub genes. Genes found as interactors in the BioGrid database are marked in red, the original geneset of *ACE2*-correlated genes is marked in green, hub genes and *ACE2* have yellow shading.

However, besides TMPRSS2 other proteases might be involved—see the heatmap in Fig. 3b which illustrates expression of transmembrane serine proteases. The highly significant co-expression of TMPRSS4 with ACE2 and its relevant role as transmembrane serine protease has enabled us to hypothesize that TMPRSS4 might also be involved in priming the SARS-CoV-2 spike protein. We therefore anticipate that inhibitors of TMPRSS4, TMPRS11D and TMPRS11E—besides those for TMPRSS2—could be a promising subject of further research.

Validation with a dataset of SARS-CoV-2 infected human bronchial organoids. We validated our results with the RNAseq dataset GSE150819 of human bronchial organoids infected with SARS-CoV-2[38]. Suppl. Table 8 shows the correlation of members of the TMPRSS gene family with ACE2 in this dataset. As in the first analysis of lung epithelial cells (Fig. 2) TMPRSS4 is also the gene most significantly correlated with ACE2 expression (p = 4.45E−05, Benjamini-Hochberg-corrected p = 0.0006, r = 0.86). Also TMPRSS2 (p = 0.014379, Benjamini-Hochberg-corrected p = 0.042562, r = 0.62) and TMPRSS11D (p = 0.000456, Benjamini-Hochberg-corrected p = 0.003185, r = 0.79) have significant p values and also Benjamini-Hochberg-corrected p values while TMPRSS11E here is not significant.

Discussion

In this meta-analysis we compared RNA-seq data of lung cells infected with SARS-CoV-2 and other lung cells with particular focus on correlated gene expression with the known SARS-CoV-2 receptor gene *ACE2*. We identified a signature of genes positively or negatively correlated with *ACE2* amongst which the most outstanding was the transmembrane serine protease *TMPRSS4*. In a recent publication Hoffmann et al.[17]. Inferred from the knowledge that the preceeding virus, SARS-CoV, uses ACE2 as receptor for entry and the serine protease TMPRSS2 for spike protein priming that the new virus SARS-CoV-2[17] would do the same. While the involvement of the receptor ACE2 appears to be established[3,16] the use of TMPRSS2 for spike protein priming appears not

fully settled as Hoffmann et al. still have to concede "that SARS-CoV-2 infection of Calu-3 cells was inhibited but not abrogated by camostat mesylate" (serine protease inhibitor with activity against TMPRSS2)[17]. The high significance (r = 0.9142, p = 4.59E−20) in our co-expression analysis with ACE2 suggests that TMPRSS4 is a considerable candidate for spike protein priming. That is in line with findings by Zang et al. who reported that TMPRSS4 besides TMPRSS2 enhances infection of small intestinal enterocytes with SARS-CoV-2[39]. However, TMPRSS4 is closely related to TMPRSS2 which both can proteolytically cleave the hemagglutinin of influenza viruses[40]. Further transmembrane serine proteases TMPRS11D (r = 0.9018, p = 9.61E−19) and TMPRS11E (r = 0.9121, p = 7.91E−20) in our analysis also emerged more significant than TMPRSS2 (r = 0.767, p = 1.3E−10). The TMPRS11 family member TMPRS11A was found to enhance viral infection with the first coronavirus SARS-CoV by spike protein cleavage in the airway[41]. Thus, we should not exclude the probability that other members of the TMPRSS gene family may be proteases for the SARS-CoV-2 spike protein. TMPRSS2 inhibitors have been proposed by Hoffmann et al.[17]—and earlier for the SARS-Cov virus by Kawase et al.[42] as working best together with cathepsin B/L inhibitors. We propose to investigate the effect of TMPRSS4 inhibitors in further research. As TMPRSS4 has been implicated in the invasion and metastasis of several cancers it has also been considered as target for cancer therapy for which a modest inhibitory effect of the above mentioned inhibitors in TMPRSS4-overexpressing SW480 cells was reported[43]. Interestingly, also *TMPRSS2* is connected with epithelial carcinogenesis as consistently over-expressed in prostate cancer[44], and later a gene fusion of *TMPRSS2* and *ERG* was reported as the predominant molecular subtype of prostate cancer[45], where *TMPRSS2* however only contributes untranslated sequence[46]. Assuming that co-expression with *ACE2* is an indication that TMPRSS4 may prime the SARS-CoV-2 spike protein we suggest that further development and testing of more effective TMPRSS4 inhibitors in in vitro and in vivo models could support the translation into clinical settings. However, we have to state the limitation that this study is a meta-analysis based solely on transcriptome and not proteome data.

Besides the identification of *TMPRSS4,* we found several significantly over-represented GOs and pathways such as Endocytosis, Papilloma virus infection and Bacterial invasion of epithelial cells for which we provide full gene lists to foster further elucidation of disease mechanisms. Genes from the constructed protein-interaction network provide a first snapshot of a comprehensive image: *IFI16* controls the interferon response to the virus[31], *LIMA1* mediating the interaction between Cadherins (*CDH1*, *CDH13*) and Actin in the context of adherens junctions potentially playing a role in virus infection, *CALM1* inhibits shedding of the ectodomain of the virus receptor *ACE2*[36].

Furthermore, GO analyses revealed several biological processes related to viral cell entry, host reaction, immune response, ROS, inflammation and apoptosis. This led us to propose a cascade of events taking place post SARS-CoV-2 entry into host cells- illustrated in Fig. 7 together with possible drug targets. The coronavirus SARS-CoV-2 docks at the receptor ACE2 on the membrane of the human epithelial cell, the early stage of infection. According to reports by Monteil et al. these processes can be blocked with recombinant hrsACE2[16]. Transmembrane serine proteases TMPRSS mediate SARS-CoV-2 cell entry via ACE2[17,47]. TMPRSS2 has been described as a mediator of ACE2-coupled endocytosis in the first SARS-CoV[48] and by a previous publication also for SARS-CoV-2[17]. However, we identified high levels of co-expression between ACE2 and TMPRSS4 and other members of the TMPRSS family and hypothesize that any of these additional family members might have the same function as the well described TMPRSS2. As a consequence, we propose that besides TMPRSS2 also other TMPRSS family members could be targets of pharmaceutical intervention warranting further research. After entering the cell, SARS-CoV-2 RNA is released, replicated and packaged again. Drugs can target the viral protease and the polymerase needed for replication[49]. Replication can further be inhibited by interferon and interferon-stimulated genes (ISG)[28] which we also found evidence for in negatively correlated replication pathways (e.g. DNA replication and homologous recombination). This depends on a healthy immune response and may be impaired in individuals with a weak immune system due to age or underlying diseases. The virus is packaged and released into the extracellular space where it can be attacked by macrophages chemo-attracted by CXCL17[24]. Also T-Cells can be involved in the immune response. We found evidence for their activation in GO analysis by associated interleukins IL1 and IL7. Although immunity is not the main focus of this manuscript it is tempting to speculate that the severity of the clinical manifestations such as the acute respiratory failure and also failure in other organs depend on the state of the immune system which decreases with age or diseases such as diabetes.The involvement of ACE2 in the renin-angiotensin system as antagonist of ACE in regulating blood pressure via Angiotensin II, vasoconstriction, dilation and its protective role against lung injury[41] are additional factors which correlate with age[50]. This is confirmed by reports from Wadmann et al.[51] about Centers for Disease Control and Prevention (CDC) data from 14 U.S. states that 50% hospitalized COVID-19 patients had pre-existing high blood pressure[51]. In their study about ACE2 in the preceeding SARS-CoV virus Imai et al.[21] found that *ACE2* protects against lung injury and acid-induced lung injury in a *Ace2*-knockout mouse can be improved by an inhibitor of the Angiotensin II receptor AT1[21]. The results of clinical studies but also statistics on hypertension and even more important statistics on pharmacological treatment of hypertension in COVID-19 patients may shed light on the discussions if treatment with ACE inhibitors and angiotensin receptor blockers (ARBs) are detrimental[52] or beneficial[53].

We conclude, that our meta-analysis of RNA-Seq data of lung cells partially infected with SARS-CoV-2 identified the transmembrane serine protease *TMPRSS4* as one of the most significantly correlated genes with the virus receptor *ACE2*. The importance of this finding is underlined by Zang et al. who simultaneously with our preprint publication reported that TMPRSS4 enhances SARS-CoV-2 infection of small intestinal enterocytes[39]. We propose that inhibitors of TMPRSS family members TMPRSS4, TMPRSS11D and TMPRSS11E besides TMPRSS2 are worthwhile testing in i*n vitro* and i*n vivo* studies. As clinicians, pathologists and scientists are struggling to get to grips with and understand the damage wrought by SARS-CoV-2 as it invades the body, we hope that our analyses and dataset will contribute to a better understanding of the molecular basis of COVID-19.

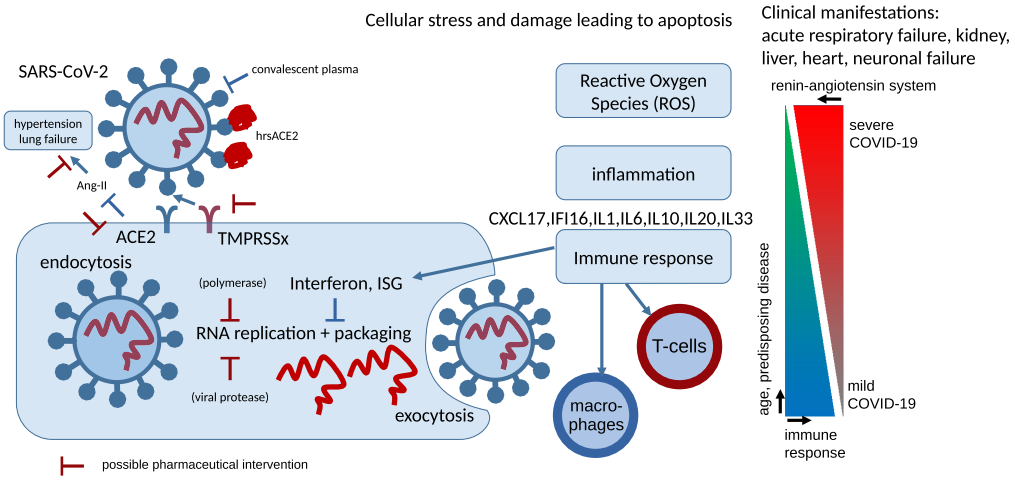


Figure 7. Scheme of SARS-CoV-2 infection. The coronavirus SARS-CoV-2 docks at the receptor ACE2 on the membrane of the human epithelial cell. Transmembrane serine proteases TMPRSSx mediate SARS-CoV-2 cell entry via ACE2. TMPRSS2 was reported for this in the first SARS-CoV and by previous publication also for SARS-CoV-2 but we hypothesize that due to co-expression with ACE2, TMPRSS4 and other members of the TMPRSS family may well perform this task. We suggest that inhibitors of TMPRSS4 and other TMPRSS family members might have therapeutic potential. Upon entry into the cell, viral RNA is released, replicated and packaged again. Replication can be inhibited by interferon and interferon stimulated genes (ISG) what we also saw in negatively correlated replication pathways (e.g. DNA replication and homologous recombination). This indicates a healthy immune response and may be impaired in persons with a weak immune system due to age or disease. The packaged virus is released from the cell and can be attacked by macrophages chemo-attracted by CXCL17—or T-cells for which we found evidence for by GO analysis and by associated interleukins IL1 and IL7. It is tempting to speculate that the severity of the clinical manifestations such as the acute respiratory failure and also failure in other organs depends on the quality of the immune system decreasing with age or diseases such as diabetes. The involvement of ACE2 in the renin-angiotensin system as antagonist of ACE in regulating blood pressure via Angiotensin II (Ang-II), vasoconstriction, dilation and its protective role against lung injury are additional factors which correlate with age[21,50,53].

Methods

Sample collection of lung cell RNA-Seq data. Next-generation sequencing datasets measured in RNA-Seq experiments with lung cells (GSE147507: Illumina NextSeq 500; GSE146482: Illumina NovaSeq 6000; GSE85121: Illumina HiSeq 2500) were downloaded from NCBI GEO (Table 1, suppl. Table 4). These datasets were provided along with studies by Blanco-Melo et al.[23] (GSE147507) and Staudt *et al.*[54] (GSE85121). A final publication related to the dataset GSE146482 is yet to materialize. From accession no. GSE85121 only small airway epithelium brushing cells were used but alveolar macrophages were excluded while from accession no. GSE147507 only human epithelial and adenocarcinoma lung cells were used but ferret cells and updates after March 24 were excluded and from accession no. GSE146482 only control epithelial cells were used but graphene oxide treated cells were excluded. After exclusion of the above mentioned datasets not fitting the target cell type, 49 samples remained useful.

Data normalization and analysis. After the excluded samples had been filtered from the downloaded RNA-Seq data data was imported into the R/Bioconductor environment[55,56]. Read counts from accesion nos. GSE147507 and GSE146482 were converted to FPKM (fragments per kilobase of exon model per million reads mapped) using trancript lengths downloaded from ENSEMBL (version GRCh38, p13). Batch effects were removed with the package sva[57] employing method ComBat[58]. Normalization was performed via the voom method[59]. Pearson correlation coefficients between samples were calculated with the R-builtin method *cor*. Dendrograms were drawn employing the dendextend package[60] filtering genes for high coefficient of variation above the 75% quantile.

The validation dataset GSE150819 of human bronchial organoids infected with SARS-CoV-2[38] downloaded from NCBI GEO was normalized with the voom method[59] before calculating correlation of each gene to ACE2 gene expression.

Detection of genes correlated with ACE2. The Pearson correlation of the normalized gene expression values for all samples was calculated between the gene *ACE2* and each other gene. The method cor.test was applied to calculate the p value for the t test for Pearson correlation. The p value was Bonferoni-corrected by divi-

sion by the number of genes and additionally adjusted via the Benjamini–Hochberg method[61]. Correlated genes were filtered with a very restrictive criterion (Bonferoni-adjusted-p < 1E−11)—in order to get human readable numbers of genes for heatmap and network generation- and conventional criteria (r > 0.6, Bonferoni-adjuste-p < 0.05) for positively correlated genes and (r < 0.6, Bonferoni-adjusted-p < 0.05) for negatively correlated genes.

Pathway and gene ontology (GO) analysis. The R package GOstats[62] was employed for over-representation analysis of positively and negatively correlated genes with the SARS-CoV-2 receptor gene *ACE2*. KEGG pathway annotations which had been downloaded from the KEGG database[27] in March 2018 were used for testing over-representation of the positively and negatively ACE2-correlated genes via the R-builtin hypergeometric test.

Dot plots showing the p value of the hypergeometric test, the ratio of significant genes to all genes in the pathway and the number of significant genes per pathway were plotted via the package *ggplot2*[63].

Protein-interaction networks. A human protein interaction network was constructed in a similar manner as we described in our previous publication[64]. However, here only direct interactors and no further interactors of interactors were extracted from the Biogrid database version 3.4.161[65] using the restrictively filtered (Bonferoni-adjusted-p < 1E−1) genes significantly correlated or anti-correlated with ACE2 gene expression. The network was reduced to the n = 30 nodes with most interactions and was plotted via the R package *network*[66] showing original genes in green and BioGrid-derived interactors in red.

Data availability

No datasets were generated during the current study. The datasets used for the meta-analysis are available at the National Center for Biotechnology Information (NCBI) Gene expression Omnibus (GEO) accessions referred to in Table 1.

Received: 2 June 2020; Accepted: 17 November 2020
Published online: 08 December 2020

References

1. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020).
2. Zhou, H. *et al.* A Novel Bat Coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Curr. Biol.* **30**(2196), e3-2203.e3 (2020).
3. Zhang, T., Wu, Q. & Zhang, Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr. Biol.* **30**(1346), e2-1351.e2 (2020).
4. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
5. Tilocca, B. *et al.* Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. *Microbes Infect.* **22**, 188–194 (2020).
6. Tilocca, B. *et al.* Immunoinformatic analysis of the SARS-CoV-2 envelope protein as a strategy to assess cross-protection against COVID-19. *Microbes Infect.* **22**, 182–187 (2020).
7. Tilocca, B. *et al.* Molecular basis of COVID-19 relationships in different species: a one health perspective. *Microbes Infect.* **22**, 218–220 (2020).
8. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* https://doi.org/10.1016/S1473-3099(20)30120-1 (2020).
9. Verity, R. *et al.* Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* https://doi.org/10.1016/S1473-3099(20)30243-7 (2020).
10. Wang, Y. *et al.* Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *The Lancet* https://doi.org/10.1016/S0140-6736(20)31022-9 (2020).
11. Wang, M. *et al.* Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res.* **30**, 269–271 (2020).
12. Lai, C.-C., Shih, T.-P., Ko, W.-C., Tang, H.-J. & Hsueh, P.-R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *Int. J. Antimicrob. Agents* **55**, 105924 (2020).
13. Yamamoto, N., Matsuyama, S., Hoshino, T. & Yamamoto, N. Nelfinavir inhibits replication of severe acute respiratory syndrome coronavirus 2 in vitro. *BioRxiv*. https://doi.org/10.1101/2020.04.06.026476 (2020)
14. Beigel, J. H. *et al.* Remdesivir for the Treatment of Covid-19—final report. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMoa2007764 (2020).
15. Casadevall, A. & Pirofski, L. The convalescent sera option for containing COVID-19. *J. Clin. Invest.* **130**, 1545–1548 (2020).
16. Monteil, V. *et al.* Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. *Cell* https://doi.org/10.1016/j.cell.2020.04.004 (2020).
17. Hoffmann, M. *et al.* SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* https://doi.org/10.1016/j.cell.2020.02.052 (2020).
18. Matsuyama, S. *et al.* Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proc. Natl. Acad. Sci.* **117**, 7001–7003 (2020).
19. Sungnak, W. *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **26**, 681–687 (2020).
20. Collin, J. *et al.* Co-expression of SARS-CoV-2 entry genes in the superficial adult human conjunctival, limbal and corneal epithelium suggests an additional route of entry via the ocular surface. *Ocul. Surf.* https://doi.org/10.1016/j.jtos.2020.05.013 (2020).
21. Imai, Y. *et al.* Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* **436**, 112–116 (2005).
22. Haga, S. *et al.* Modulation of TNF- -converting enzyme by the spike protein of SARS-CoV and ACE2 induces TNF- production and facilitates viral entry. *Proc. Natl. Acad. Sci.* **105**, 7809–7814 (2008).
23. Blanco-Melo, D. *et al.* Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**(1036), e9-1045.e9 (2020).
24. Maravillas-Montero, J. L. *et al.* Cutting edge: GPR35/CXCR8 is the receptor of the mucosal chemokine CXCL17. *J. Immunol. Baltim. Md* **1950**(194), 29–33 (2015).

25. Thomas, A. C., Tattersall, D., Norgett, E. E., O'Toole, E. A. & Kelsell, D. P. Premature terminal differentiation and a reduction in specific proteases associated with loss of ABCA12 in Harlequin ichthyosis. *Am. J. Pathol.* **174**, 970–978 (2009).
26. Wilk, E. *et al.* RNAseq expression analysis of resistant and susceptible mice after influenza A virus infection identifies novel genes associated with virus replication and important for host resistance to infection. *BMC Genomics* **16**, 655 (2015).
27. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **47**, D590–D595 (2019).
28. Schoggins, J. W. & Rice, C. M. Interferon-stimulated genes and their antiviral effector functions. *Curr. Opin. Virol.* **1**, 519–525 (2011).
29. Gillespie, K. A., Mehta, K. P., Laimins, L. A. & Moody, C. A. Human papillomaviruses recruit cellular DNA repair and homologous recombination factors to viral replication centers. *J. Virol.* **86**, 9520–9526 (2012).
30. Zhang, Y. *et al.* Clinical and coagulation characteristics of 7 patients with critical COVID-2019 pneumonia and acro-ischemia. *Zhonghua Xue Ye Xue Za Zhi Zhonghua Xueyexue Zazhi* **41**, E006 (2020).
31. Thompson, M. R. *et al.* Interferon γ-inducible protein (IFI) 16 transcriptionally regulates type I interferons and other interferon-stimulated genes and controls the interferon response to both DNA and RNA viruses. *J. Biol. Chem.* **289**, 23568–23581 (2014).
32. Chan, M. C. *et al.* Tropism and innate host responses of a novel avian influenza A H7N9 virus: an analysis of ex-vivo and in-vitro cultures of the human respiratory tract. *Lancet Respir. Med.* **1**, 534–542 (2013).
33. Jamaluddin, M., Tian, B., Boldogh, I., Garofalo, R. P. & Brasier, A. R. Respiratory syncytial virus infection induces a reactive oxygen species-MSK1-phospho-ser-276 RelA pathway required for cytokine expression. *J. Virol.* **83**, 10605–10615 (2009).
34. Meng, W. & Takeichi, M. Adherens junction: molecular architecture and regulation. *Cold Spring Harb. Perspect. Biol.* **1**, a002899 (2009).
35. Lin, L.-T. & Richardson, C. The host cell receptors for measles virus and their interaction with the viral hemagglutinin (H) protein. *Viruses* **8**, 250 (2016).
36. Lambert, D. W., Clarke, N. E., Hooper, N. M. & Turner, A. J. Calmodulin interacts with angiotensin-converting enzyme-2 (ACE2) and inhibits shedding of its ectodomain. *FEBS Lett.* **582**, 385–390 (2008).
37. Hein, M. Y. *et al.* A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163**, 712–723 (2015).
38. Suzuki, T. *et al.* Generation of human bronchial organoids for SARS-CoV-2 research. *biorxiv*. https://doi.org/10.1101/2020.05.25.115600 (2020).
39. Zang, R. *et al.* TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. *Sci. Immunol.* https://doi.org/10.1126/sciimmunol.abc3582 (2020).
40. Bertram, S., Glowacka, I., Steffen, I., Kühl, A. & Pöhlmann, S. Novel insights into proteolytic cleavage of influenza virus hemagglutinin. *Rev. Med. Virol.* **20**, 298–310 (2010).
41. Kam, Y.-W. *et al.* Cleavage of the SARS coronavirus spike glycoprotein by airway proteases enhances virus entry into human bronchial epithelial cells in vitro. *PLoS ONE* **4**, e7870 (2009).
42. Kawase, M., Shirato, K., van der Hoek, L., Taguchi, F. & Matsuyama, S. Simultaneous treatment of human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute respiratory syndrome coronavirus entry. *J. Virol.* **86**, 6537–6545 (2012).
43. de Aberasturi, A. L. & Calvo, A. TMPRSS4: an emerging potential therapeutic target in cancer. *Br. J. Cancer* **112**, 4–8 (2015).
44. Bugge, T. H., Antalis, T. M. & Wu, Q. Type II transmembrane serine proteases. *J. Biol. Chem.* **284**, 23177–23181 (2009).
45. Tomlins, S. A. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644–648 (2005).
46. Tomlins, S. A. *et al.* Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia N. Y. N* **10**, 177–188 (2008).
47. Meng, T. *et al.* The insert sequence in SARS-CoV-2 enhances spike protein cleavage by TMPRSS. *biorxiv*. https://doi.org/10.1101/2020.02.08.926006 (2020).
48. Glowacka, I. *et al.* Differential downregulation of ACE2 by the spike proteins of severe acute respiratory syndrome coronavirus and human coronavirus NL63. *J. Virol.* **84**, 1198–1205 (2010).
49. Huang, J., Song, W., Huang, H. & Sun, Q. Pharmacological therapeutics targeting RNA-dependent RNA polymerase, proteinase and spike protein: from mechanistic studies to clinical trials for COVID-19. *J. Clin. Med.* **9**, 1131 (2020).
50. Conti, S., Cassis, P. & Benigni, A. Aging and the renin-angiotensin system. *Hypertension* **60**, 878–883 (2012).
51. Wadman, M., Couzin-Frankel, J., Kaiser, J. & Matacic, C. A rampage through the body. *Science* **368**, 356–360 (2020).
52. Fang, L., Karakiulakis, G. & Roth, M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection?. *Lancet Respir. Med.* **8**, e21 (2020).
53. Vaduganathan, M. *et al.* Renin–angiotensin–aldosterone system inhibitors in patients with Covid-19. *N. Engl. J. Med.* **382**, 1653–1659 (2020).
54. Staudt, M. R., Salit, J., Kaner, R. J., Hollmann, C. & Crystal, R. G. Altered lung biology of healthy never smokers following acute inhalation of E-cigarettes. *Respir. Res.* **19**, 78 (2018).
55. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2018).
56. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
57. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinforma. Oxf. Engl.* **28**, 882–883 (2012).
58. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostat. Oxf. Engl.* **8**, 118–127 (2007).
59. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
60. Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720 (2015).
61. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
62. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
63. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis* (Springer, Berlin, 2009).
64. Wruck, W. & Adjaye, J. Meta-analysis of human prefrontal cortex reveals activation of GFAP and decline of synaptic transmission in the aging brain. *Acta Neuropathol. Commun.* **8**, 26 (2020).
65. Chatr-aryamontri, A. *et al.* The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* **45**, D369–D379 (2017).
66. Butts, C. T. Network: a package for managing relational data in R. *J. Stat. Softw.* **24**, 1–24 (2008).

Acknowledgments

James Adjaye acknowledges support from the Medical faculty of the Heinrich-Heine University, Duesseldorf.

Author contributions

W.W. analysed the data and wrote the manuscript. J.A. supervised the work, co-wrote the manuscript and gave the final approval.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-78402-2.

Correspondence and requests for materials should be addressed to J.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2.1.10 Transmission of SARS-COV-2 from China to Europe and West Africa: a detailed phylogenetic analysis.

bioRxiv 2020.10.02.323519v1; doi: doi.org/10.1101/2020.10.02.323519

Wruck W(1), Adjaye J(1).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Moorenstr.5, 40225, Düsseldorf, Germany. james.adjaye@med.uni-duesseldorf.de.

Background: SARS-CoV-2, the virus causing the Covid-19 pandemic emerged in December 2019 in China and raised fears that it could overwhelm healthcare systems worldwide. In June 2020, all African countries registered human infections with SARS-CoV-2. The virus is mutating steadily and this is monitored by a well curated database of viral nucleotide sequences from samples taken from infected individual thus enabling phylogenetic analysis and phenotypic associations.

Methods: We downloaded from the GISAID database, SARS-CoV-2 sequences established from four West African countries Ghana, Gambia, Senegal and Nigeria and then performed phylogenetic analysis employing the nextstrain pipeline. Based on mutations found within the sequences we calculated and visualized statistics characterizing clades according to the GISAID nomenclature.

Results: We found country-specific patterns of viral clades: the later Europe-associated G-clades predominantly in Senegal and Gambia, and combinations of the earlier (L, S, V) and later clades in Ghana and Nigeria. Contrary to our expectations, the later Europe-associated G-clades emerged before the earlier clades. Detailed analysis of distinct samples showed that some of the earlier clades might have circulated latently and some reflect migration routes via Mali and Tunisia.

Conclusions: The distinct patterns of viral clades in the West African countries point at its emergence from Europe and China via Asia and Europe. The observation that the later clades emerged before the earlier clades could be simply due to founder effects or due to latent circulation of the earlier clades. Only a marginal correlation of the G-clades associated with the D614G mutation could be identified with the relatively low case fatality (0.6-3.2).

DOI: 10.1101/2020.10.02.323519

URL: https://www.biorxiv.org/content/10.1101/2020.10.02.323519v1

Contribution: Wasco Wruck came up with the concept of the study, wrote the manuscript, analysed and interpreted the data.

Transmission of SARS-COV-2 from China to Europe and West Africa: a detailed phylogenetic analysis.

Wasco Wruck[1] and James Adjaye[1,*]

[1]Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany

[*]corresponding author: James Adjaye (james.adjaye@med.uni-duesseldorf.de)

Abstract

Background: SARS-CoV-2, the virus causing the Covid-19 pandemic emerged in December 2019 in China and raised fears that it could overwhelm healthcare systems worldwide. In June 2020, all African countries registered human infections with SARS-CoV-2.

The virus is mutating steadily and this is monitored by a well curated database of viral nucleotide sequences from samples taken from infected individual thus enabling phylogenetic analysis and phenotypic associations.

Methods: We downloaded from the GISAID database, SARS-CoV-2 sequences established from four West African countries Ghana, Gambia, Senegal and Nigeria and then performed phylogenetic analysis employing the nextstrain pipeline. Based on mutations found within the sequences we calculated and visualized statistics characterizing clades according to the GISAID nomenclature.

Results: We found country-specific patterns of viral clades: the later Europe-associated G-clades predominantly in Senegal and Gambia, and combinations of the earlier (L, S, V) and later clades in Ghana and Nigeria. Contrary to our expectations, the later Europe-associated G-clades emerged before the earlier clades. Detailed analysis of distinct samples showed that some of the earlier clades might have circulated latently and some reflect migration routes via Mali and Tunisia.

Conclusions: The distinct patterns of viral clades in the West African countries point at its emergence from Europe and China via Asia and Europe. The observation that the later clades emerged before the earlier clades could be simply due to founder effects or due to latent circulation of the earlier clades. Only a marginal correlation of the G-clades associated with the D614G mutation could be identified with the relatively low case fatality (0.6-3.2).

Key messages

- Ghana and Nigeria have a combination of earlier (L, V, S) and later Europe-associated G-clades of SARS-CoV-2, therefore pointing to multiple introductions while in Senegal and Gambia Europe-associated G-clades predominate pointing to introductions mainly from Europe.
- Surprisingly, the later G-clades emerged before the earlier clades (L, V, S)
- Detailed phylogenetic analysis points at latent circulation of earlier clades before the first registered cases.
- Phylogenetic analysis of some cases points at migration routes to Europe via Tunisia, Egypt and Mali.
- A marginal correlation of r=0.28 between the percentage of the D614G mutation defining the G-clades and case-fatality can be detected.

Introduction

The Covid-19 pandemic resulting from the SARS-CoV-2 coronavirus infection which emerged in December 2019 in Wuhan, China, spread all over the world and after a delay of a few months also appeared on the African continent. Early in June 2020, all African countries registered human infections with SARS-CoV-2. Starting with the first sequenced human sample of SARS-CoV-2, several mutations of the virus sequence arose which could be grouped into clades allowing associations with

regional prevalences. In this study, we focus on samples from West Africa which was the region from which the first African SARS-CoV-2 sequences became available. We aimed at analysing phylogenetic characteristics possibly giving clues about the distribution between countries and eventually even about putative severity changes between specific clades. We use the nomenclature of clades suggested by the GISAID initiative and adopted in several publications. Previous studies reported potential impact of the D614G amino acid mutation which is induced by the A23403G single nucleotide polymorphism (SNP) (1) and associated with the branch of the phylogenetic tree referred to as clade G.

Brufsky hypothesized that the higher number of deaths on the East coast of the United States compared to the West coast could be due to the higher prevalence of the D614G mutation on the East coast (2). The D614G mutation has been suggested to affect the adherence of the virus to the cell membrane and as a consequence results in higher virulence. Supportive evidence was reported in mice (3), (4). Korber et al. hypothesized that the D614 amino acid on the surface of the spike protein protomer region S1 of the virus could have a hydrogen bond to the T859 amino acid in the S2 region residing on the membrane (1). Furthermore, they showed that clade G rapidly starts to replace other clades associated with the D614 amino acid in each country entered (1). In mid-March 2020, the G clade was found almost exclusively in Europe (5) but soon after spread all over the world. Korber et al. see its origin from China or Europe (1). In China, four early samples carried the D614G mutation. One sample from January 24th 2020 had only the A23403G (D614G) but not the C3037T and C14408T mutations which usually associate with A23403T in clade G. Three samples with the D614G were related to the first German sample. In Europe the first German sample from January 28th carried the A-to-G mutation at nucleotide position 23403 (D614G) mutation and the C-to-T mutation at position 3037, but not the mutation at position 14408. The first sample carrying all of the above mentioned mutations plus the C241T in the Untranslated Region (UTR) was identified in Italy on February 20th 2020 (1). Another interesting feature of the G clades is that the associated C14408T

mutation adjacent to the RNA dependent RNA polymerase (RdRp) putatively increases the mutation rate as Pachetti *et al.* report (5).

Detailed analyses of virus evolution have been performed for some countries, e.g. France (6), New York (7) and India (8). For France it could be deduced by the distinction between clade G and the earlier phylogenetic branches that the first SARS-CoV-2 did not lead to local transmission while the clade G was circulating for a considerable time before the first recorded case which was of clade G and had no travel events or traveller contact (6).

Since the declaration of Covid-19 as pandemic by the WHO on March 11th 2020, fears were expressed that it could overwhelm weaker healthcare systems as existing in many African countries. Furthermore, hygiene, social distancing and lockdown face many challenges in countries with high percentages without clean running water, cramped confines and the dependence on a daily income. However, Africa has the advantage of a very young population, e.g. in Sub-Saharan Africa with a median age of 19.7 years (9) for which in average milder etiopathologies can be expected. Additionally, for the early outbreak a study evaluating air traffic from affected regions in China calculated relatively low transmission risks for most African countries except South Africa and Ethiopia (10). For later phases, Cabore et al. proposed a model estimating risk of exposure for African countries based on a Hidden Markov model which accounts for factors such as gathering, weather, distribution and hygiene with the conclusion that with respect to the high calculated infection rates effective containment is indispensable (11).

Here, we analysed SARS-CoV-2 nucleotide sequenes from samples obtained from the West African countries of Gambia, Ghana, Nigeria and Senegal in order to identify characteristic mutations and to dissect their patterns of distribution.

Methods

Sample collection

We downloaded SARS-CoV-2 viral sequences for West African samples and reference samples from European, North and South American countries and China from the GISAID database of June 2020. The samples used in this study are shown in Table 1.

Construction of the phylogenetic tree

The phylogenetic tree was constructed using a pipeline adapted from the Zika virus pipeline on the nextstrain.org web page (12) employing the Augur (12), the MAFFT (13) and the IQ-tree (14) software. Details of steps which were performed:

First all West African and reference sequences in FASTA format were aligned employing the augur command:

augur align --sequences westafrica.fasta --reference-sequence sars_cov2_referencesequence.gb --output wa_aligned.fasta --fill-gaps

which called MAFFT (13) with the command:

mafft --reorder --anysymbol --nomemsave --adjustdirection --thread 1 wa_aligned.fasta.to_align.fasta 1> wa_aligned.fasta 2> wa_aligned.fasta.log

Metadata was extracted from the sequences FASTA via the augur command:

augur parse --sequences wa_aligned.fasta --fields strain accession date --output-sequences wa_aligned_parsed.fasta --output-metadata metadata.tsv

Then a tree was built via the augur command:

augur tree --alignment wa_aligned_parsed.fasta --output wa_tree_raw.nwk

Calling the IQ-tree algorithm (14) via this command:

iqtree -ninit 2 -n 2 -me 0.05 -nt 1 -s wa_aligned_parsed-delim.fasta -m GTR > wa_aligned_parsed-delim.iqtree.log

The tree was refined via the Augur software calling TreeTime (15) for Maximum-Likelihood analysis inferring a time resolved phylogeny tree:

augur refine --tree wa_tree_raw.nwk --alignment wa_aligned_parsed.fasta --metadata metadata.tsv --output-tree wa_tree.nwk --output-node-data wa_branch_lengths.json --timetree --coalescent opt --date-confidence --date-inference marginal --clock-filter-iqd 4 --keep-polytomies

Here, the command from the Zika pieline was adapted to --keep-polytomies to keep all samples.

Metadata information was manually supplemented with country and region information and associated with the tree via a call to Augur:

augur traits --tree wa_tree.nwk --metadata metadata_countries.tsv --output wa_traits.json --columns region country --confidence

Augur was called to infer ancestral states of discrete character again using TreeTime (15) :

augur ancestral --tree wa_tree.nwk --alignment wa_aligned_parsed.fasta --output-node-data wa_nt_muts.json --inference joint

Amino acid mutations were identified with the augur translate command:

augur translate --tree wa_tree.nwk --ancestral-sequences wa_nt_muts.json --reference-sequence sars_cov2_referencesequence.gb --output wa_aa_muts.json

Results were exported via the augur command:

augur export v2 --tree wa_tree.nwk --metadata metadata_countries.tsv --node-data wa_branch_lengths.json wa_traits.json wa_nt_muts.json wa_aa_muts.json --colors colors.tsv --lat-longs lat_longs.tsv --auspice-config auspice_config.json --output wa_cov19.json

Visualization of the phylogenetic tree and annotation with mutations and clades

The phylogenetic tree was annotated with crucial mutations using the tool FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). Branches of the tree corresponding to clades following the nomenclature of GISAID were coloured distinctly.

World map chart

The world map chart was built using the R-package Rworldmap (16). Clade distribution pie charts were copied to the distinct country locations. Connections between countries were based on the nextstrain Africa analysis and our own auspice analysis. Further connections between countries were retrieved from literature on virus introductions into countries or regions. The first patient on the West Coast of the United States returned from a journey to Wuhan, China (17). The first introductions in New York came from multiple independent infected individuals mainly from Europe (7). The first cases in France and Europe were Chinese travellers from the predominantly affected Hubei province who entered the county in mid-January and were tested positive on January 24th 2020 (18). Patient zero in Germany was a Chinese resident from Wuhan visiting Germany (19). The Italian outbreak started with two Chinese travellers who arrived in Milan-Lombardy, went to Rome later on and were tested positive on January 31st 2020 (20). The first Italian citizen was confirmed for Covid-19 on February 21st 2020 in Lombardy (20). In the Netherlands, the first patient diagnosed on February 27th 2020 had probably infected himself on a trip to Northern Italy between February 18th and 21st (21). The first cases in the UK returned from travels to the Chinese Hubei province and were tested positive for SARS-CoV-2 on January 30th 2020 (22).

Results

Phylogenetic tree and diversity

The phylogenetic tree shown in Figure 1a displays similarities of West-African virus sequences with representative reference sequences from China and multiple European countries. The tree can be

divided into two major branches resulting from the A23403G (D614G) mutation. The branch at the bottom is directly associated with the first recorded sequences from Wuhan, China and does not carry the D614G mutation. The Nigerian samples cluster with these early Chinese samples in the bottom branch of the tree. The branch on top is associated with sequences prevalent in Europe as demonstrated by reference sequences from Germany, France, Italy, Austria, Netherlands and UK. Ghanaian samples are about equally distributed between the top (European) and bottom branch of the tree. Senegalese samples cluster close with the French reference sample at the top of the tree.

The phylogenetic tree can be viewed interactively via the nextstrain.org framework under the URL: https://nextstrain.org/community/wwruck/wa

The split of the tree by the A23403G (D614G) mutation into two major branches corresponds to the highest diversity found at that location (Figure 1b). This mutation resides within the spike protein.

Association with clades

We associated the West-African and reference samples via their characteristic mutations with clades according to the GISAID nomenclature. The phylogenetic tree in Figure 2 is coloured by these clades. The West-African samples are distributed over all clades suggesting introductions from China and European countries. However, each of the investigated countries has a specific pattern: most Senegalese samples have close similarity with the French reference, most Nigerian samples cluster in early Chinese-based clade S and Ghanaian samples are spread over all clades, the three Gambian samples are distributed over clades V, GR and GH. Within the clade S, there are putatively specific West-African mutations at the branches at C24370T and G22468T. Ghanaian samples predominate in the branch associated with the C24370T mutation. The branch determined by the mutation G22486T (Supplementary Figure 1) may reflect migration routes because in the nextstrain analysis of whole Africa there are also samples from Mali and Tunisia in this branch (https://nextstrain.org/ncov/africa?f_region=Africa, accessed August 14 2020). Two of the non-

French-related Senegalese samples emanate from the C24370T and G22468T branches while the other (Senegal/136) has strong similarity with Spanish end-February samples from the early clade S (Supplementary Figure 2) pointing at multiple introductions to Senegal from France, Spain and African countries.

Timeline of clade distribution

In the temporal course of the clade distribution in Figure 3, the increased share of the Europe-associated G-clades becomes obvious. The G-clades harbor the putatively more infective D614G mutation (1). Surprisingly, the later Europe-associated G-clades (G, GH, GR) emerged before the earlier clades L, S and V in West African sequenced samples. This could be due to founder effects by introductions from France closely connected to Senegal and displaying a similar clade distribution and by migration and travel routes such as in the first registered Nigerian case infected in Italy (23). Furthermore, the China-based L-,V- and S-clade samples were obtained in mid-March, a time point within the Wuhan lockdown and when the epidemic in China was nearly totally over. Thus, the virus may have circulated in several countries before the first samples were sequenced. Surprisingly, the abundance of the S-clade is relatively high mainly due to the contribution from Nigeria and Ghana. However, without the S-clade distribution, the change in abundance resembles the global one with a delay of about 2-4 weeks.

Country-specific patterns of clade distribution

Figure 4 shows that West African countries have acquired distinct patterns of China-and Europe-based clades. The first row contains the clade distribution charts of the West African countries investigated here whilst the second row contains charts of countries with comparable distributions. Nigeria has the highest percentage of the China-based early clades (L,S,V). Ghana has nearly equally distributed percentages of China and Europe-based clades (G,GH,GR) and in that sense has

similarities with the German distribution. Senegal's clade distribution resembles the one from France but includes also a few samples from the early China-based clades. There were only three sequences from Gambian, two from Europe-based clades GR and GH and one from China-based clade V. That pattern resembles the one from Italy when the clade G is substituted by the G-derived GH clade which however does not infer a connection to Italy but instead a similar combination of Chinese and European-related clades. Also the UK distribution in the last row has similarity with the Gambian distribution but as it includes also Chinese clades it also resembles the one from Ghana. The Dutch distribution which is quite similar to the German also resembles the clade distribution from Ghana. Last but not least, there are the quite distinct distribution from the US West and East Coast (California, CA and New York, NY). The Californian chart has similarity with the Nigerian because of the high percentage of Chinese-based clades while the chart from New York has a comparable high percentage of clade GH as the ones from France and Senegal.

Geographic distribution

The world map in Figure 5 reveals the distinct combinations of introduction of China-and Europe-based clades in West African countries. Nigeria has the highest percentage of the early clades (L,S,V) which were based in China but subsequently distributed to Europe and to the West Coast of USA. Ghana possesses nearly equally distributed percentages of the early clades and the Europe-based clades (G,GH,GR) comparable to the West Coast of USA. Senegal has a similar clade distribution like France and only few samples from the early China-based clades might be more comparable to the US East Coast.

We set out to further explore the above-mentioned surprising observation (Figure 3) that in West Africa the early clades emerged after the later Europe-associated G-clades. Possible explanations could be (i) latent circulation of the early clades in West Africa or (ii) later introduction of the earlier clades. With the aim to find evidence for one of these alternatives, we looked into detail of the

phylogeny of samples from the earlier clades. We picked two samples from the early clades: sample Senegal/136/2020 comes from a phylogenetic branch predominated by Spanish samples but also including samples from Asia and Latin America (suppl. Figure 2), several West African samples from Nigeria (dated March 29th, 2020), Ghana and Senegal in the phylogenetic branch in suppl. Figure 3 have a long latency time of about 2 months to the estimated common predecessor estimated on January 29th, 2020. Thus, there is evidence for a combination of both explanations : SARS-CoV-2 samples of the early clades may have circulated latently in West Africa since January 2020 but additionally there might have been introductions of the early clades from Europe and Asia or via maritime trade.

Discussion

In this phylogenetic analysis of SARS-CoV-2 sequences from the West African countries Gambia, Ghana, Nigeria and Senegal we identified country-specific patterns of earlier (L, S, V) and later Europe-associated (G,GR, GH) clades. In Senegal and Gambia, the later Europe-associated clades were predominant, in Ghana earlier and later clades were more equally distributed and in Nigeria the earlier clades were the predominant samples downloaded from the GISAID database in June 2020. This would suggest multiple introductions mainly from Europe into Senegal and Gambia, from Europe and directly or indirectly via other Asian or European countries from China then to Ghana and Nigeria. The introductions from China to Nigeria and Ghana are in line with a study by Haider *et al.* in which both countries, but not Senegal and Gambia, appear in a table of estimations of SARS-CoV-2 transmission risk from China based on air traffic statistics (10). However, they are at low risk in the second quartile - with the fourth quartile having the highest risk. There was a lack of data for Senegal and Gambia therefore hinting to no or only low-level air traffic connection to China, thus suggesting a predominant introduction from Europe.

Against our expectations, we found that the later European-associated clades (G, GR, GH) emerged before the earlier Chinese-based clades (L, S, V) in the registered cases in the investigated West African countries. We propose the following hypothesis as an explanation to this surprising observation: the early clades were already circulating within the populations before the later European-associated clades were introduced. A higher disease severity of the later European clades might then be a possible explanation for their earlier detection. Intriguingly, most of the cases investigated in this study occurred within the time interval of the Wuhan lockdown between January 23rd and April 8th 2020. Thus, transmission of the early clades must have taken place very early or via intermediate countries or other Chinese provinces. Besides the later Europe-associated G-clades, the early clades were also circulating in Europe and the US West coast of USA, for example, the Senegal sample no. 136 from the early clade S has similarity with Spanish samples (suppl. Figure 2). Other explanations for the relatively long latency may be founder effects that by chance individuals infected with the later clades travelled to West Africa before individuals infected with the earlier clades – or slower means of transportation such as ships commuting between China, America, Europe and West Africa.

Based on previous reports (1), it might probably be that the later G clades will replace the early clades in Nigeria and Ghana. The question if that correlates with the severity of the disease still needs to be addressed, Brufsky infers it from the higher mortality at the East Coast of USA with predominantly D614G-carrying G-clades compared to the West Coast with the predominant early clades (2). Becerra-Flores et al. found significant correlations between the percentage of D614G and case-fatality on a country by country basis (24). However, others find evidence for higher transmissibility and also higher viral-load but no evidence for higher disease severity (1), (25), (26). A correlation of the mutation D614G associated with the G-clades and case fatality in the West African countries can only be identified at a marginal level of r=0.28 (Supplementary Table 1). The case fatality is fortunately rather low ranging from 0.6 in Ghana up to 3.2 in Gambia. Other factors such as climate, sunlight exposure (27) and associated Vitamin D (28), medical infrastructure and

demographics might influence the etiopathology even more. There are also perspectives of decreased disease severity as Benedetti *et al.* argue that SARS-CoV-2 will mutate continuously and attenuate naturally to become endemic at a low mortality rate (29), as has been observed with earlier viruses (30).

The limitations of this study are the sample size, possible selection bias of the samples and the intrinsic incompleteness of the phylogenetic analysis which may lead to altered results when more samples are included. Nonetheless, this is the first study of its kind, the data and concept should form the basis for a more extensive analysis due to an increased number of sequenced samples.

In conclusion, in this phylogenetic analysis of SARS-CoV-2, we found distinct patterns of viral clades: the later Europe-associated G-clades are predominant in Senegal and Gambia, and combinations of the earlier (L, S, V) and later clades in Ghana and Nigeria. Intriguingly, the later clades emerged before the earlier clades which could simply be due to founder effects or due to latent circulation of the earlier clades. Only a marginal correlation of the G-clades in the West African countries can be associated with mortality which fortunately is at a rather low level therefore disproving fears that the pandemic would massively overwhelm the health systems in Africa. The rather young population and the climate might be factors favoring this low fatality rate in comparison to Western countries but nevertheless a cautious balance between health protection and economics might prevent future disastrous outbreaks.

Author Contributions

JA and WW came up with the concept of the study. WW and JA wrote the manuscript. WW analysed and WW and JA interpreted the data. JA supervised the work.

Funding

James Adjaye acknowledges financial support from the Medical Faculty, Heinrich-Heine-University-Düsseldorf, Germany.

Acknowledgments

We gratefully acknowledge all authors from the originating laboratories and from the submitting laboratories of the sequences from GISAID and GenBank which are listed in Supplementary Table 2. James Adjaye acknowledges financial support from the medical faculty of Heinrich Heine University-Düsseldorf, Germany.

Conflicts of Interest

The authors declare no conflict of interest.

Tables

Table 1: SARS-CoV-2samples used for the phylogenetic analysis

region	name	accession	date	clade
West Africa	hCoV-19/Gambia/GC19-029/2020	EPI_ISL_428857	20.04.2020	GH
	hCoV-19/Senegal/611/2020	EPI_ISL_420076	20.03.2020	GH
	hCoV-19/Senegal/003/2020	EPI_ISL_418206	28.02.2020	GH
	hCoV-19/Senegal/016/2020	EPI_ISL_418207	02.03.2020	GH
	hCoV-19/Senegal/026/2020	EPI_ISL_418209	03.03.2020	GH
	hCoV-19/Senegal/020/2020	EPI_ISL_418208	04.03.2020	GH
	hCoV-19/Senegal/382/2020	EPI_ISL_420073	19.03.2020	GH
	hCoV-19/Senegal/370/2020	EPI_ISL_420072	18.03.2020	GH
	hCoV-19/Nigeria/OY008-CV29/2020	EPI_ISL_455429	27.03.2020	GH
	hCoV-19/Nigeria/KW017-CV24/2020	EPI_ISL_455362	10.04.2020	GH
	hCoV-19/Ghana/2333_S5/2020	EPI_ISL_422394	27.03.2020	GH
	hCoV-19/Ghana/2986_S10/2020	EPI_ISL_422401	31.03.2020	GH
	hCoV-19/Ghana/1622_S2/2020	EPI_ISL_422384	24.03.2020	GH
	hCoV-19/Nigeria/OS060-CV9/2020	EPI_ISL_455419	29.03.2020	GH
	hCoV-19/Senegal/610/2020	EPI_ISL_420075	20.03.2020	GH
	hCoV-19/Senegal/328/2020	EPI_ISL_420071	17.03.2020	GH
	hCoV-19/Senegal/640/2020	EPI_ISL_420079	20.03.2020	GH
	hCoV-19/Gambia/GC19-015/2020	EPI_ISL_428855	17.03.2020	GR
	hCoV-19/Ghana/2914_S8/2020	EPI_ISL_422399	30.03.2020	GR
	hCoV-19/Nigeria/NG57752/2020	EPI_ISL_462992	2020-03	GR
	hCoV-19/Ghana/1565_S13/2020	EPI_ISL_422404	24.03.2020	GR
	hCoV-19/Nigeria/Lagos01/2020	EPI_ISL_413550	27.02.2020	GR
	hCoV-19/Senegal/315/2020	EPI_ISL_420070	17.03.2020	G
	hCoV-19/Senegal/600/2020	EPI_ISL_420074	20.03.2020	G
	hCoV-19/Senegal/073/2020	EPI_ISL_418210	10.03.2020	G
	hCoV-19/Senegal/082/2020	EPI_ISL_418211	11.03.2020	G
	hCoV-19/Senegal/087/2020	EPI_ISL_418212	11.03.2020	G
	hCoV-19/Senegal/094/2020	EPI_ISL_418213	12.03.2020	G
	hCoV-19/Senegal/119/2020	EPI_ISL_418215	12.03.2020	G
	hCoV-19/Ghana/2853_S7/2020	EPI_ISL_422398	29.03.2020	G
	hCoV-19/Senegal/139/2020	EPI_ISL_418217	13.03.2020	G
	hCoV-19/Ghana/2944_S9/2020	EPI_ISL_422400	30.03.2020	G
	hCoV-19/Ghana/1659_S14/2020	EPI_ISL_422405	25.03.2020	G
	hCoV-19/Senegal/102/2020	EPI_ISL_418214	12.03.2020	G
	hCoV-19/Senegal/618/2020	EPI_ISL_420077	20.03.2020	S
	hCoV-19/Senegal/136/2020	EPI_ISL_418216	13.03.2020	S
	hCoV-19/Senegal/620/2020	EPI_ISL_420078	20.03.2020	S
	hCoV-19/Ghana/1651_S3/2020	EPI_ISL_422387	25.03.2020	S
	hCoV-19/Ghana/2850_S15/2020	EPI_ISL_422406	29.03.2020	S
	hCoV-19/Ghana/2828_S6/2020	EPI_ISL_422397	29.03.2020	S
	hCoV-19/Ghana/3177_S12/2020	EPI_ISL_422403	30.03.2020	S
	hCoV-19/Ghana/3176_S11/2020	EPI_ISL_422402	30.03.2020	S
	hCoV-19/Nigeria/OS030-CV5/2020	EPI_ISL_455415	29.03.2020	S
	hCoV-19/Nigeria/OS085-CV14/2020	EPI_ISL_455424	29.03.2020	S
	hCoV-19/Nigeria/OS055-CV8/2020	EPI_ISL_455418	29.03.2020	S
	hCoV-19/Nigeria/OS070-CV11/2020	EPI_ISL_455422	29.03.2020	S
	hCoV-19/Nigeria/OS075-CV12/2020	EPI_ISL_455423	29.03.2020	S
	hCoV-19/Nigeria/OS122-CV18/2020	EPI_ISL_455426	29.03.2020	S
	hCoV-19/Nigeria/OS125-CV20/2020	EPI_ISL_455427	29.03.2020	S
	hCoV-19/Gambia/GC19-026/2020	EPI_ISL_428856	21.03.2020	V
	hCoV-19/Nigeria/OG007-CV22/2020	EPI_ISL_455412	29.03.2020	V
	hCoV-19/Nigeria/OY045A-CV35/2020	EPI_ISL_455431	02.04.2020	V
	hCoV-19/Nigeria/OS016-CV3/2020	EPI_ISL_455413	27.03.2020	V
	hCoV-19/Senegal/306/2020	EPI_ISL_420069	17.03.2020	L
	hCoV-19/Ghana/2230_S4/2020	EPI_ISL_422390	25.03.2020	L
	hCoV-19/Nigeria/OY035-CV34/2020	EPI_ISL_455430	02.04.2020	L
	hCoV-19/Ghana/1513_S1/2020	EPI_ISL_422382	24.03.2020	L
	hCoV-19/Nigeria/OS029-CV4/2020	EPI_ISL_455414	29.03.2020	L
	hCoV-19/Nigeria/OS116-CV17/2020	EPI_ISL_455425	29.03.2020	L
	hCoV-19/Nigeria/NG02/2020	EPI_ISL_450507	09.04.2020	L
others	hCoV-19/France/HF1463/2020	EPI_ISL_429968	21.02.2020	GH
	hCoV-19/England/SHEF-BFD27/2020	EPI_ISL_416737	03.03.2020	GR
	hCoV-19/Brazil/SPBR-13/2020	EPI_ISL_416035	05.03.2020	GR
	hCoV-19/Austria/CeMM0045/2020	EPI_ISL_437932	24.02.2020	GR
	hCoV-19/Netherlands/Berlicum_1363564/2020	EPI_ISL_413565	24.02.2020	GR
	hCoV-19/Italy/CDG1/2020	EPI_ISL_412973	20.02.2020	G
	hCoV-19/Germany/BavPat1/2020	EPI_ISL_406862	28.01.2020	G
	hCoV-19/Italy/PV-5314-N/2020	EPI_ISL_451307	21.02.2020	G
	hCoV-19/Luxembourg/LNS0000001/2020	EPI_ISL_419562	29.02.2020	G
	hCoV-19/DRC/3632/2020	EPI_ISL_447607	23.04.2020	G
	hCoV-19/HongKong/HKU-908a/2020	EPI_ISL_434569	27.01.2020	V
	hCoV-19/DRC/3632/2020	EPI_ISL_447607	23.04.2020	V
	hCoV-19/Wuhan-Hu-1/2019	EPI_ISL_402125	31.12.2019	L
	hCoV-19/USA/CA2/2020	EPI_ISL_406036	22.01.2020	L
	hCoV-19/Wuhan/WH01/2019	EPI_ISL_406798	26.12.2019	L

Figure legends

Fig. 1: Phylogenetic tree revealing similarities of West-African viral sequences with Chinese and multiple European countries. (a) Nigerian samples cluster with the early Chinese samples within the bottom branch of the tree, Ghanaian samples are about equally distributed between the top (European) and bottom branch of the tree. Senegalese samples cluster closer with the French reference sample on the top of the tree. (b) Highest diversity is at the A23403G (D614G) mutation splitting the tree in the bottom (Chinese) and top (European) branch. This mutation was reported to increase infectivity.

Fig. 2: Phylogenetic tree colored by clades shows distribtution of West-African samples over all clades suggesting introductions from China and European countries. Patterns are country.specific, e.g. most Senegalese samples have close similarity with the French reference, most Nigerian samples cluster in early Chinese-based clade S and Ghanaian samples are spread over all clades. Within the clade S, there are putatively specific West-African mutations at the branches at C24370T and G22468T. G22486T may reflect migration routes because in the nextstrain analysis of whole Africa there are also Tunisian samples in this branch (https://nextstrain.org/ncov/africa?f_region=Africa, accessed Jun 26th, 2020). Two of the non-French related Senegalese samples come from these branches while the other (Senegal/136) has strong similarity with Spanish-end of February samples from the early clade S pointing at multiple introductions to Senegal from France, Spain and African countries.

Fig. 3: Temporal course of clade distribution confirms gaining of share of the Europe-associated G-clades harboring the putatively more infective D614G mutation. Interestingly the younger Europe-associated G-clades emerged earlier in West African sequenced samples. This could be due to founder effects by introductions from France closely connected to Senegal and displaying a similar clade distribution. Furthermore, the China-based L-,V- and S-clade samples start in mid-March a time when the epidemic in China was nearly totally suppressed. Thus, the virus may have circulated in several countries before the first samples were sequenced. Surprisingly, the abundance of the S-

clade is relatively high mainly due to Nigeria and Ghana but without that exception the clade distribution resembles the global one with a delay of about 2-4 weeks.

Fig. 4: West African countries display distinct patterns of China-and Europe-based clades. Nigeria has the highest percentage of the China-based early clades (L,S,V) and Ghana has nearly equally distributed percentages of China and Europe-based clades (G,GH,GR). Senegal has a similar clade distribution as France but also a few samples from the early China-based clades. In Gambia there were only three sequences, two from Europe-based clades GR and GH and one from China-based clade V.

Fig. 5: Geographic map reveals distinct patterns of introduction of China-and Europe-based clades in West African countries. Nigeria with the highest percentage of the China-based early clades (L,S,V) and Ghana with nearly equally distributed percentages of China and Europe-based clades (G,GH,GR) might be comparable with the US West Coast while Senegal with a similar clade distribution like France and few samples from the early China-based clades may be more comparable to the US East Coast. It will be interesting to observe if the later G clades replace the early clades in Nigeria and Ghana and if that correlates with the severity of the disease as was postulated for the US.

Supplementary Material

Supplementary Table 1: Case fatality and percentage of mutation D614G in West African countries

Supplementary Table 2: Acknowledgement table of sequence samples from the GISAID database

Supplementary Figure 1: Detailed phylogenetic analysis of the Senegal/618 and several Nigerian samples point at introduction through travel or migration routes via Tunisia, Egypt and Mali.

Supplementary Figure 2: Detailed phylogenetic analysis of the Senegal/136 sample suggests introduction from Spain.

Supplementary Figure 3: Detailed phylogenetic analysis of Nigerian, Ghanian and Senegalese samples points at long latent circulation of early clades of SARS-CoV-2 in these countries between end of January until end of March 2020.

References

1. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell [Internet]. 2020 Jul [cited 2020 Aug 3]; Available from: https://linkinghub.elsevier.com/retrieve/pii/S0092867420308205

2. Brufsky A. Distinct viral clades of SARS-CoV-2: Implications for modeling of viral spread. J Med Virol [Internet]. 2020 Jun 24 [cited 2020 Aug 3]; Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.25902

3. Krueger DK, Kelly SM, Lewicki DN, Ruffolo R, Gallagher TM. Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. J Virol. 2001 Mar;75(6):2792–802.

4. Geoghegan JL, Holmes EC. The phylogenomics of evolving virus virulence. Nat Rev Genet. 2018 Dec;19(12):756–69.

5. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J Transl Med

[Internet]. 2020 Dec [cited 2020 Aug 3];18(1). Available from: https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-020-02344-6

6. Gámbaro F, Behillil S, Baidaliuk A, Donati F, Albert M, Alexandru A, et al. Introductions and early spread of SARS-CoV-2 in France, 24 January to 23 March 2020. Eurosurveillance [Internet]. 2020 Jul 2 [cited 2020 Aug 3];25(26). Available from: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.26.2001200

7. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary H, Obla A, et al. Introductions and early spread of SARS-CoV-2 in the New York City area. Science. 2020 May 29;eabc1917.

8. Bhattacharjee B, Pandit B. Phylogenetic clustering of the Indian SARS-CoV-2 genomes reveals the presence of distinct clades of viral haplotypes among states [Internet]. Genomics; 2020 May [cited 2020 Aug 3]. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.05.28.122143

9. Quaresima V, Naldini MM, Cirillo DM. The prospects for the SARS -CoV-2 pandemic in Africa. EMBO Mol Med [Internet]. 2020 Jun 8 [cited 2020 Aug 5];12(6). Available from: https://onlinelibrary.wiley.com/doi/abs/10.15252/emmm.202012488

10. Haider N, Yavlinsky A, Simons D, Osman AY, Ntoumi F, Zumla A, et al. Passengers' destinations from China: low risk of Novel Coronavirus (2019-nCoV) transmission into Africa and South America. Epidemiol Infect [Internet]. 2020 [cited 2020 Aug 5];148. Available from: https://www.cambridge.org/core/product/identifier/S0950268820000424/type/journal_article

11. Cabore JW, Karamagi HC, Kipruto H, Asamani JA, Droti B, Seydi ABW, et al. The potential effects of widespread community transmission of SARS-CoV-2 infection in the World Health Organization African Region: a predictive model. BMJ Glob Health. 2020 May;5(5):e002647.

12. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018 Dec 1;34(23):4121–3.

13. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002 Jul 15;30(14):3059–66.

14. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015 Jan;32(1):268–74.

15. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol [Internet]. 2018 Jan 1 [cited 2020 Jun 26];4(1). Available from: https://academic.oup.com/ve/article/4/1/vex042/4794731

16. South A. rworldmap: A New R package for Mapping Global Data. R J. 2011 Jun;3(1):35–43.

17. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First Case of 2019 Novel Coronavirus in the United States. N Engl J Med. 2020 05;382(10):929–36.

18. Lescure F-X, Bouadma L, Nguyen D, Parisey M, Wicky P-H, Behillil S, et al. Clinical and virological data of the first cases of COVID-19 in Europe: a case series. Lancet Infect Dis. 2020 Jun 1;20(6):697–706.

19. Böhmer MM, Buchholz U, Corman VM, Hoch M, Katz K, Marosevic DV, et al. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. Lancet Infect Dis [Internet]. 2020 May 15 [cited 2020 Jun 19]; Available from: http://www.sciencedirect.com/science/article/pii/S1473309920303145

20. Carinci F. Covid-19: preparedness, decentralisation, and the hunt for patient zero. BMJ [Internet]. 2020 Feb 28 [cited 2020 Jun 19];368. Available from: https://www.bmj.com/content/368/bmj.m799

21. Kluytmans-van den Bergh MFQ, Buiting AGM, Pas SD, Bentvelsen RG, van den Bijllaardt W, van Oudheusden AJG, et al. Prevalence and Clinical Presentation of Health Care Workers With Symptoms of Coronavirus Disease 2019 in 2 Dutch Hospitals During an Early Phase of the Pandemic. JAMA Netw Open. 2020 May 21;3(5):e209673.

22. Lillie PJ, Samson A, Li A, Adams K, Capstick R, Barlow GD, et al. Novel coronavirus disease (Covid-19): The first two patients in the UK with person to person transmission. J Infect. 2020 May 1;80(5):578–606.

23. Awoyelu EH, Oladipo EK, Adetuyi BO, Senbadejo TY, Oyawoye OM, Oloke JK. Phyloevolutionary analysis of SARS-CoV-2 in Nigeria. New Microbes New Infect. 2020 Jul;36:100717.

24. Becerra-Flores M, Cardozo T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. Int J Clin Pract. 2020 May 6;e13525.

25. Grubaugh ND, Hanage WP, Rasmussen AL. Making Sense of Mutation: What D614G Means for the COVID-19 Pandemic Remains Unclear. Cell. 2020 Aug;182(4):794–5.

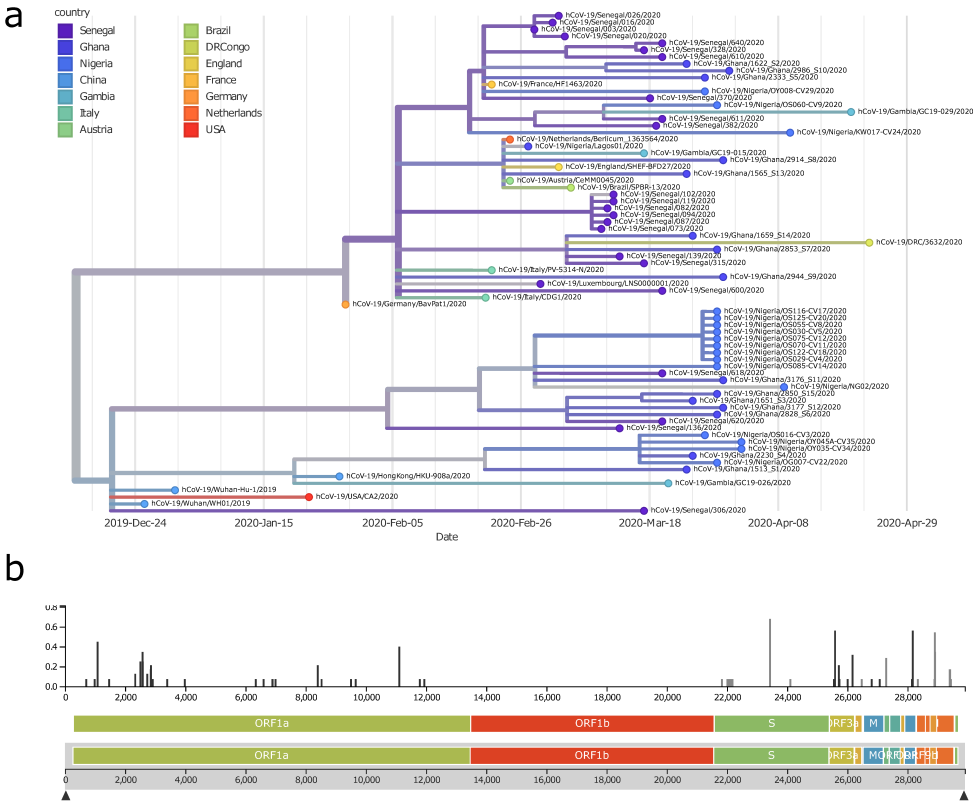
26. Lorenzo-Redondo R, Nam HH, Roberts SC, Simons LM, Jennings LJ, Qi C, et al. A Unique Clade of SARS-CoV-2 Viruses is Associated with Lower Viral Loads in Patient Upper Airways [Internet]. Infectious Diseases (except HIV/AIDS); 2020 May [cited 2020 Aug 24]. Available from: http://medrxiv.org/lookup/doi/10.1101/2020.05.19.20107144

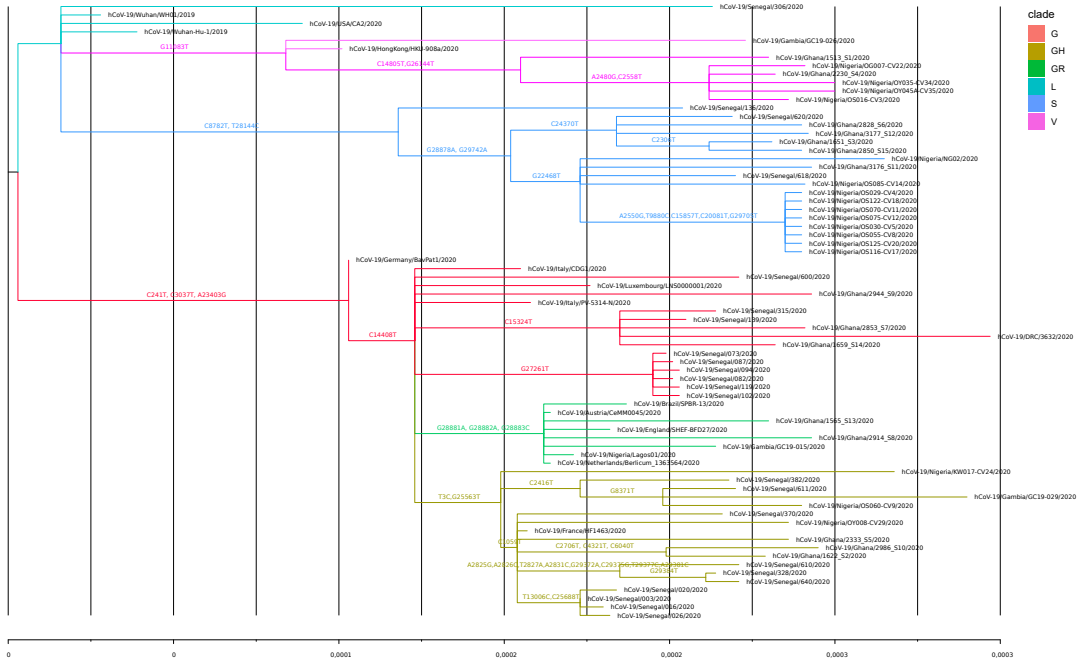
27. Whittemore PB. COVID-19 fatalities, latitude, sunlight, and vitamin D. Am J Infect Control [Internet]. 2020 Jun [cited 2020 Aug 6]; Available from: https://linkinghub.elsevier.com/retrieve/pii/S0196655320305745

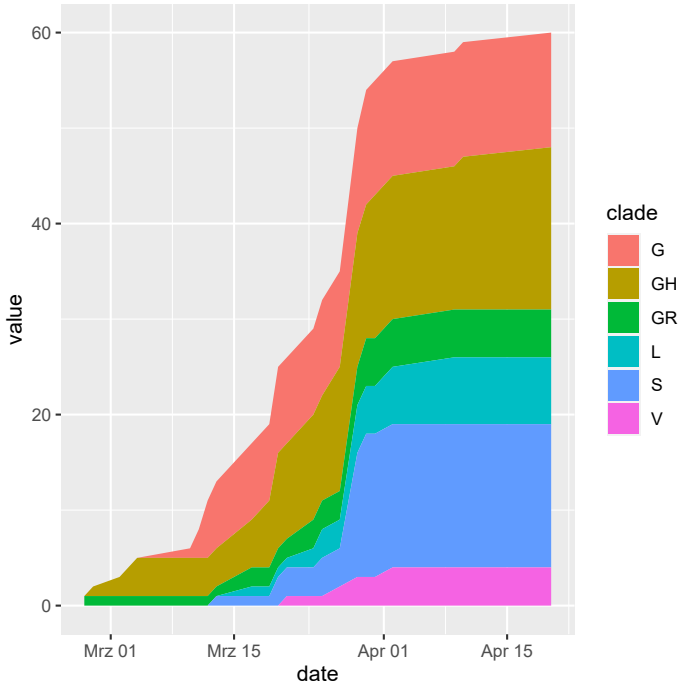
28. Ilie PC, Stefanescu S, Smith L. The role of vitamin D in the prevention of coronavirus disease 2019 infection and mortality. Aging Clin Exp Res. 2020 Jul;32(7):1195–8.

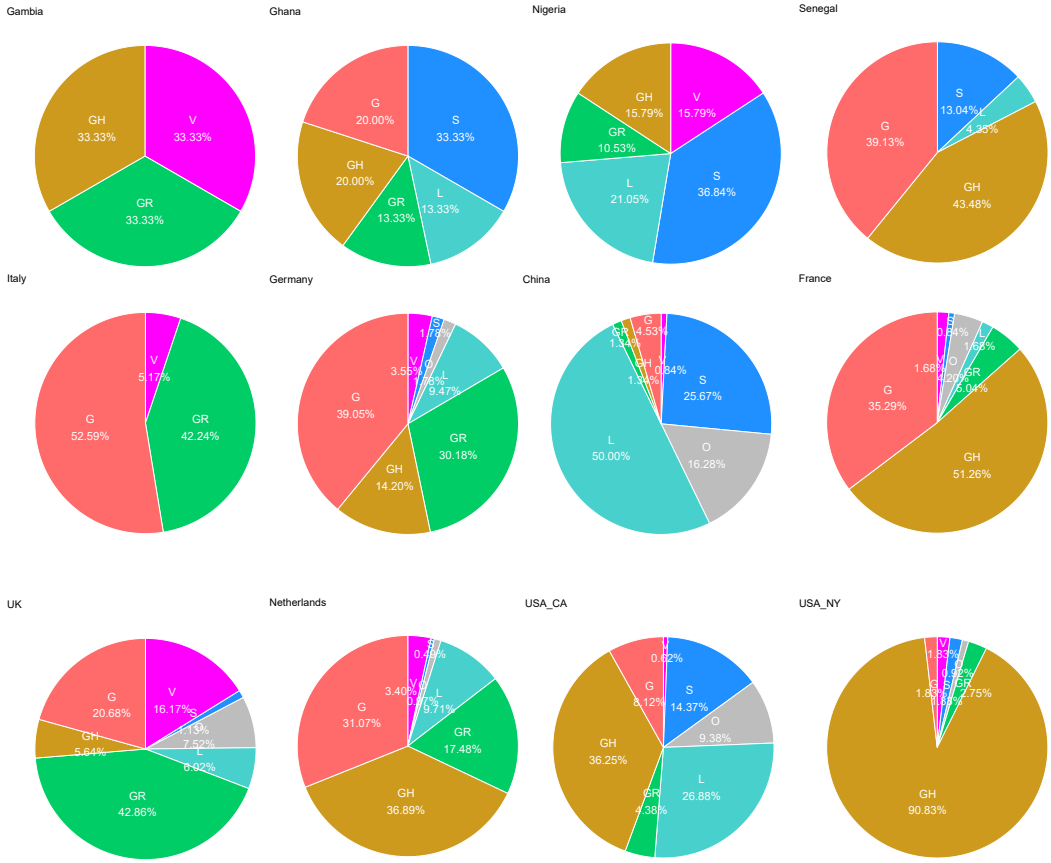
29. Benedetti F, Pachetti M, Marini B, Ippodrino R, Ciccozzi M, Zella D. SARS-CoV-2: March toward adaptation. J Med Virol [Internet]. 2020 Jul 11 [cited 2020 Aug 21]; Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.26233

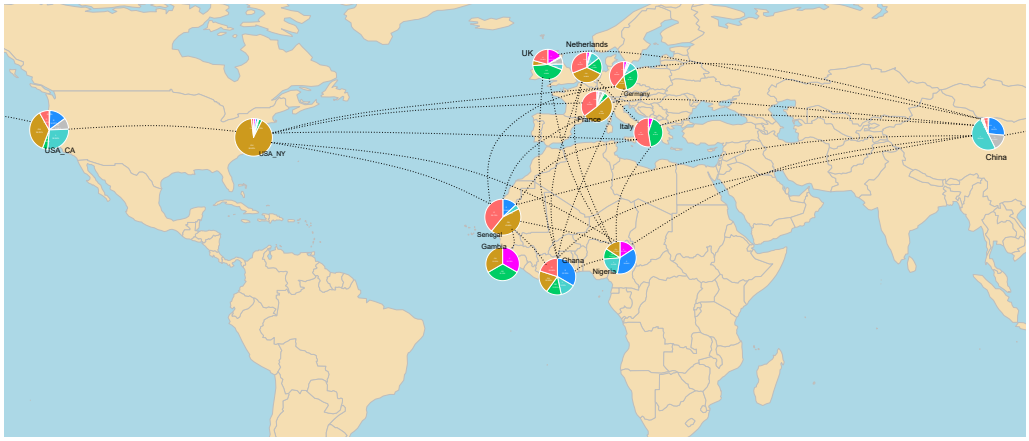
30. Armengaud J, Delaunay-Moisan A, Thuret J-Y, van Anken E, Acosta-Alvear D, Aragón T, et al. The importance of naturally attenuated SARS-CoV-2in the fight against COVID-19. Environ Microbiol. 2020;22(6):1997–2000.


a
country
Senegal
Ghana
Nigeria
China
Gambia
Italy
Austria
Brazil
DRCongo
England
France
Germany
Netherlands
USA
2019-Dec-24
2020-Jan-15
2020-Feb-05
2020-Feb-26
2020-Mar-18
2020-Apr-08
2020-Apr-29
Date
b
ORF1a
ORF1b
S
ORF3a
M
E


clade
G
GH
GR
L
S
V
hCoV-19/Wuhan/WH01/2019
hCoV-19/USA/CA2/2020
hCoV-19/Wuhan-Hu-1/2019
G11083T
hCoV-19/HongKong/HKU-908a/2020
C14805T,G26144T
hCoV-19/Senegal/306/2020
hCoV-19/Gambia/GC19-026/2020
hCoV-19/Ghana/1513_S1/2020
hCoV-19/Nigeria/OG007-CV22/2020
hCoV-19/Ghana/2230_S4/2020
A2480G,C2558T
hCoV-19/Nigeria/OY035-CV34/2020
hCoV-19/Nigeria/OY045A-CV35/2020
hCoV-19/Nigeria/OS016-CV3/2020
hCoV-19/Senegal/136/2020
hCoV-19/Senegal/620/2020
C8782T, T28144C
C24370T
hCoV-19/Ghana/2828_S6/2020
hCoV-19/Ghana/3177_S12/2020
C2304T
hCoV-19/Ghana/1451_S3/2020
hCoV-19/Ghana/2850_S15/2020
G28878A, G29742A
hCoV-19/Nigeria/NG02/2020
hCoV-19/Ghana/3176_S11/2020
G22468T
hCoV-19/Senegal/618/2020
hCoV-19/Nigeria/OS085-CV14/2020
hCoV-19/Nigeria/OS029-CV4/2020
hCoV-19/Nigeria/OS122-CV18/2020
hCoV-19/Nigeria/OS070-CV11/2020
A2550G,T9880C,C15857T,C20081T,G29703T
hCoV-19/Nigeria/OS075-CV12/2020
hCoV-19/Nigeria/OS038-CV5/2020
hCoV-19/Nigeria/OS055-CV8/2020
hCoV-19/Nigeria/OS125-CV20/2020
hCoV-19/Nigeria/OS118-CV17/2020
hCoV-19/Germany/BavPat1/2020
hCoV-19/Italy/CDG1/2020
hCoV-19/Senegal/600/2020
hCoV-19/Luxembourg/LNS0000001/2020
hCoV-19/Ghana/2944_S9/2020
C241T, C3037T, A23403G
hCoV-19/Italy/PV-S314-N/2020
hCoV-19/Senegal/315/2020
C15324T
hCoV-19/Senegal/139/2020
hCoV-19/Ghana/2853_S7/2020
C14408T
hCoV-19/DRC/3632/2020
hCoV-19/Ghana/1659_S14/2020
hCoV-19/Senegal/073/2020
hCoV-19/Senegal/087/2020
G27261T
hCoV-19/Senegal/094/2020
hCoV-19/Senegal/082/2020
hCoV-19/Senegal/110/2020
hCoV-19/Senegal/102/2020
hCoV-19/Brazil/SPBR-13/2020
hCoV-19/Austria/CeMM0045/2020
hCoV-19/Ghana/1565_S13/2020
G28881A, G28882A, G28883C
hCoV-19/England/SHEF-BFD27/2020
hCoV-19/Ghana/2914_S8/2020
hCoV-19/Gambia/GC19-015/2020
hCoV-19/Nigeria/Lagos01/2020
hCoV-19/Netherlands/Berlicum_136364/2020
hCoV-19/Nigeria/KW017-CV24/2020
hCoV-19/Senegal/382/2020
C2416T
hCoV-19/Senegal/611/2020
G8371T
T3C,G25563T
hCoV-19/Gambia/GC19-029/2020
hCoV-19/Nigeria/OS060-CV9/2020
hCoV-19/Senegal/370/2020
hCoV-19/Nigeria/OY008-CV29/2020
hCoV-19/France/IF1463/2020
hCoV-19/Ghana/2333_S5/2020
C1059T
C2706T, C4321T, C6040T
hCoV-19/Ghana/2986_S10/2020
hCoV-19/Ghana/1622_S2/2020
A2825G,A2826C,T2827A,A2831C,G29727A,C29736G,T29737G,G29801C
hCoV-19/Senegal/610/2020
G29384T
hCoV-19/Senegal/328/2020
hCoV-19/Senegal/640/2020
hCoV-19/Senegal/020/2020
T13006C,C25688T
hCoV-19/Senegal/003/2020
hCoV-19/Senegal/004/2020
hCoV-19/Senegal/026/2020
0
0
0,0001
0,0002
0,0002
0,0003
0,0003


60
40
20
0
value
Mrz 01
Mrz 15
Apr 01
Apr 15
date
clade
G
GH
GR
L
S
V


Gambia
GH 33.33%
V 33.33%
GR 33.33%
Ghana
G 20.00%
S 33.33%
GH 20.00%
GR 13.33%
L 13.33%
Nigeria
GH 15.79%
V 15.79%
GR 10.53%
L 21.05%
S 36.84%
Senegal
S 13.04%
L 4.35%
G 39.13%
GH 43.48%
Italy
V 5.17%
G 52.59%
GR 42.24%
Germany
G 39.05%
GR 30.18%
GH 14.20%
China
S 25.67%
L 50.00%
O 16.28%
France
G 35.29%
GH 51.26%
UK
G 20.68%
V 16.17%
GH 5.64%
GR 42.86%
L 6.92%
Netherlands
G 31.07%
GR 17.48%
GH 36.89%
L 9.71%
USA_CA
G 8.12%
S 14.37%
O 9.38%
GH 36.25%
GR 4.38%
L 26.88%
USA_NY
GH 90.83%


UK
Netherlands
Germany
France
Italy
USA_CA
USA_NY
China
Senegal
Gambia
Ghana
Nigeria

2.2 Publications as contributing author

2.2.1 Induced pluripotent stem cell-derived neuronal cells from a sporadic Alzheimer's disease donor as a model for investigating AD-associated gene regulatory networks.

BMC Genomics. 2015 Feb 14;16:84. doi: 10.1186/s12864-015-1262-5.

Hossini AM(1), Megges M(2)(3)(4), Prigione A(5)(6), Lichtner B(7), Toliat MR(8), Wruck W(9), Schröter F(10), Nuernberg P(11), Kroll H(12), Makrantonaki E(13)(14), Zouboulis CC, Adjaye J(16)(17).

Author information: (1)Departments of Dermatology, Venereology, Allergology and Immunology, Dessau Medical Center, 06847, Dessau, Germany. amir.hossini@klinikum-dessau.de. (2)Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University Duesseldorf, Moorenstr. 5, 40225, Duesseldorf, Germany. megges@molgen.mpg.de. (3)Molecular Embryology and Aging Group, Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, 14195, Berlin, Germany. megges@molgen.mpg.de. (4)Department of Biology, Chemistry and Pharmacy, Institute of Chemistry and Biochemistry, Freie Universität Berlin, Thielallee 63, 14195, Berlin, Germany. megges@molgen.mpg.de. (5)Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University Duesseldorf, Moorenstr. 5, 40225, Duesseldorf, Germany. Alessandro.Prigione@mdc-berlin.de. (6)Current address: Max Delbrueck Center for Molecular Medicine (MDC), Robert Roessle Str. 10, D-13125, Berlin, Germany. Alessandro.Prigione@mdc-berlin.de. (7)Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University Duesseldorf, Moorenstr. 5, 40225, Duesseldorf, Germany. lichtner@molgen.mpg.de. (8)Cologne Center for Genomics (CCG), Institute for Genetics, University of Cologne, 50931, Cologne, Germany. mohammad.toliat@uni-koeln.de. (9)Molecular Embryology and Aging Group, Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, 14195, Berlin, Germany. Wasco.Wruck@med.uni-duesseldorf.de. (10)Molecular Embryology and Aging Group, Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, 14195, Berlin, Germany. Friederike.Schroeter@med.uni-duesseldorf.de. (11)Cologne Center for Genomics (CCG), Institute for Genetics, University of Cologne, 50931, Cologne, Germany. peter.nuernberg@uni-koeln.de. (12)Institute for Transfusion Medicine Dessau, Red Cross Blood Transfusion Service NSTOB, 06847, Dessau, Germany. hartmut.kroll@bsd-nstob.de. (13)Departments of Dermatology, Venereology, Allergology and Immunology, Dessau Medical Center, 06847, Dessau, Germany. emakrantonaki@yahoo.com. (14)Geriatrics Research Group, Department of Geriatric Medicine, Charité Universitätsmedizin Berlin, Reinickendorfer Str. 61, 13447, Berlin, Germany. emakrantonaki@yahoo.com. (15)Departments of Dermatology, Venereology, Allergology and Immunology, Dessau Medical Center, 06847, Dessau, Germany. christos.zouboulis@klinikum-dessau.de. (16)Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University Duesseldorf, Moorenstr. 5, 40225, Duesseldorf, Germany. james.adjaye@med.uni-duesseldorf.de. (17)Molecular Embryology and Aging Group, Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, 14195, Berlin, Germany. james.adjaye@med.uni-duesseldorf.de.

Erratum in BMC Genomics. 2015;16:433. Zoubouliss, Christos C [corrected to Zouboulis, Christos C].

BACKGROUND: Alzheimer's disease (AD) is a complex, irreversible neurodegenerative disorder. At present there are neither reliable markers to diagnose AD at an early stage nor therapy. To investigate underlying disease mechanisms, induced pluripotent stem cells (iPSCs) allow the generation of patient-derived neuronal cells in a dish. RESULTS: In this study, employing iPS technology, we derived and characterized iPSCs from dermal fibroblasts of an 82-year-old female patient affected by sporadic AD. The AD-iPSCs were differentiated into neuronal cells, in order to generate disease-specific protein association networks modeling the molecular pathology on the transcriptome level of AD, to analyse the reflection of the disease phenotype in gene expression in AD-iPS neuronal cells, in particular in the ubiquitin-proteasome system (UPS), and to address expression of typical AD proteins. We detected the expression of p-tau and GSK3B, a physiological kinase of tau, in neuronal cells derived from AD-iPSCs. Treatment of neuronal cells differentiated from AD-iPSCs with an inhibitor of γ-secretase resulted in the down-regulation of p-tau. Transcriptome analysis of AD-iPS derived neuronal cells revealed significant changes in the expression of genes associated with AD and with the constitutive as well as the inducible subunits of the proteasome complex. The neuronal cells expressed numerous genes associated with sub-regions within the brain thus suggesting the usefulness of our in-vitro model. Moreover, an AD-related protein interaction network composed of APP and GSK3B among others could be generated using neuronal cells differentiated from two AD-iPS cell lines. CONCLUSIONS: Our study demonstrates how an iPSC-based model system could represent (i) a tool to study the underlying molecular basis of sporadic AD, (ii) a platform for drug screening and toxicology studies which might unveil novel therapeutic avenues for this debilitating neuronal disorder.

DOI: 10.1186/s12864-015-1262-5 PMCID: PMC4344782 PMID: 25765079 [Indexed for MEDLINE]

URL: https://link.springer.com/article/10.1186/s12864-015-1262-5

Contribution of Wasco Wruck: microarray data processing, generation of heatmaps, cluster analysis, allocation of AD related genes to brain regions, generation of protein association networks.

Hossini et al. BMC Genomics (2015) 16:84
DOI 10.1186/s12864-015-1262-5

BMC Genomics

RESEARCH ARTICLE Open Access

Induced pluripotent stem cell-derived neuronal cells from a sporadic Alzheimer's disease donor as a model for investigating AD-associated gene regulatory networks

Amir M Hossini[1†], Matthias Megges[2,5,6†], Alessandro Prigione[2,8], Bjoern Lichtner[2], Mohammad R Toliat[3], Wasco Wruck[5], Friederike Schröter[5], Peter Nuernberg[3], Hartmut Kroll[4], Eugenia Makrantonaki[1,7], Christos C Zouboulis[1] and James Adjaye[2,5*]

Abstract

Background: Alzheimer's disease (AD) is a complex, irreversible neurodegenerative disorder. At present there are neither reliable markers to diagnose AD at an early stage nor therapy. To investigate underlying disease mechanisms, induced pluripotent stem cells (iPSCs) allow the generation of patient-derived neuronal cells in a dish.

Results: In this study, employing iPS technology, we derived and characterized iPSCs from dermal fibroblasts of an 82-year-old female patient affected by sporadic AD. The AD-iPSCs were differentiated into neuronal cells, in order to generate disease-specific protein association networks modeling the molecular pathology on the transcriptome level of AD, to analyse the reflection of the disease phenotype in gene expression in AD-iPS neuronal cells, in particular in the ubiquitin-proteasome system (UPS), and to address expression of typical AD proteins.
We detected the expression of p-tau and GSK3B, a physiological kinase of tau, in neuronal cells derived from AD-iPSCs. Treatment of neuronal cells differentiated from AD-iPSCs with an inhibitor of γ-secretase resulted in the down-regulation of p-tau. Transcriptome analysis of AD-iPS derived neuronal cells revealed significant changes in the expression of genes associated with AD and with the constitutive as well as the inducible subunits of the proteasome complex. The neuronal cells expressed numerous genes associated with sub-regions within the brain thus suggesting the usefulness of our *in-vitro* model. Moreover, an AD-related protein interaction network composed of APP and GSK3B among others could be generated using neuronal cells differentiated from two AD-iPS cell lines.

Conclusions: Our study demonstrates how an iPSC-based model system could represent (i) a tool to study the underlying molecular basis of sporadic AD, (ii) a platform for drug screening and toxicology studies which might unveil novel therapeutic avenues for this debilitating neuronal disorder.

Keywords: Alzheimer's disease, Skin cells, γ-secretase, Induced pluripotent stem cells, *TNFRSF1A*, Neuronal cells, p-tau, Transcriptome, Proteasome

* Correspondence: james.adjaye@med.uni-duesseldorf.de
†Equal contributors
[2]Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University Duesseldorf, Moorenstr. 5, 40225 Duesseldorf, Germany
[5]Molecular Embryology and Aging Group, Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
Full list of author information is available at the end of the article

Background

Alzheimer's disease (AD) is characterized by histopathological changes, designated as senile plaques and fibrillary deposits, which ultimately lead to the death of neuronal cells in particular in the cerebral cortex of the brain [1,2]. The familial form of AD is rare, affecting less than five percent of AD patients and has been associated with mutations of *Presenilin 1* (*PSEN1*), *Presenilin 2* (*PSEN2*) and *Amyloid Precursor Protein* (*APP*) [3,4]. These mutations result in incorrect cleavage of the protein, producing a deposited protein of amyloid-β (Aβ) that is more likely to form plaques [1,5]. Little is known about the molecular basis of multifactorial sporadic AD. In post-mortem examination of patients with AD, massive accumulation of two types of amyloid fibril senile plaques (Aβ40, Aβ42) and hyperphosphorylated tau forming paired helical filaments could be detected [6,7]. Both types of amyloid fibrils are mainly created enzymatically by β- and γ-secretase activity from the APP [8].

The most widely accepted theory for the onset of sporadic AD is the accumulation of extracellular Aβ42 in an aggregated state in the brain, subsequently leading to the formation of neurofibrillary tangles (NFT) containing hyperphosphorylated tau proteins and consequently to its inactivation, thus leading to inhibition of binding to the spindle apparatus and hence disrupted axonal transport [9,10]. The major modification of tau is its phosphorylation. Its hyperphosphorylation has been shown to be the critical step in the formation of NFTs [11,12]. One of the kinases that phosphorylates tau *in-vivo* is glycogen synthase kinase-3β (GSK3B), which is widely expressed in all tissues with elevated expression in developing brains [13]. Unlike many other kinases, GSK3B is believed to be permanently active in resting cells and in neurons without extracellular stimulation and can be inactivated by Ser9 phosphorylation [14].

Moreover, the ubiquitin-proteasome system (UPS) has been shown to be involved in the pathogenesis of AD [15-18].

The UPS consists of the 26S proteasome and the small protein ubiquitin, a post-translational modification, and is operative in all eukaryotes for intracellular protein homeostasis and quality [19,20]. The alternative form of the constitutive proteasome is the immunoproteasome complex [21]. It was demonstrated in *in-vitro* experiments that the accumulation of Aβ peptide in *APP*/*PSEN1* mutant neuronal cell culture leads to the inhibition of the proteasome as well as the de-ubiquitinating enzymes (DUBs) [15]. Despite increasing knowledge on AD-associated pathology, the molecular mechanisms underlying the cause of sporadic and familial AD are still not completely understood. This limitation is primarily due to limited access and availability of viable neuronal cells from AD patients because of ethical and practical reasons. Human induced pluripotent stem (iPSCs) cells enables the generation of clinically relevant neuronal cells *in-vitro*. Patient-derived skin cells are an easily accessible source for reprogramming and the obtained neuronal cells can be used to investigate the pathogenesis of neuronal disorders, including Parkinson disease and Alzheimer's disease [22-26].

In our study, we reprogrammed dermal fibroblasts obtained from an 82-year-old female patient diagnosed with late-stage AD to iPSCs in one reprogramming experiment. The two derived iPSC lines showed pluripotency-associated properties similar to human embryonic stem cells (hESC) and they could be successfully differentiated into neuronal cells *in-vitro*. The differentiated neuronal cells seemed to reflect the sporadic AD phenotype in the brain of the patient, including the expression of p-tau proteins, the up-regulation of GSK3B protein and its phosphorylation in contrast to the parental dermal fibroblast cells. In addition, numerous AD-related genes were found to be down-regulated, as revealed by microarray-based gene expression analysis of one neuronal differentiation experiment per AD-iPS cell line. Most notably these genes could be allocated to brain regions affected by Alzheimer's disease.

We demonstrated a down-regulation of p-tau proteins in AD neuronal cells with an inhibitor of γ-secretase. In addition, neuronal cells differentiated from the patient iPSCs showed an up-regulation of a number of neuronal and biological processes, which include development of the nervous system, neurogenesis, WNT signaling pathway, the lysosome, glutathion metabolism as well as the alanine, aspartate and glutamate metabolism. Furthermore, the down-regulation of AD-related genes enabled us to successfully construct a protein association network using the STRING database reflecting the presence of AD-related disease mechanisms in our iPSCs model. Finally, we could show that gene regulation of the constitutive as well as of the inducible subunits of the proteasome complex is affected in iPSC-derived neurons from the AD patient compared to the healthy subject. Further investigations are needed to better understand the molecular basis of the onset and progression of Alzheimer's disease. Elucidating the molecular mechanism of sporadic AD by modeling it via iPSC technology and protein association networks could provide valuable information needed to uncover appropriate strategies against the early onset of the disease.

Results

Generation and characterization of sporadic AD-iPSCs

Dermal fibroblasts were isolated from an 82-year-old woman diagnosed with final stage AD. The cell line was named NFH-46, and lack of AD-related mutations, such

as *APP*, *PSEN1* and *PSEN2* [1,5], was confirmed by direct sequencing analysis (Additional file 1). HLA haplotype analysis in the AD donor did not reveal any association of HLA alleles to Morbus Alzheimer. The HLA-alleles HLA-A*01:01,*03:01; B*08,*35, C*04:01,*07:01, DRB1*03:01, *11:01 were found in NFH-46. However, the Alzheimer-related HLA-alleles HLA-A*02, HLA-B*07 and HLA-C*07:02 could not be detected.

AD-iPSCs were generated by retroviral transduction using the classical Yamanaka cocktail [27], which includes the four transcription factors OCT4, KLF4, SOX2, and c-MYC, as demonstrated previously [28]. In a single reprogramming experiment several colonies exhibiting hESC-like morphologies were identified and manually picked for expansion and characterization. Two iPSC lines, AD-iPS5 and AD-iPS26B, were successfully established from this reprogramming experiment and characterized with respect to pluripotency-associated properties. Both lines exhibited hESC-like morphologies (Figure 1), telomerase activity (Additional file 2), alkaline phosphatase (AP) activity (Additional file 3a), expression of pluripotency-associated markers NANOG, SSEA4, TRA-1-60, and TRA-1-81 (Figure 2), expression of pluripotency-associated genes such as *NANOG*, *POU5F1*, *SOX2*, *LIN28*, *TDGF1*, *DPPA4*, *FGF4*, *GDF3*, *LEFTY1*, *LEFTY2* (Additional file 4) and the genetic fingerprinting pattern of the parental NFH-46 fibroblasts (Additional file 3b).

Finally, the transcriptomes of the AD-iPSC lines are similar to hESCs (H1 and H9) and to iPS lines previously generated from control NFH-2 fibroblasts [28] (Additional file 5).

The ability to differentiate into almost all tissue types as a hallmark of human pluripotent stem cells was analyzed employing embryoid bodies (EBs) based differentiation *in-vitro* and teratoma formation *in-vivo*. The AD-iPSC lines were able to differentiate *in-vitro* into all three embryonic germ layers, as detected by the expression of marker proteins specific for ectoderm (b-TUBULIN III and NESTIN), for mesoderm (Smooth Muscle Actin (SMA) and T/Brachyury), and endoderm (Alpha feto protein (AFP) and SOX17) (Additional file 6).

Finally, both AD-iPSC lines successfully generated teratomas (Additional file 7). For AD-iPS5, the presence of known endoderm-associated structures appeared unclear. However, this must not necessarily imply an impairment of this line towards endoderm differentiation *in-vivo*, since the teratoma assay itself is not standardized [29]. Moreover, in the *in-vitro* differentiated cells from AD-iPS5, SOX17 and AFP, both protein markers representative of endoderm, could be detected. Thus, we consider AD-iPS5 to be pluripotent.

Chromosomal analysis of AD-iPS5 revealed the loss of a gonosome, probably the X chromosome, because during mitosis they revealed a normal female karyotype. This is in agreement with our previous study showing

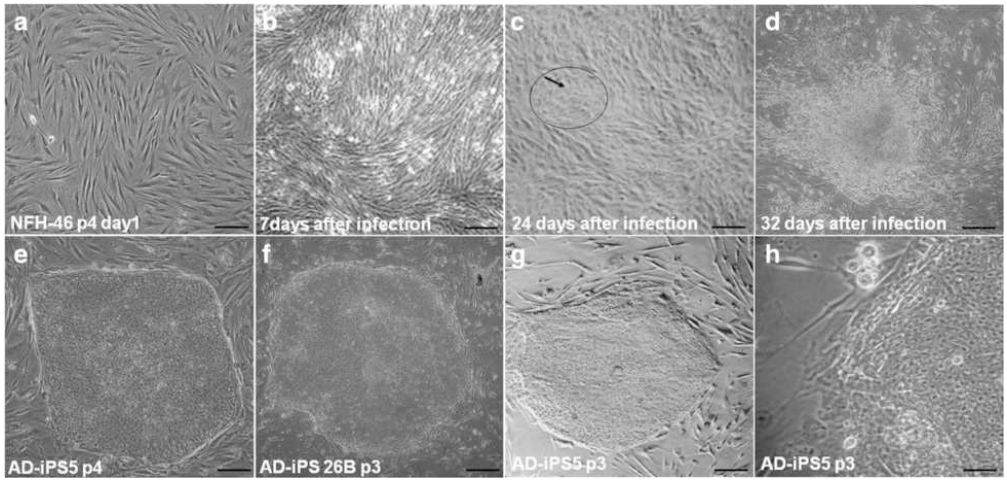


Figure 1 Generation of human iPSCs from skin fibroblasts of a sporadic Alzheimer patient. (a): Morphology of fibroblasts NFH-46 in passage 4 (p4) before viral transduction. **(b)**: Changes in morphology of NFH-46 seven days after infection with retroviruses. **(c)**: Changes of NFH-46 on day 24 after infection shown in circle with arrow. **(d)**: Typical image of non-embryonic stem cell like colony. **(e, f)**: Typical morphology of AD-iPS colonies (AD-iPS-5, passage 4; AD-iPS-26B, passage 3) of one reprogramming experiment. **(g)**: Typical morphology of AD-iPS colony in passage 3(p3). **(h)**: AD-iPSC structure in high magnification. Scale bar, 100 μm.

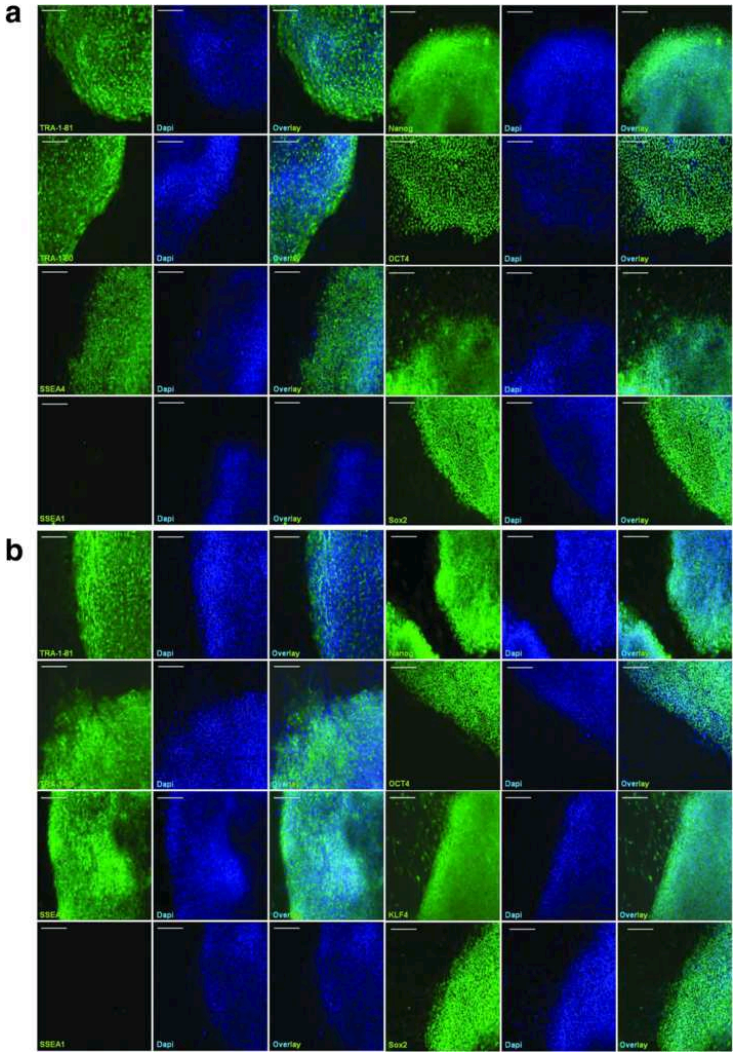


Figure 2 AD-iPSCs express key pluripotency-associated proteins. Two AD-iPSC lines were successfully generated with one reprogramming experiment: AD-iPS5 **(a)** and AD-iPS26B **(b)**. Both lines exhibited hESC-like morphologies, were positive for pluripotency-associated marker proteins, such as TRA-1-81, TRA-1-60, SSEA4, and NANOG, and were negative for the differentiation-specific marker SSEA1. Scale bar, 100 μm.

that iPSC lines generated from old donors are more likely to contain chromosomal aberrations [28]. In nine mitoses, a very small supernumerary marker chromosome (sSMC) was found besides the monosomy X (Additional file 8). It is unlikely that the presence of a small supernumaray maker chromosome has an effect on AD-iPS5 as sSMCs are a common phenomenon in human. The karyotypes of the second iPSC line AD-iPS26B and the parental cells NFH-46 were normal (Additional file 8).

Generation of neuronal cells from AD-iPSCs (AD-iPS neurons)

We derived neuronal cells from AD-iPSCs in one experiment to address the potential of these to reflect

neuropathological features found in neuronal cells of sporadic AD patients. As a control, neuronal cells were derived from the female hESC line H9 in one differentiation experiment. The neuronal cells were generated following a recently published protocol, which requires the exposure to TGF-β receptor (SB431542) and MEK1/2 (PD0325901) inhibitors [30]. AD-iPSC lines (AD-iPS5 and AD-iPS26B) and H9 were successfully differentiated into neuronal cells. The efficiency of differentiation varied, as AD-iPS5 showed more pronounced neuronal differentiation than AD-iPS26B. All induced neuronal cells were positive for neuronal cell markers PAX6, NESTIN, and b-TUBULIN III as shown in Figure 3. Most of the neuronal marker genes in the heatmap shown in Figure 4a are expressed in a similar manner in AD-iPSC neurons and H9 neurons, hence confirming a neuronal differentiation of comparable quality across all used pluripotent cell lines.

Expression of neuronal marker genes in AD-iPS neurons and H9 neurons

The heatmap in Figure 4a shows the expression pattern of pre-synaptic and post-synaptic genes as well as markers of distinct subtypes of neural progenitors and mature neuronal cell types in AD-iPS neurons and H9 neurons of one differentiation experiment conducted. Neuronal markers *FABP7*, *HES5*, *SOX2*, *PROM1* and *ASCL1* are expressed in AD-iPS5 and AD-iPS26B neurons, however, *FABP7* was not detected in H9 neurons. *GALC*, a marker of oligodendrocyte progenitor cells is expressed in all samples but lower in AD-iPS26B and H9 neurons. *MAP2*, a marker of neuronal dendrites is expressed in AD-iPS5 neurons but not in neuronal cells of AD-iPS26B and H9. Moreover, markers of retinal ganglion cells (*POU4F2*), dopaminergic neurons (*TH*) and glutamatergic neurons (*SLC17A6*) are expressed in AD-iPS5 neurons (Figure 4a). In addition, neural genes such

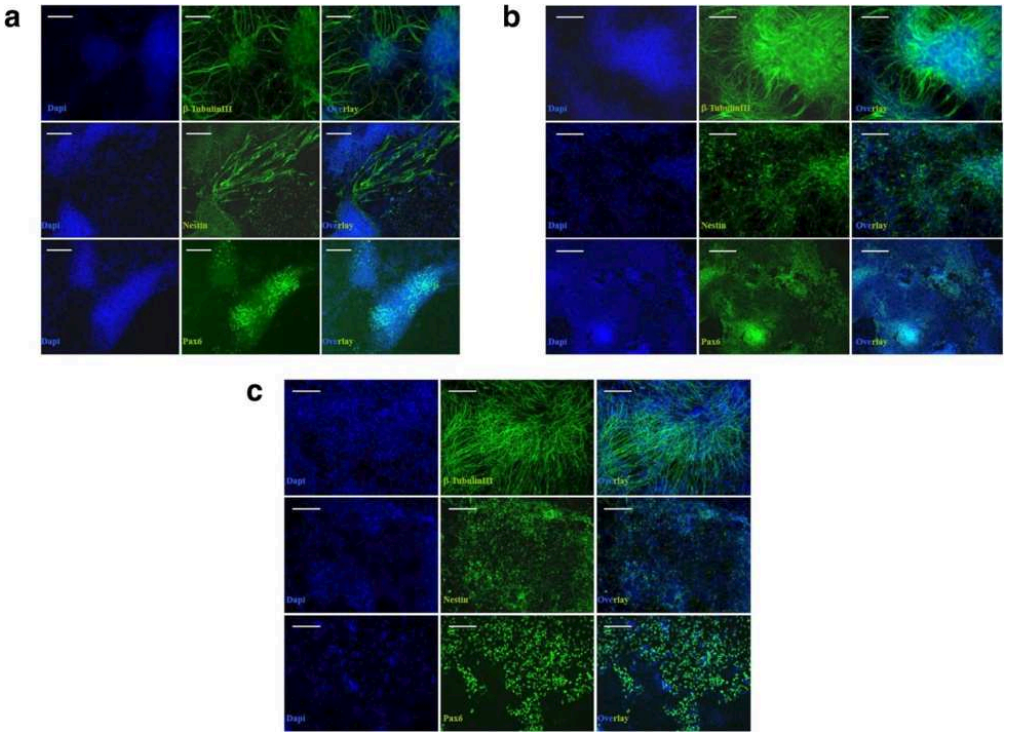


Figure 3 Neuronal differentiation of AD-iPSCs. Induction of neuronal cells by simultaneous treatment with inhibitors of transforming growth factor (TGF)-β receptor and MEK1/2 for 4 weeks. Neuronal cells derived from AD-iPSCs **(a)**: AD-iPS5 neurons and **(b)**: AD-iPS26B neurons showed the expression of neuronal markers, including PAX6, NESTIN and -β-TUBULIN III in a similar fashion as neuronal cells derived from the hESC line **(c)**: H9. The neuronal differentiation was carried out once per AD-iPS cell line or H9 cell line. Scale bar, 100 μm.

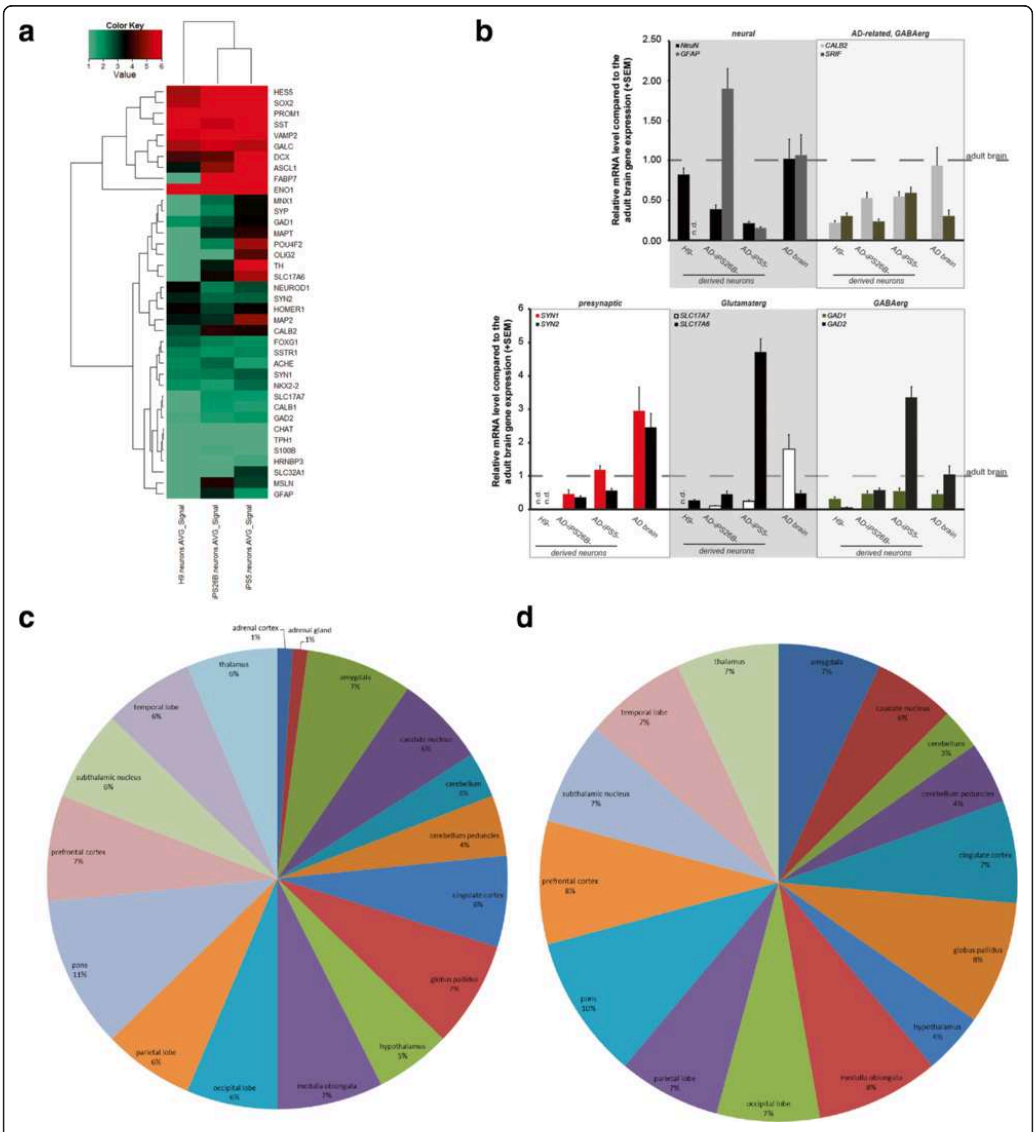


Figure 4 Expression of neuronal marker genes and genes associated with Alzheimer-related brain regions in AD-iPS neurons. **(a)** Cluster analysis of neuronal marker genes in AD-iPS neurons and H9 neurons based on Euclidean distance of microarray-based gene expression values. **(b)** A number of genes of the heatmap data in **(a)** were confirmed by real- time PCR analysis. Bars indicate the RNA level normalized to β-ACTIN first and compared to gene expression of adult brain tissue (non-diseased, male, 21 years; Amsbio). The samples of AD-iPS neurons and H9 neurons consisted of cRNA generated from RNA isolated from a single well of a single neuronal differentiation respectively. In addition, to our data we analyzed AD brain RNA (male, 87 years, diagnosed AD; Amsbio). **(c)** Brain regions associated with genes from the AD KEGG pathway (but neither in Huntington disease nor in Parkinson disease) down-regulated in AD-iPS5 compared to H9 neurons. **(d)** Brain regions associated with genes from the AD KEGG pathway (but neither in Huntington nor in Parkinson disease) down-regulated in AD-iPS26B compared to H9 neurons.

as *NeuN* (*HRNBP3* or *FOX1*), *GFAP* and *GAD1*, *GAD2* (GABA-ergic genes) on the one hand and specific AD-related neuronal genes such as *CALBINDIN1* and *2* (also known as calretinin) as well as *SST* (somatostatin or *SRIF*) on the other were analyzed. SRIF-positive interneurons are inhibitory neurons which express *GAD1* and/or *GAD2* as well as *CALB2* [31] and are the most affected subtypes of neurons in AD [32]. To confirm the array-derived heatmap data we analyzed relative gene expression by real-time PCR (Figure 4b) using the samples of the same neuronal differentiation experiment which were hybridized for transcriptome analysis. By matching the gene expression to adult brain RNA, again the H9-derived neuronal cells did not express subtype-specific neuronal genes. The astrocyte-specific gene *GFAP* was barely detected in all neural cells compared to the mRNA level of the adult and AD brain (Figure 4b). The neuronal cells from AD-iPS5 and AD-iPS26B were positive for *SYNAPSIN I*, *vGLUT2* (*SLC17A6*) and *GAD2* and have the same transcript level as the AD brain for *CALB2* and *GAD1* (Figure 4b).

Proof-of-principle drug discovery using sporadic AD-iPSC derived neuronal cells

In addition to gene expression based analysis of Alzheimer-related genes we evaluated the possible medical relevance of our neuronal cell model in terms of drug discovery and selection of an appropriate therapy for sporadic AD. For this purpose, we subjected the induced neuronal cells to treatment with the γ-secretase-inhibitor Compound E (CE). The experiment was carried out once. Two distinct concentrations were employed: low 10 nM and high 100 nM. After one week of treatment, cells were lysed directly and the protein expression levels of p-tau, tau, p-GSK3B and GSK3B were investigated isolating samples from one well of one inhibitor treatment experiment conducted. Neuronal cells derived from both AD-iPSC lines (AD-iPS5 and AD-iPS26B) exhibited the expression of tau and p-tau, which were undetectable in the parental fibroblasts (NFH-46) (Figure 5). The results were confirmed using two antibodies, one recognizing only p-tau and the other binding to both the phosphorylated and non-phosphorylated forms of tau (Figure 5). Drug treatment did not result in any reduction of p-tau in AD-iPS5 derived neuronal cells. On the other hand, we observed a significant reduction of p-tau and tau expression in neuronal cells differentiated from AD-iPS26B compared to untreated cells following high doses of CE (Figure 5). The expression of both p-GSK3B and GSK3B was significantly higher in neuronal cells compared to parental fibroblasts NFH-46 (Figure 5). However, no evident change in their expression could be identified following CE treatment. Neuronal cells were identified based on the expression of NESTIN and b-TUBULIN III, however, expression was also detected in their parental fibroblast cells, thus confirming previous observations in fibroblasts [33].

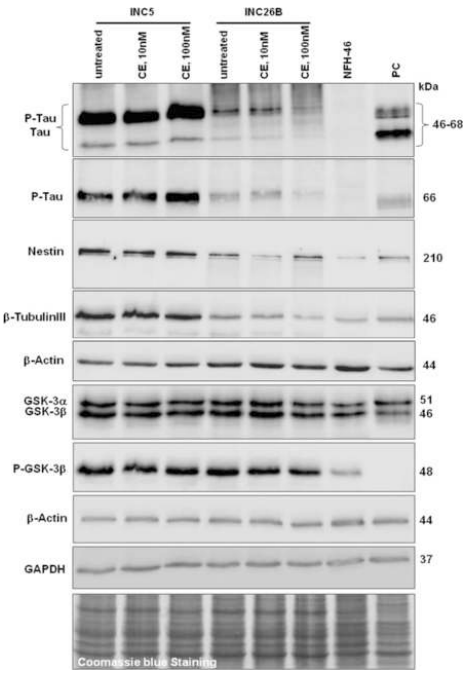


Figure 5 Treatment of AD neuronal cells with the γ-secretase-inhibitor compound E (CE). Western blot analysis was used to monitor the expression of phosphorylated tau (p-tau), non-phosphorylated tau, GSK3α/β and phosphorylated GSK3β in AD-iPS5 neurons (INC5) and AD-iPS26B neurons (INC26B). The samples were derived from one well of a single neuronal differentiation. The parental fibroblast cells (NFH-46) were included as control in addition to the positive control (PC) for p-tau represented by lysate from neuroblastoma. β-ACTIN and GAPDH and coomassie blue staining were used to confirm similar protein loading across samples.

Differential gene expression associated with Alzheimer-related pathways and biological processes in AD-iPSC neurons compared to H9 neurons

Using microarray based gene expression analysis we looked at the changes in the biological processes within the AD-iPS neuronal cells compared to H9 neurons as control. The hybridized samples were isolated from one neuronal differentiation experiment. Processes related to WNT signaling pathway and the alanine, aspartate and glutamate metabolism, in the case of AD-iPS5 neurons as well as the lysosome pathway and glutathion metabolism in the case of AD-iPS26B appeared to be up-regulated compared to H9 neurons. Pathways related to

Alzheimer's disease, Huntington's disease, Parkinson's disease and the proteasome were down-regulated in both AD-iPS neurons compared to H9 neurons (Additional files 9 and 10). AD-iPS5 and AD-iPS26B neurons showed up-regulated gene expression for biological processes such as neuronal fate commitment, neuron maturation, response to oxygen radical and/or response to reactive oxygen species (Additional file 9). In contrast to that, the UPS, apoptosis, and oxidative phosphorylation emerged as down-regulated biological processes (Additional file 10). Overall, these data suggest that AD neuronal cells exhibit alterations in key signaling pathways related to cell death, anabolism and catabolism in comparison to the healthy control.

AD-iPSC neurons show a distinct gene expression pattern of Alzheimer-associated genes of genome wide association studies compared to H9 neurons

To further analyze the reflection of Alzheimer-specific gene expression patterns in our iPSC-based model system we performed data mining to extract disease relevant gene expression using an Alzheimer gene list recently published by the European Alzheimer's Disease Initiative (EADI) [34]. The cluster analysis in Figure 6 showed that AD-iPS5 neurons and AD-iPS26 neurons were more similar to each other than to H9 neurons. Basically there are six gene clusters: (i) a cluster of genes expressed in all experiments such as *APOE* and *APP*, (ii) a cluster of genes expressed in no experiment such as *CASS4* and *CR1*, (iii) a cluster of genes expressed in both AD-iPSC neuron experiments but not in the H9 neuron experiment such as *PTK2B* and *PICALM*, (iv) a singleton cluster of *HLA-DRB5* not expressed in both AD-iPS neuron experiments but expressed in the H9 neurons, (v) a singleton cluster of *MEF2C* not expressed in AD-iPS26B experiments but expressed in AD-iPS5 and the H9 neurons and (vi) a cluster of genes expressed in AD-iPS26B neurons but not in AD-iPS5 and H9 neurons containing genes such as *SLC24A4* and *ABCA7*.

AD-iPSC neurons show down-regulation of genes involved in Alzheimer's, Huntington's and Parkinson's disease compared to H9 neurons

Analysis of differences in the iPS-derived neurons when compared to the annotations Alzheimer's disease, Parkinson's disease and Huntington's disease revealed that most genes down-regulated in AD-iPS5 vs. H9 neurons (Figure 7) and AD-iPS26B vs. H9 neurons (Figure 8) were common to all three neural disorders. Exclusively associated with Alzheimer's disease were 16 genes in AD-iPS5 vs. H9 neurons and 10 genes in in AD-iPS26B vs. H9 neurons. These genes were *APP*, *APOE*, *PSENEN*, *CDK5*, *HSD17B10*, *TNFRSF1A*, *PPP3CB*, *PPP3CC*, *CHP*, *GAPDH*, *CAPN2*, *CAPN1*, *ATP2A2*, *GSK3B*, *CALM3* and *CALM2* in the experiment AD-iPS5 vs. H9 neurons and *APP*, *CDK5*, *HSD17B10*, *CHP*, *GAPDH*, *NAE1*, *ATP2A2*, *GSK3B*, *CALM3* and *CALM2* in the experiment AD-iPS26B vs. H9 neurons.

Brain allocation of Alzheimer-specific genes down-regulated in AD-iPSC neurons compared to H9 neurons

The expression in different brain regions of the Alzheimer-exclusive genes that were found to be down-regulated in AD-iPSC neurons of one differentiation experiment were investigated using the GNF/Atlas organism part. The expression of the largest set of genes was allocated to pons with 12% in the case of AD-iPS5 neurons for the genes *APOE*, *APP*, *ATP2A2*, *CALM2*, *CALM3*, *CAPN2*, *CDK5*, *GAPDH*, *GSK3B* and *PPP3CB* (Figure 4c) and 13% in AD-iPS26B neurons for the genes *APP*, *ATP2A2*, *CALM2*, *CALM3*, *CDK5*, *GAPDH* and *GSK3B* (Figure 4d). This was followed by globus pallidus with 9% in AD-iPS5 neurons for the genes *APP*, *ATP2A2*, *CALM2*, *CALM3*, *CDK5*, *GAPDH* and *PPP3CB* and 9% in AD-iPS26B neurons for the genes *APP*, *ATP2A2*, *CALM2*, *CALM3*, *CDK5* and *GAPDH*. The percentages for medulla oblongata, prefrontal cortex and amygdala were found to be 5–8% (Figure 4c and d).

An Alzheimer-relevant functional protein association network can be built using an Alzheimer-specific gene set down-regulated in AD-iPSC neurons compared to H9 neurons

To further specify the reflected Alzheimer-related phenotype in our iPSC-based neuronal disease model we constructed protein association networks by means of the gene expression data generated from one differentiation experiment. Therefore, protein-interaction networks were generated using genes annotated with AD pathway that are down-regulated in AD-iPS5 neurons vs. H9 neurons, in AD-iPS26B neurons vs. H9 neurons as well as the overlap of both datasets. We successfully modeled the association of Alzheimer-related proteins within our cellular system in both AD-iPS neuronal differentiation experiments through the subsequent comparison to non-AD embryonic stem cell line H9 neuronal cells and further network construction by applying STRINGv9. The generated networks in Figures 9, 10 and 11 depict associations between proteins in a color code. The color of the line between proteins represents the following evidence categories: neighborhood in the genome (dark green line), gene fusion (red line), co-occurrence across genomes (dark blue line), co-expression (black line), experimental/biochemical data (purple line), association in curated databases (light blue line) and co-mentioning in PubMed abstracts/textmining evidence (light green line).

The protein association networks built from genes down-regulated in AD-iPS5 vs. H9 neurons in Figure 9

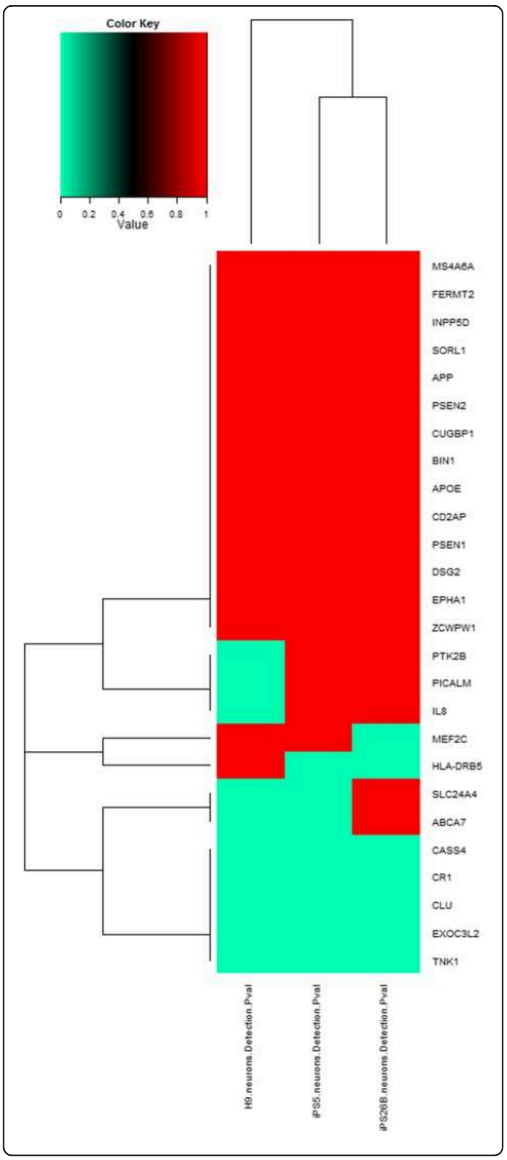


Figure 6 Expression of Alzheimer risk genes in AD-iPSC derived neurons. Cluster analysis of Alzheimer risk genes in experiments AD-iPS5 neurons, iPS26B neurons and H9 neurons of one neuronal differentiation experiment each. Up and down-regulated transcripts are depicted in red and green, respectively. RNA from AD-iPS5 neurons, AD-iPS26B neurons and H9 neurons was hybridized onto an Illumina human-8 BeadChip version 3. Alzheimer-associated genes known from genome wide association studies were filtered from the microarray experiments of AD-iPS5 neurons, AD-iPS26B neurons and H9 neurons. Illumina detection p-values were mapped to a binary scale (0 = not expressed if p-value > 0.05, 1 = expressed if p-value < = 0.05). These values were clustered via the R heatmap2 function using Euclidean distance as distance measure.

contains 18 more Alzheimer-related proteins than the AD-iPS26B vs. H9 neurons network in Figure 10. These proteins are NDUFB10, NDUFA9, NDUFB8, NDUFB9, ATP5G1, CAPN2, UQCRQ, NDUFA1, CAPN1, NDUFS7, SDHB, TNFRSF1A, CASP3, APOE, SDHD, PPP3CB, PPP3CC and PSENEN (Figure 9). However, the two proteins NDUFA6 and NAE1 are only part of a network built from the AD-iPS26B vs. H9 neurons dataset (Figure 10). Interestingly, both interaction networks contain APP and GSK3B as well as CDK5 and HSD17B10. In the AD-iPS5 vs. H9 neurons and not in the AD-iPS26B vs. H9 neurons network APP is depicted to be associated with CASP3 and APOE by experimental evidence and textmining evidence

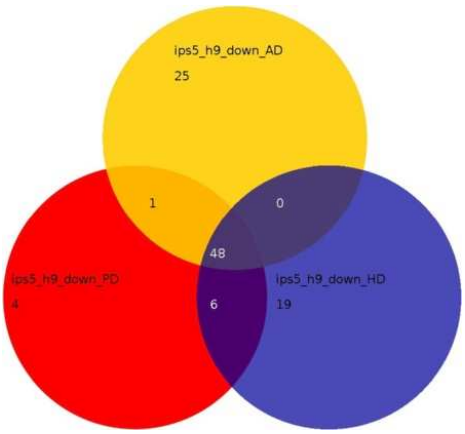


Figure 7 Gene expression associated with Alzheimer, Parkinson and Huntington disease in AD-iPS5 neurons. Venn diagram of down-regulated genes in the comparison AD-iPS5 vs. H9 neurons in KEGG pathways Alzheimer disease (AD), Parkinson disease (PD) and Huntington disease (HD). RNA from AD-iPS5 neurons and H9 neurons of one neuronal differentiation experiment was hybridized onto an Illumina human-8 BeadChip version 3. Functional annotation of significantly down-regulated genes from the experiments AD-iPS5 neurons vs. H9 neurons was performed with the DAVID Bioinformatics Resources 6.7. ips5 vs. H9: numbers of significant genes from KEGG pathways for ALzheimer, Parkinsons and Huntington.

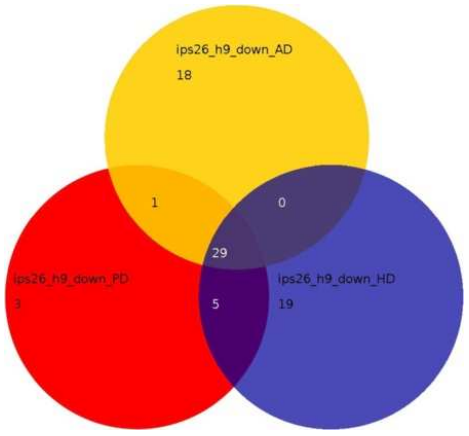


Figure 8 Gene expression associated with Alzheimer, Parkinson and Huntington disease in AD-iPS26B neurons. Venn diagram of down-regulated genes in the comparison AD-iPS26B vs. H9 neurons in KEGG pathways Alzheimer disease (AD), Parkinson disease (PD) and Huntington disease (HD). RNA from AD-iPS26B neurons and H9 neurons was hybridized onto an Illumina human-8 BeadChip version 3. Functional annotation of significantly down-regulated genes from the experiments AD-iPS26B neurons vs. H9 neurons was performed with the DAVID Bioinformatics Resources 6.7. ips26 vs. H9: numbers of significant genes from KEGG pathways for ALzheimer, Parkinsons and Huntington.

as well as with PSENEN, however, only via textmining evidence (Figure 9). In addition, only the AD-iPS26B vs. H9 neurons network depicts the association of APP and NAE1 through co-expression, database, experimental and textmining evidence (Figure 10). Furthermore, only in the AD-iPS5 vs. H9 neurons and not in the AD-iPS26B vs. H9 neurons network functional associations between CASP3 and CDK5, CAPN2, TNFRSF1A, CALM3, APOE and COX4I1 that are proven by textmining evidence could be found. TNFRSF1A which is only a part of the network in Figure 9 is associated with CDK5, CALM3 by textmining evidence and with CASP3 with additional database evidence. Exclusively in this protein association network APOE, which plays a major role in Alzheimer pathogenesis, is associated with CASP3, CALM3, CALM2 and GAPDH by textmining evidence whereas CAPN2 is associated with CASP3, CALM2 and CALM3 by textminig evidence. Surprisingly, no associations could be found for CAPN1 in this network. Additional proteins only part of the AD-iPS5 vs. H9 neurons network are PPP3CC and PPP3CB both of which are associated with each other by co-occurrence, database, experimental and textmining evidence. They are associated with CALM2 and CALM3 with experimental and textmining evidence. PSENEN, a subunit of the γ-secretase complex, occurs only in this protein association network and is associated with UQCRH via textmining and co-expression evidence. Additional proteins exclusively part of the AD-iPS5 vs. H9 neurons network are NDUFS7, NDUFA9, NDUFB8, NDUFB10, NDUFA1, NDUFB9, UQCRQ, ATP5G1, SDHB and SDHD. These are part of a complex protein association network-cluster mainly consisting of proteins involved in oxidative phosphorylation (Figure 9). Interestingly, NDUFA6 is only part of the protein association network based on the genes in the AD-iPS26B vs. H9 neurons dataset (Figure 10).

The network built from the genes overlapping between the AD-iPS5 vs. H9 neurons and AD-iPS26B vs. H9 neurons datasets in Figure 11, shows associations with experimental and textmining evidence of APP with HSD17B10, GAPDH, CDK5, GSK3B and SNCA. In addition, database evidence to prove the association of APP, SNCA and GAPDH could be found. Associations with textmining evidence between CDK5 with CALM2 and CALM3 as well as between GAPDH and ATP2A, CALM2, CALM3, SDHA, ATP5B and ATP5J could be found by our method. Furthermore, the interaction of GAPDH with ATP5B and ATP5J is associated with co-expression evidence in this network. GAPDH is associated with CHP via experimental evidence. Additional Alzheimer-related genes are interconnected to a complex protein association network cluster similar to Figures 9 and 10. consisting of proteins like NDUFB3, ATP5E, NDUFB5, UQCRC, NDUFB6, NDUFB7, ATP5B, NDUFAB1, NDUFB2 that are mainly involved in oxidative phosphorylation or in the electron transport chain in mitochondria. In our Alzheimer-related protein association network in Figure 11 we found experimental evidence association between SNCA and NDUFB6 as well as the co-expression evidence interactions of HSD17B10 with NDUFV2, NDUFB7, UQCRH and ATP5J. These associations connect Alzheimer-specific APP, GSK3B, CDK5, CALM2, CALM3 and ATP2A2 with proteins involved in Alzheimer related failure of the function of mitochondrial processes of the respiratory chain of the protein association cluster. A further association to proteins involved in oxidative phorphorylation are the interactions of GAPDH with ATP5J, ATP5B and SDHA (Figure 11).

A subset of UPS-related genes is down-regulated in AD-iPSC neurons compared to H9 neurons

The cluster analysis of UPS-associated genes assembled both AD-iPS neurons datasets into a cluster separated from the H9 neurons. Genes were divided into three clusters (i) 36 genes which have lower gene expression values in both AD-iPS neuron compared to H9 neurons, among them *PSMC1*, *PSMA5*, *NEDD8*. (ii) 2 genes characterized by higher expression in both AD-iPS neuron compared to H9 neurons: *PSMD5*, *PSMB9*. (iii) 25 genes the expression of which varies between the three samples. The last cluster

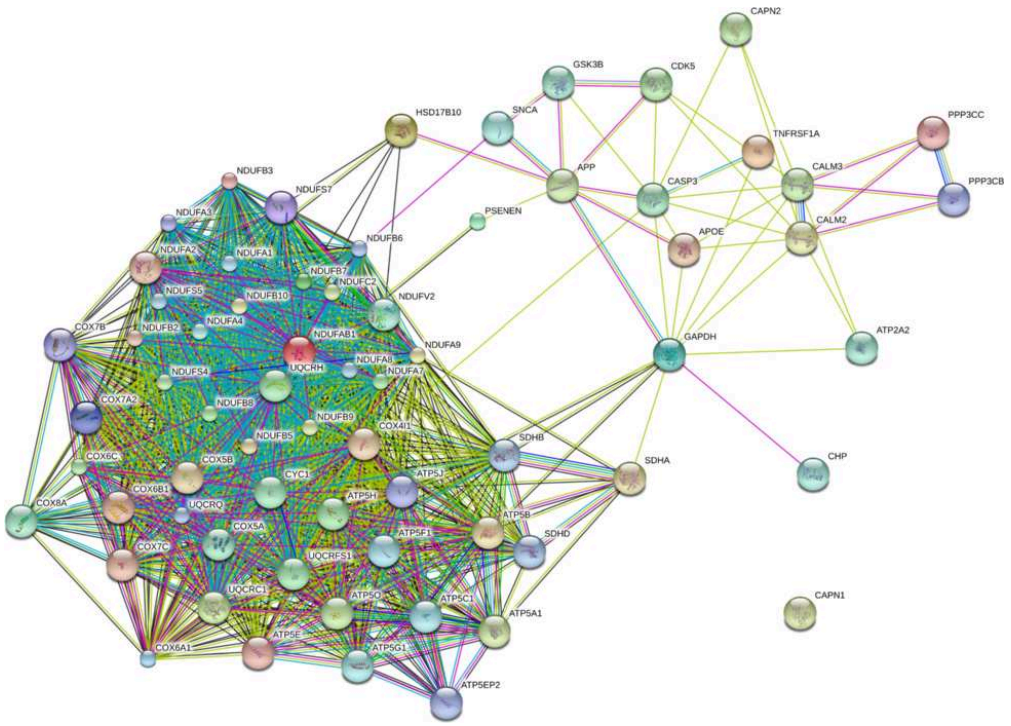


Figure 9 Alzheimer-related protein association network in AD-iPS5 neurons. Protein association network retrieved from STRING v9 using genes from the Alzheimer disease pathway down-regulated in the AD-iPS5 neurons vs. H9 neurons comparison of one neuronal differentiation each. The network circles represent proteins. The lines between the circles show the functional association. Co-expression evidence: black, database evidence: light blue, textmining evidence: yellow, experimental evidence: purple, co-occurrence evidence: blue, neighborhood evidence: green, fusion evidence: red.

is subdivided into (i) 5 genes with high expression values in H9 neurons and AD-iPS5 neurons but not in AD-iPS26B neurons - *PSME2*, *PSMD8*, (ii) 9 genes with high expression values in H9 neurons and AD-iPS26B neurons but not in AD-iPS5 neurons - *PSMD14*, *PSMD4*, (iii) 5 genes with a low gene expression values in H9 neurons and AD-iPS26B neurons and a high gene expression in AD-iPS-5 neurons - *PSMD4*, *SUGT1*, (iv) 6 genes with low gene expression values in H9 neurons and AD-iPS5 neurons and a high gene expression values in AD-26B neurons - *PSME1*, *PSMB10* (Figure 12).

Discussion

Little is known about the clinical onset and course of sporadic AD due to the limited insight and access to brain-derived neuronal cells from patients afflicted with neurodegenerative diseases. Therefore, it is essential that we develop new *in-vitro*-based experimental models that may reflect affected nerves in the brain. Thus, an early diagnosis could help in this regard to treat the affected individuals effectively and to test preventive approaches of sporadic Alzheimer's disease treatment. In recent time, several research groups independently and successfully differentiated somatic cells of AD patients directly or by iPSC-based approaches into neuronal cells, and examined them with respect to the molecular basis of disease development [22-26].

The results of these studies lead to valuable insights regarding understanding of AD molecular disease, given that these innovative and predictive patient cell models displayed the AD phenotype [22]. In our current study, we confirmed that dermal fibroblasts derived from sporadic AD patients could be induced to a pluripotent state (iPSCs) and further differentiated into neuronal cells in one reprogramming and one neuronal differentiation experiment per derived AD-iPSC clone. The transcriptomes of the derived neuronal cells were characterized by the expression of neuronal markers such as *GALC*,

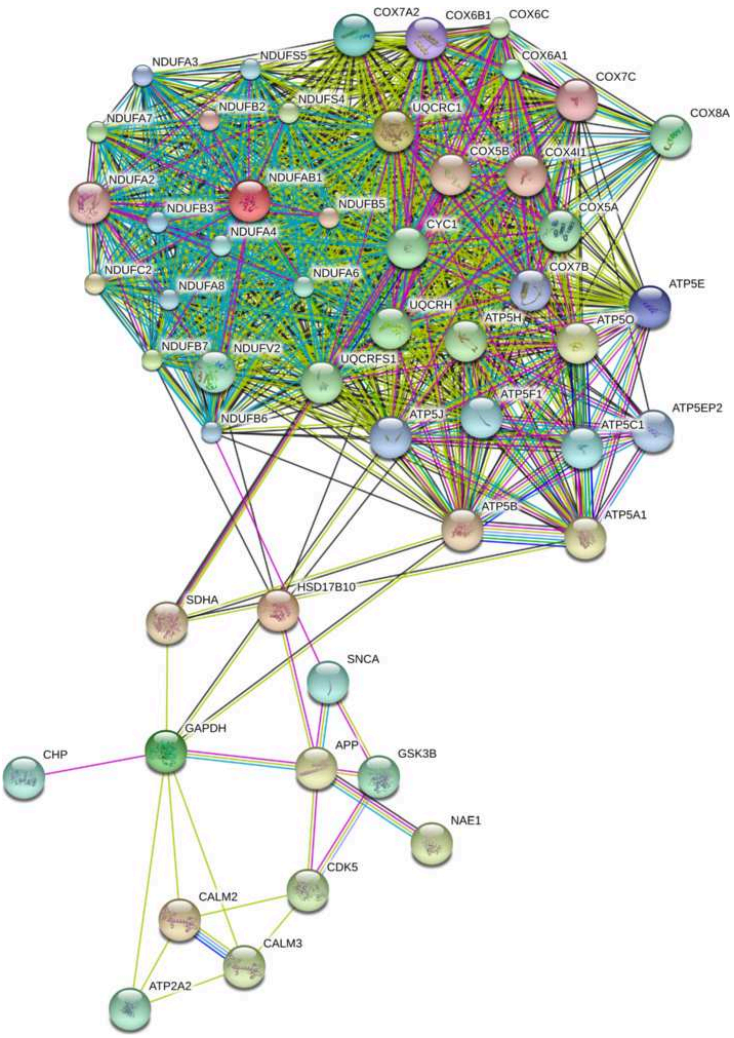


Figure 10 Alzheimer-related protein association network in AD-iPS26B neurons. Protein association network retrieved from STRING v9 using genes from the Alzheimer disease pathway down-regulated in the AD-iPS26B neurons vs. H9 neurons comparison of one neuronal differentiation each. The network circles represent proteins. The lines between the circles show the functional association. Co-expression evidence: black, database evidence: light blue, textmining evidence: yellow, experimental evidence: purple, co-occurrence evidence: blue, neighborhood evidence: green, fusion evidence: red.

MAP2, VAMP2, HES5, SOX2, PROM1 and AD-specific gene expression patterns when compared to control neuronal cells. We successfully generated an Alzheimer's disease-related protein association network using detected AD-related alterations of the transcriptome. As a control we used neuronal cells generated from the female embryonic stem cell line H9 in our study. This is a very clean background, it might be argued that some of the disease associated effects we see in our model come from differences between embryonic stem cells and iPS cells in general or are related to the advanced age of the AD patient (82-years-old) whose cells were used in this study. An effect of the age-related high mutation load in the parental fibroblast

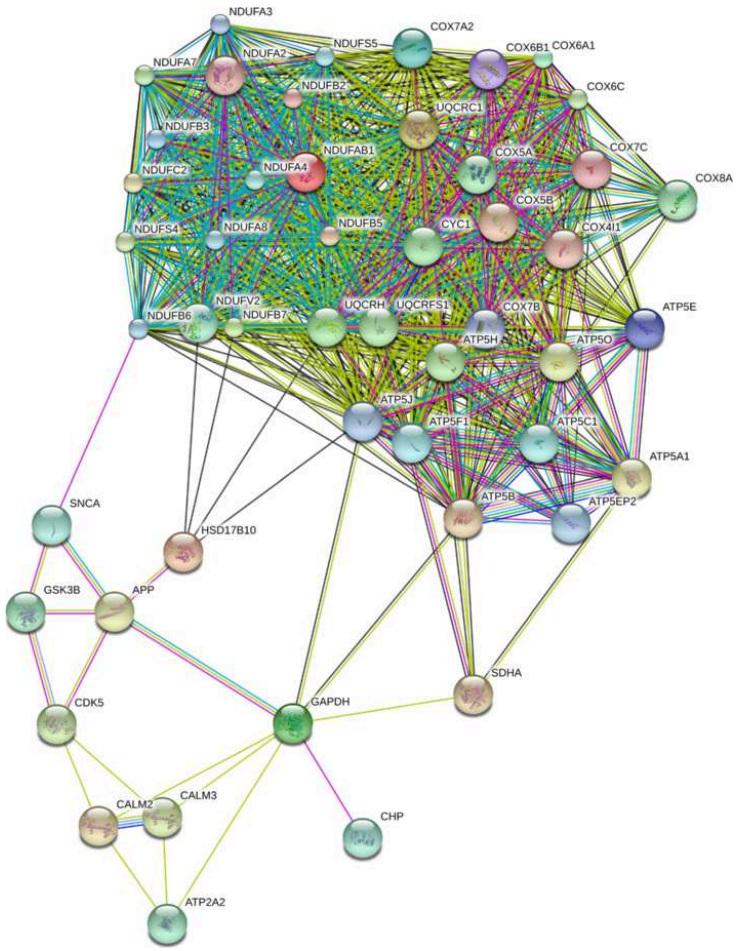


Figure 11 Overlapping Alzheimer-related protein association network of AD-iPS5 neurons and of AD-iPS26B neurons. Protein association network of genes down-regulated in both AD-iPS5 neurons vs. H9 neurons and in AD-iPS26B neurons vs. H9 neurons comparisons of one neuronal differentiation. Protein interaction network retrieved from STRING v9 using genes from the Alzheimer disease pathway down-regulated in AD-iPS5 vs. H9 neurons and in AD-iPS26B vs. H9 neurons experiments. The network circles represent proteins. The lines between the circles show the functional association. Co-expression evidence: black, database evidence: light blue, textmining evidence: yellow, experimental evidence: purple, co-occurrence evidence: blue, neighborhood evidence: green, fusion evidence: red.

compared to H9 cannot be excluded but could be significant since AD is an aged-related disease and as such one would expect a higher mutation load in patients. An age-matched control iPS cell line would have been an alternative to H9 for our study. Many iPSC-based disease models compare the effect of mutations in a disease associated gene within the same genetic background. The parental fibroblasts used for AD-iPSC generation in this study did not carry any mutations in Alzheimer-related genes which excluded the possibility to use the same genetic background in our case. Based on the fact that *FABP7* a gene involved in neuronal development [35] is not expressed in H9 neurons, one could argue that the mixture of neuronal subtypes found in the neuronal differentiation of H9 is distinct from the neuronal differentiations of AD iPSCs. Despite possible differences in neuronal differentiation efficiencies or mixture of neuronal subtypes we see altered expression of AD-related genes in two neuronal differentiations of AD patient fibroblast derived

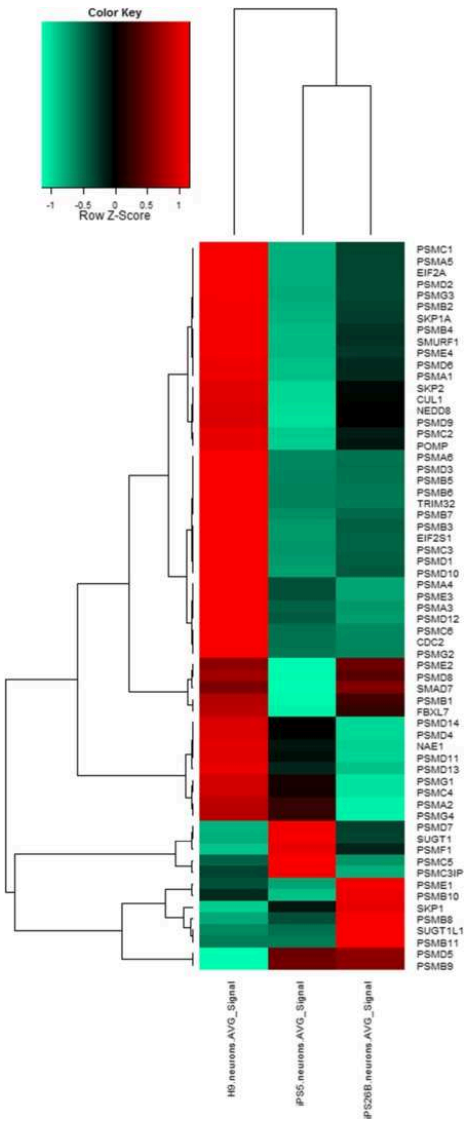


Figure 12 UPS-associated gene expression in AD-iPSC neurons. Cluster analysis based on Pearson correlation of UPS-related genes in AD-iPS neurons and H9 neurons. Up and down-regulated transcripts are depicted in red and green, respectively. RNA from AD-iPS5 neurons, AD-iPS26B neurons and H9 neurons of one well of one neuronal differentiation each was hybridized onto an Illumina human-8 BeadChip version 3. Known proteasome-related genes were selected from the microarray experiments of AD-iPS5 neurons, AD-iPS26B neurons and H9 neurons. Their Illumina average signal intensities were transformed to a logarithmic scale (log2) and clustered with the R heatmap2 function using Pearson correlation as similarity measure.

iPS cell lines compared to the neuronal differentiation of H9 used as a control.

At the protein level, the neuronal cells which we derived from sporadic AD-iPSCs in a single neuronal differentiation experiment expressed p-tau and GSK3B, both valid as neuropathological proteins. Analysis of the brain of sporadic AD patients often shows intracellular accumulation of hyperphosphorylated tau proteins, an early event preceding the appearance of NFT in AD [6]. Physiological, GSK3β is a multifunctional protein kinase that phosphorylates a variety of substrates including the tau protein, which is associated with neuronal-specific microtubules. Surprisingly, in our cell model we observed the up-regulation of both isoforms of the GSK3α and GSK3β protein as well as clear formation of abnormal p-tau (Thr 231) in AD-iPSC derived neuronal cells in comparison to their parental fibroblast cells, which expressed minute levels of GSKαβ and no tau.

In a recently published study on neuronal cells derived from AD-iPSC of familial AD, abnormal p-tau expression was not detected by western blotting, probably due to the short time scale in culture [25]. Nevertheless, another study could show increased p-tau in both familial and one sporadic AD sample, however measured by the MSD phospho tau kit [22]. In the same study high amounts of GSK3β were measured in induced neuronal cells from an AD donor, this is in accordance with the observation in our AD-iPSC derived neuronal cells.

Global gene expression analysis revealed up-regulation of AD-related pathways in AD-iPS neurons of one neuronal differentiation experiment for each AD-iPS cell line such as WNT, the lysosome signalling pathway and glutathion metabolism all of which have been shown to be altered in Alzheimer's disease [36-38]. Interestingly, we detected altered expression of genes that are involved in the alanine, aspartate and glutamate metabolism. A very recent work showed through meta-analysis of genome-wide association studies involving 2540 Alzheimer cases that changes in glutamate metabolism are overrepresented in data from patients with AD [39].

In addition, biological processes involving response to oxygen radical and response to oxidative stress were up-regulated in our sporadic AD model. Indeed these processes are known to be up-regulated in Alzheimer's disease [40]. Moreover, biological processes such as neurogenesis appeared to be up-regulated, underlining successful neuronal differentiation. The mixture of neuronal subtypes generated by our neuronal differentiation experiments might vary as we did not carry out subtype specific differentiations.

A cluster analysis of the expression of AD-associated genes recently published confirmed differential expression in the generated AD-iPS neurons compared to H9 neurons [34].

Even though 19 of 26 genes of the cluster analysis of Alzheimer-related genes seemed to be expressed in a similar manner in H9 neurons compared to AD-iPS neurons, the transcriptomes of AD-iPS neurons were more similar to each other than to H9 neurons despite a possible variation of neuronal subtype mixtures in the conducted neuronal differentiation experiments. Differential behavior between both AD-iPS neurons experiments and the H9 neurons experiment points to genes whose up-regulation (*PTK2B*, *PICALM*, *IL8*) or down-regulation (*HLA-DRB5*) may play a major role in development of the disease in this patient.

While *PTK2B*, *IL8* and *HLA-DRB5* are clearly involved in Alzheimer pathology [34,41,42], there are controversial studies about the involvement of PICALM in Alzheimers disease [43,44].

The genes *EXOC3L2*, *CLU*, *CR1* and *TNK1* although associated with AD were not found to be expressed in AD-iPS neurons. The gene *EXOC3L2* has been associated with late onset Alzheimer disease (LOAD) in GWAS [45]. However, it could be shown that the association is likely to be caused by the close location to *APOE* and there was found no more evidence after adjustment for *APOE* [46]. Our results showing no expression for *EXOC3L2* in AD-iPS5 and AD-iPS26B neurons are in line with that finding.

The AD-association with *TNK1* which was also not expressed significantly in our experiments is unclear as several studies report ambiguous results [47,48].

In contrast to that, differentially expressed genes in AD-iPS neurons revealed the down-regulation of pathways annotated to Alzheimer's disease, Huntington's disease and Parkinson's disease. The overlap of down-regulated gene expression related to Alzheimer's disease, Huntington's disease and Parkinson's disease in our AD-iPS neuronal cells in Figures 7 and 8 supports the notion of a pathological mechanism common to these three neurodegenerative diseases [49].

Sixteen Alzheimer's disease-specific genes could be confirmed to be down-regulated in AD-iPS5 (Figure 7) and 10 genes in AD-iPS26B compared to H9 neurons (Figure 8). These genes could be allocated to brain regions which are affected by Alzheimer's disease depicted in Figure 4b and c. In both AD-iPS neuron experiments the largest number of Alzheimer-related genes was allocated to pons. Indeed, pons has been reported to have a smaller volume in patients with familial Alzheimer's disease [50]. The second largest gene set was allocated to the brain region globus pallidus for both AD-iPS neuronal cells, a brain region which was reported to be involved in Alzheimers disease [51]. Finally, the medulla oblongata and amygdala brain regions which show deregulated gene expression in AD iPS neurons in our model were reported to be affected in AD [52,53].

We could successfully generate a protein association network consisting of AD-specific genes down-regulated in AD-iPS neurons compared to H9 neurons. The networks reflects the differences between the two AD-iPS neurons as there are more Alzheimer risk genes (for example, PSEN and APOE) present in the network built from AD-related genes down-regulated in AD-iPS5 neurons against healthy H9 neurons. Despite these differences the construction of a STRING-based protein association network representing the AD phenotype was possible with our approach for both AD-iPS neuron cell populations. APP, a protein which plays a central role in the pathology of Alzheimer's disease, is part of the protein association network in both AD-iPS neuron cultures as well as in the association network built from the overlap of both experiments. Several studies report protein interaction networks characterizing the molecular disease phenotype in post mortem brains of sporadic AD patients or patients with familial Alzheimer disease [54,55]. Despite the usefulness of these networks for gaining insights into molecular changes in the final stage of AD, no information can be drawn about the early molecular pathology of AD with this approach. Generating iPS from AD patients in the early stage of the disease would allow modeling disease-specific changes in the AD-related protein association network over time, which provides valuable information of the development of new therapy approaches at early stages. So far, our study is the first demonstration of a protein interaction network of AD-iPS neurons derived from skin cells from an 82-year-old sporadic AD patient. Modeling sporadic AD using iPS technology as presented here enables us to formulate hypotheses to increase our understanding of AD pathogenetic mechanisms and test them by monitoring the effect on the protein association network.

Next to differential gene expression of AD-specific genes in AD-iPS derived neurons we detected the down-reglation of genes in AD-iPS neuronal cells that play a major role in the UPS. Our data revealed that the majority of the UPS-related genes are down-regulated in AD-iPS neurons compared to H9 neurons, suggesting that UPS functionality is lowered in our AD-iPS neurons but not in healthy H9 neurons. Indeed, UPS deficiency has been associated with AD pathology [18]. These results suggest that our Alzeimer model very likely reflects UPS-related features of AD pathology which are most probably present in the neuronal cells of the sporadic AD patients.

Furthermore, the UPS-related gene expression data suggests that both the constitutive and inducible proteasome play a role in AD pathology. This is reflected in the lower gene expression of the main constitutive subunits of the proteasome *PSMB5/6/7* in iPS-derived neuronal cells from the AD patient compared to H9 neurons and the

higher gene expression of *PSMB9* in both AD-iPS neurons compared to H9 neurons. The genes *PSMB8* and *PSMB10*, the expression of which is higher than in H9 neurons only in the case of AD-iPS26B neurons are probably less important in AD-related pathology driven by the UPS as their expression showed a higher variation. Indeed, constitutive proteolytic activities have been reported to be decreased in AD brains, meanwhile the composition of the proteasome complex is not affected [56]. Interestingly, in contrast to the constitutive proteolytically-active subunits, the inducible ones have been reported to be highly expressed in the hippocampus (HC) of severe diseased AD patients (Braak stage ≥ III) [57,58].

In addition, we found that the expression of the gene *NEDD8* is down-regulated in AD-iPS neurons compared to the control. It is known that NEDD8 plays a role in AD pathology. The APP binding protein-1 (APP-BP1) is also increased in the AD-affected HC [16]. APP-BP1 associates with UBA3 resulting in an E1-like activating enzyme for the process of NEDD8, an ubiquitin-like protein [16]. However, NEDD8 was additionally found to be present in high amounts in neurofibrillary tangles (NFTs) and senile plaques from a patient with AD [59,60], which would suggest that NEDD8 is up-regulated in AD, which we do not see in our iPS-based model. Our data suggests that UPS dysfunction may occur early in AD pathogenesis eventually leading to cellular protein aggregates later on.

Using our neuronal cell model, we provide a proof of principle that neuronal cells differentiated from patient dermal fibroblasts-derived iPS cells offer a valuable tool for modeling early molecular pathology of AD, screening and development of appropriate drugs for the treatment of AD in the future. However, our results suggest that the different iPSC clones even derived from the same individual may give rise to different responses which is reflected in the differences between the two protein association networks generated from the gene expression data of the two different AD-iPS clones differentiated to neuronal cells in one experiment each. In our study we generated two sporadic AD-iPSC lines: one (AD-iPS26B) exhibited complete teratoma formation and normal karyotype whilst the other cell line (AD-iPS5) showed karyotype abnormalities and failed to differentiate into endoderm *in-vivo*. Due to the fact that the teratoma assay is not standardized [29] the trend now is towards a transcriptome-based classification of pluripotency (PluriTest) rather than the traditional teratoma-based assay [61]. Interestingly, only AD-iPS26B responded to the **γ**-secretase-inhibitor CE treatment in one inhibitor treatment experiment. This would suggest that the selection of iPSC clones is critical to enable the generation of results that are clinically relevant. However, neurons from AD-iPS5 showed a higher number of AD-related genes that are deregulated compared to AD-iPS26B neurons as reflected in the different protein association networks built from these two neuronal cell differentiation experiments. Therefore, the failure to respond to CE could be a reflection of this difference in AD-related gene expression between the two clones. Our findings, although based on a limited number of AD-iPSCs, have highlighted the fact that molecular pathology of sporadic Alzheimer can be modelled in disease-related protein association networks by means of iPSC technology and transcriptome analysis. Abnormalities of the karyotype in the parental AD-iPS cells should be avoided but as in our case do not necessarily lead to a distortion of gene expression-based AD protein association networks with this approach.

Conclusion

In summary, we have generated patient-specific pluripotent stem cells from skin fibroblasts of an 82-year-old woman suffering from sporadic AD and induced these to differentiate into neuronal cells. The patient-derived cells recapitulated key features of the disease, including the expression of p-tau, GSK3β, down-regulation of AD-related genes and altered biological features caused by differential expression of genes involved in e.g. the UPS or response to oxidative stress.

The majority of the UPS genes are down-regulated in AD neurons, thus supporting the idea that dysfunctions in this system may occur early in AD pathogenesis and then lead to cellular protein aggregates at later stages of the disease. Additionally, a few genes were distinctly up-regulated in neurons derived from the two AD-iPSC lines. This may suggest that these genes are less important for AD pathogenesis as their expression varies among different neuronal lines.

In essence, we have successfully generated an Alzheimer-related protein association network using characteristic gene expression patterns detected in AD-iPSCs compared to a healthy control of one neuronal differentiation experiment. Our results lend further support to the fact that neuronal cells differentiated from iPSCs from sporadic AD patients in part recapitulate the neuropathological processes of the disease. We anticipate that iPSC-based modeling of AD as demonstrated here can be useful for formulating testable hypotheses that might eventually enhance our meager knowledge of the molecular basis of its progression and should eventually lead to the development of new drugs to prevent or treat this disease.

Methods

Ethics statement

Full-thickness skin biopsy was resected from the forearm of the patient undergoing surgery. The Charité University Medicine Berlin ethics committee specifically approved

this study. Ethical agreement was preliminarily obtained from the guardian of the participant including written informed consent.

Cell culture

Adult dermal fibroblasts from an 82-year-old woman (NFH-46) suffering from late-stage AD under no medication and with no family history were obtained from 6 mm full-thickness skin biopsies originating from the sun-protected forearm inner side. The skin specimens were incubated in dispase solution (2.4 U/ml) overnight at 4°C. After separation of epidermis, primary dermal fibroblasts were isolated from dermis by enzymatic digestion and were expanded within 4 weeks. The dermal fibroblasts were cultured in DMEM supplemented with 10% fetal calf serum (FCS), nonessential amino acids, L-glutamine, penicillin/streptomycin and sodium pyruvate (all from Invitrogen, Carlsbad, CA). The human embryonic stem cell (hESC) lines H1 and H9 were purchased from WiCell, Madison, WI (#WA01 and #WA09, respectively). Control iPSCs (OiPS3 and OiPS6) have recently been generated from the skin of an 84-year-old female (NFH-2 fibroblasts) [62]. hESCs and iPSCs were cultured in hESCs media containing KO-DMEM supplemented with 20% knockout serum replacement, nonessential amino acids, L-glutamine, penicillin/streptomycin, sodium pyruvate, 0.1 mM β-mercaptoethanol (all from Invitrogen) and 8 ng/ml basic fibroblast growth factor (bFGF) (Preprotech; Rocky Hill, NJ). Cultures were maintained on mitomycin C-inactivated mouse embryonic fibroblasts (MEFs) and passaged manually. For experiments feeder layer-free iPSCs and hESCs were grown on dishes coated with Matrigel (BD; San Diego, CA) in MEF-conditioned media (CM). All cultures were kept in a humidified atmosphere of 5% CO2 at 37°C. Experiments were carried out with primary fibroblasts at passages 3–6 and AD-iPS cells at passages 8–18.

Retroviral transduction into fibroblasts and derivation of iPSCs

NFH46-derived iPSCs (AD-iPS5 and AD-iPS26B) were obtained using the Yamanaka retroviral cocktail [27] in one reprogramming experiment. Briefly, pMX vector-based OCT4, KLF4, SOX2 and c-MYC retroviruses were generated using 293 T cells, according to the conventional $CaCl_2$ transfection protocol. 200,000 fibroblasts were used as input for reprogramming experiments and seeded into six wells of a six well plate. Four weeks after transduction, hESC-like colonies were manually picked and expanded for characterization as previously described [27,28].

DNA fingerprinting analysis

In order to confirm somatic origin and to exclude cross reaction with hESCs, DNA fingerprinting analysis was performed as previously described [28]. 100 ng of genomic DNA isolated from one well per AD-iPS cell line were used for PCR amplification, following this program: 94°C for 1 min, 55°C for 1 min, 72°C for 1 min, for 40 cycles, using Dyad thermal cycler (BioRad, Hercules, CA). PCR products were resolved in 2.8% agarose gels to examine the differential amplicon mobility for each primer set: D7S796, repeat (GATA)n, average heterozygosity = 0.95. The primer sequences are listed in Additional file 11.

Quantitative real-time polymerase chain reaction

Quantitative real-time polymerase chain reaction based gene expression analysis of pluripotency-associated and neural genes was carried out using the ABI PRISM SDS 2.1 software (Applied Biosystems, Foster City, CA) and Microsoft Excel [63]. The primer sequences are listed in Additional file 11. The data are presented as relative gene expression based on the ΔΔCt calculation over NFH-46 (Additional file 4) or adult brain tissue (Figure 4) with respect to standard error of mean (SEM). The real time PCR analysis was performed using triplicates for each repetition (n = 3). Real time PCR to confirm pluripotency gene expression was performed on 3 independent wells of AD-iPS5 and AD-iPS26B respectively. Each iPS line was split from one well into the three wells, and expanded, prior to RNA isolation. The RNA samples were not pooled. The real time PCR confirming the expression of neuronal markers was performed using cRNA derived from RNA of one neuronal differentiation of a single well of AD-iPS 5, AD-iPS 26B and H9 respectively.

Confirmation of functional pluripotency *in-vitro*

For *in-vitro* differentiation, embryoid bodies (EBs) were generated from AD-iPSCs in one differentiation experiment each by cell harvesting and seeding onto low-attachment dishes in DMEM supplemented with 10% FCS, nonessential amino acids, L-glutamine, penicillin/streptomycin and sodium pyruvate (all from Invitrogen) without bFGF supplementation. One week later, EBs were plated onto gelatin-coated tissue culture dishes, grown for additional ten days, and analyzed by immunofluorescence-based detection of the expression of germ layer-specific proteins.

Confirmation of functional pluripotency *in vivo*

In-vivo teratoma assays were performed by EPO-Berlin GmbH, Berlin-Buch, Germany. AD-iPSCs were collected by trypsinization, washed and injected s.c. into NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ mice, commonly known as NOD scid gamma (NSG). Histological analysis was performed at the Institute for Animal Pathology, Berlin, Germany.

Neuronal differentiation

Pluripotent stem cells (iPSCs and hESCs) were mechanically dissociated and grown in MEF-conditioned medium (CM) on matrigel-coated dishes for 72 h. Induction of neuronal cells was performed by adding 10 μM SB431542 (SB, TGFβ receptor inhibitor) and 1 μM PD0325901 (PD, MEK1/2 inhibitor) (both from Sigma-Aldrich, Deisenhofen, Germany) in the absence of bFGF [30]. The cells were grown for additional four weeks under daily medium change. The obtained neuronal cells were fixed for immunostaining or used for drug treatment. The neuronal differentiation and subsequent inhibitor treatment were carried out once per AD-iPS cell line and for the control H9.

Alkaline phosphatase analysis and immunofluorescence staining

Alkaline phosphatase (AP) activity was visualized by the commercial AP staining kit (Millipore #SCR004; Schwalbach, Germany) according to the manufacturer's instructions. In order to characterize pluripotency of all AD-iPSC colonies, the cells were fixed in phosphate-buffered saline (PBS) containing 4% paraformaldehyde (Science; Munich, Germany) for 20 min at room temperature, subsequently washed twice with PBS without Ca2+ and Mg2+, blocked with 10% FCS serum (Vector; Loerrach, Germany) and 0.1% Triton X-100 (Sigma-Aldrich, Germany) in PBS and proceeded to immunocytochemistry with primary antibodies against OCT4, SOX-2, KLF-4, SSEA1, SSEA4, TRA-1-60 and TRA-1-81 from the hESC characterization kit (all 1:100, Millipore #SCR004), NANOG (1:100, Abcam #ab62734, Cambridge, UK), Smooth-Muscle-Actin (SMA) (1:100, Dako #M0851, Hamburg, Germany), Alpha-Fetoprotein (AFP) (1:100, Sigma-Aldrich #WH0000174M1), SOX17 (1:50, R&D #AF1924, Minneapolis, MN), PAX6 (1:300, Covance #PRB-278P, Münster, Germany), Nestin (1:200, Chemicon #MAB5326, Nürnberg, Germany), b-Tubulin III (1:1000, Sigma-Aldrich #T8660), Brachyury (T) (1:50, R&D #AF2085). Alexa-488-conjugated secondary antibodies were used (1:300, Invitrogen #A11001). Nuclei were counter-stained with DAPI (200 ng/ml, Invitrogen #H357) and visualized using the confocal microscope LSM510 (Carl Zeiss, Jena, Germany).

Western blot

Human AD-iPSCs and their corresponding fibroblast cells were detached from cell culture dishes by incubation with AccutaseTM (Millipore). For protein extraction, cells were harvested 48 h after compound-E treatment and lysed in RIPA-buffer supplemented with complete protease and phosphatase inhibitors cocktail (Roche, Penzberg, Germany). The protein extracts were derived from one well of a six well plate of a single neuronal differentiation. Extracts were homogenized and centrifuged at 10000 × g for 10 min. SDS–polyacrylamide gel electrophoresis (PAGE) and Western blot analysis of total proteins were performed [64]. Western blots were incubated with anti-tau monoclonal antibody (1:1000, Cell Signaling #4019; Frankfurt, Germany), anti p-tau (1:200, Santa Cruz #sc-32276; Heidelberg, Germany), anti GSK3A/B (1:1000, Cell Signaling #5676), anti-p-GSK3B (1:1000, Cell Signaling #5558), anti-b-Tubulin-III (1:1000, Sigma Aldrich #T8660) anti-Nestin (1:200, Chemicon #MAB5326), anti-Actin, anti-GAPDH (1:1000, Cell Signaling #5142). Neuroblastoma cell lysates were used as positive control for p-tau detection (Santa Cruz #SC-2410). Following incubation with a peroxidase-labelled anti-rabbit and anti-mouse secondary antibody (1:5000, Dako, #p0448, #p0447), antigen–antibody complexes were detected by ECL Western blotting detection reagents (Peqlab; Erlangen, Germany) for 1 min and exposed to imaging with the Fusion-FX7 imaging system (Peqlab).

Global gene expression analysis

Total RNA isolated from one well of one reprogramming or neuronal differentiation experiment was quality-checked by Nanodrop analysis (Nanodrop; Wilmington, DE, USA) and 500 ng were used as input. Biotin-labeled cRNA was produced using a linear amplification kit (Ambion; Austin, TX, USA). Hybridizations, washing, Cy3-streptavidin staining and scanning were performed on the Illumina BeadStation 500 platform (Illumina; San Diego, CA, USA) according to the manufacturer's instruction. cRNA samples were hybridized onto Illumina human-8 BeadChips version 3. The following samples were hybridized in duplicate: H9, NFH-46, NFH-2, AD-iPS5, AD-iPS26B, AD-iPS5 neurons, iPS26B neurons and H9 neurons. All basic expression data analysis was carried out using the BeadStudio software 3.0. Raw data were background-subtracted and normalized using the "rank invariant" algorithm and then filtered for significant expression based on negative control beads. Gene expression results were deposited in the Gene Expression Omnibus (GEO) repositary website http://www.ncbi.nlm.nih.gov/geo/; GEO data access number: GSE42492.

Cluster analysis of expression of Alzheimer-associated genes

Genes associated with Alzheimer in a published genome-wide association study (GWAS) [34] and known neuronal marker genes were filtered from the microarray experiments of AD-iPS5 neurons, AD-iPS26 neurons and H9 neurons from one neuronal differentiation experiment each. Illumina detection p-values were mapped to a binary scale (0 = not expressed if p-value > 0.05, 1 = expressed if p-value < = 0.05). These values were clustered via the R heatmap2 function using Euclidean distance as distance measure.

Differences and commonalities between neural disorders Alzheimer disease, Huntington disease and Parkinson disease

Functional annotation of significantly regulated genes from the experiments AD-iPS5 neurons vs. H9 neurons and AD-iPS26 neurons vs. H9 neurons from one neuronal differentiation was performed with the DAVID tool [65,66]. We focused the analysis on Alzheimer disease-related KEGG pathways Alzheimer disease, Huntington disease and Parkinson disease. Only down-regulated genes (compared to the reference of H9 neurons) were significantly enriched in these pathways and were therefore intersected via Venn diagrams to find overlaps and differences in the iPS cells concerning these neural disorders. All disjunct sets from this analysis are comprised in an excel table.

Analysis of brain regions associated with regulated genes

Association of genes with tissues was downloaded from the Ensembl/Biomart Human genes 75 (GRCh37.p13) GNF/Atlas organism part annotation data set accounting for gene expression in organism parts. Only brain regions were used for follow-up processing. These brain regions were mapped to genes significantly down-regulated in the AD-iPS neuron experiments which were carried out once per AD-iPS cell line and were in the KEGG Alzheimer disease pathway but neither in Huntington disease nor in Parkinson disease pathways.

Cluster analysis of proteasome-specific genes

Genes associated with the proteasome in the literature and in the gene definition annotation were selected from the microarray experiments of AD-iPS5 neurons, AD-iPS26B neurons and H9 neurons from one neuronal differentiation each. Their Illumina average signal intensities were transformed to a logarithmic scale (log2) and clustered with the R heatmap2 function using Pearson correlation as similarity measure.

Building of protein association networks

Differentially expressed genes were filtered from the microarray data by comparing the signal intensities of AD-iPS neurons and H9 neurons from one neuronal differentiation experiment each. Genes with signal intensity ratios below 0.8 were considered as down-regulated. The list of official symbols of these genes was used as input for gene annotation analysis using DAVID Bioinformatics Resources 6.7 [64,65]. Subsequently, a gene list with the KEGG annotation Alzheimers disease was used as input for building of the protein association networks using STRING v9.1 [67,68].

Sequence analysis

In this study, we performed a systematic analysis of the entire coding region including flanking intron sequences of the genes *APP*, *PSEN1* and *PSEN2* by direct sequencing. The target fragments were amplified by polymerase chain reaction (PCR) using intronic primers designed from genomic sequence with the Primer 3 software. PCR products were purified by Exo/SAP digestion (Exonuclease I; New England; Beverly, MA; shrimp alkaline phosphatase; Promega, San Diego, CA, USA) and directly sequenced using ABI-PRISM BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) and the ABI-PRISM 3730 DNA Analyzer, as described by the manufacturer. Sequences were analyzed using Mutation Surveyor software v3.24 (SoftGenetics LLC, State College, PA). The primer sequences are listed in Additional file 12.

Measurement of telomerase activity

The enzyme activity of telomerase was determined using the TraPEZE RT Telomerase Detection Kit (Millipore #S7710) [61].

Karyotype analysis

For detection of possible karyotype abnormalities in two AD-iPS cell lines, chromosomal analysis after GTG-banding was performed at the Human Genetic Center, Berlin, Germany. For each cell line, 25 metaphases were counted and 6 (NFH-46), 8 (AD-iPS26B) and 10 (AD-iPS5) karyograms were analyzed [28].

Inhibition of γ-secretase

In order to test the pharmacological response capabilities of AD-derived neurons, the same cells were treated with compound E, 2S-2-{[(3,5-difluorophenyl)acetyl]amino}-N-[(3S)-1-methyl-2-oxo-5-phenyl-2,3-dihydro-1H-1,4-benzodiazepin-3-yl]propanamide (γ-secretase-inhibitor; Calbiochem; Darmstadt, Germany). Two distinct concentrations were used: low 10 nM and high 100 nM. The inhibitor was added to the medium every alternate day for four weeks. The treated neuronal cells were then harvested in RIPA-buffer for protein analysis. One neuronal differentiation and subsequent inhibitor treatment was conduted. Protein was isolated from one well of the respective experiment.

Detection of HLA haplotype

In order to further characterize the NFH-46 cell line and its immunogenetic association with AD, genomic DNA was isolated with the FlexiGene DNA kit (Qiagen; Hilden, Germany). The purified DNA was then applied to genotyping of the HLA–A,B,C and DRB1 genes by PCR with sequence-specific primers (HLA DNA typing kit; Olerup; Vienna, Austria) and sequence-specific oligonucleotide hybridization (LABType HD; BmT, Meerbusch, Germany) according to the manufacturer's instructions.

Statistical analysis

Data were analyzed using BeadStudio and Microsoft Excel and are expressed as mean and standard deviation. Data comparisons between two groups were performed by two-tailed unpaired Student's *t* test, and P-values ≤ 0.05 in combination with ratios outside the interval [0.8, 1.25] were considered statistically significant.

Additional files

Additional file 1: Sequencing analysis of Alzheimer-related genes *APP, PSEN1, PSEN2.* Representative example for DNA sequencing of APP gene exon 16 of patient NFH-46, lack of mutations Lys670Asn and Met671Leu (a) and for exon 17 lack of mutation Val717Ile (b).

Additional file 2: Telomerase activity in Alzheimer donor-derived AD-iPS cells. The telomerase activity was low in the somatic fibroblast cells NFH-46 from which the two AD-iPS lines AD-iPS5 and AD-iPS26B were derived. Upon induction of pluripotency, the enzyme was reactivated in both iPS lines. Human embryonic stem cell lines H1 and H9 and the telomerase positive control cells (TPC) provided by the kit served as positive controls. The minus telomerase control (MTC, only CHAPS lysis buffer), no template control (NTC, only water) and heat inactivated cell extracts served as negative controls. The standard deviation is indicated by error bars.

Additional file 3: Alkaline phosphatase (AP) staining and DNA fingerprinting of sporadic AD-iPS cell lines. (a): The two iPS cell lines derived from the sporadic Alzheimer fibroblasts NFH-46 were positive for the pluripotency-associated alkaline phosphatase (AP) staining. Morphologies of both AD-iPSCs are shown in low and high magnification. (b): DNA fingerprinting confirmed the somatic origin of the two AD-iPS cell lines, AD-iPS5 and AD-iPS26B, and the lack of cross-contamination with hESC lines H1 and H9. The AD-iPS cell lines were derived in one reprogramming experiment.

Additional file 4: Pluripotency-associated genes are expressed in AD-iPS cells in a similar manner as in ESC line H1. Quantitative real-time PCR to analyze the expression of the most common pluripotency genes in the two generated AD-iPS lines (AD-iPS5 and AD-iPS26B) and embryonic stem cell line H1. Bars indicate the RNA level normalized to β-ACTIN first and compared to gene expression of NFH-46 (plus standard error of mean SEM; n = 3). Each AD-iPS cell line was split from one well into the three wells for expansion before the RNA was isolated. The RNA samples were not pooled. Both AD iPS cell lines were generated in one reprogramming experiment.

Additional file 5: Microarray-based gene expression profiling of AD-iPS cells, control iPS cells and related parental fibroblast cells. (a): AD-iPS cells (AD-iPS5 and AD-iPS26B), both from one well of one reprogramming experiment, cluster with control iPS cells (OiPS3 and OiPS6) and with hESCs (H1 and H9), and are far apart from AD fibroblasts (NFH-46) and control fibroblasts (NFH-2). (b): Table showing all the Pearson correlation values r^2 between all the single samples analyzed. For color coding, five distinct degrees of correlation are represented: red for $r^2 = 1$, orange for $1 < r^2 < 0.9$, yellow for $0.9 < r^2 < 0.8$, light yellow for $0.8 < r^2 < 0.75$, and grey for $r^2 < 0.75$.

Additional file 6: *In-vitro* differentiation of sporadic AD-iPS cell lines. Both lines (a) AD-iPS5 and (b) AD-iPS26B could be successfully differentiated into all three embryonic germ layers *in-vitro* through an embryoid body (EB) based differentiation approach. Indicated are the expression of marker proteins specific for ectoderm, mesoderm, and endoderm. Scale bars, 100 μm.

Additional file 7: Teratoma formation of sporadic AD-iPS cell lines. The differentiation potential of the two AD-iPS cell lines was tested *in-vivo* with the teratoma formation assay. AD-iPS26B successfully gave rise to teratoma containing derivatives of all three germ layers. For AD-iPS5 endodermal cells could not be clearly determined.

Additional file 8: Karyotype of sporadic AD-iPSCs. Karyotyping analysis of the AD-reprogrammed cells was performed. AD-iPS26B exhibited a normal female karyotype, in a similar fashion to the parental fibroblast cells NFH-46. AD-iPS5 was found to harbour next to monosomy of the X chromosome, small supernumerary marker chromosoms (circle).

Additional file 9: Pathways and biological processes significantly up-regulated in AD-iPSC neuronal cells versus neuronal cells derived from embryonic stem cells of one neuronal differentiation experiment each.

Additional file 10: Pathways and biological processes significantly down-regulated in AD-iPSC neuronal cells versus neuronal cells derived from embryonic stem cells of one neuronal differentiation experiment each.

Additional file 11: List of primers used for quantitative real-time PCR, DNA fingerprinting.

Additional file 12: List of primers used for sequencing of *APP, PSEN1, PSEN2* in sporadic AD-patient derived fibroblasts used to derive neuronal cells.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AMH: Establishment of fibroblasts from skin biopsies of the patient with Alzheimer's disease, AD-iPSC generation and characterization, neuronal differentiation and treatment with compound E, protein analysis, interpretation of results, manuscript writing. MM: AD-iPSC generation and characterization, neuronal differentiation and treatment with compound E, real-time PCR, microarray data processing, generation of protein association networks, interpretation of results, manuscript writing. AP: AD-iPSC generation, microarray data processing, manuscript editing. BL: Telomerase activity in fibroblasts and pluripotent stem cells, support in culturing of pluripotent stem cells. MRT: Mutation analysis. WW: microarray data processing, generation of heatmaps, cluster analysis, allocation of AD related genes to brain regions, generation of protein association networks. FS: UPS-related microarray and real-time PCR data processing, writing of the manuscript. PN: Supervision of mutation analysis. HK: Supervision of HLA analysis. EM: Scientific conception and interpretation, patient selection and recruitment, application for ethics committee approval, medical evaluations, surgical performance of skin biopsies, manuscript editing. CCZ: Scientific conception and interpretation, supervision of patient recruitment and fibroblast generation, clinical interpretation, manuscript revision. JA: Scientific conception and interpretation, supervision of iPSC derivation and neuronal differentiation, manuscript writing and editing. All authors read and approved the final manuscript.

Authors' information

James Adjaye, Eugenia Makrantonaki and Christos C Zouboulis are senior authors.

Acknowledgements

This study was funded by German Federal Ministry of Education and Research (BMBF) grants (01GN0807) to C.C.Z, E.M and J.A.. The authors thank Dietrich Trebing and Aikatherini I. Liakou, Departments of Dermatology, Venereology, Allergology and Immunology, Dessau Medical Center for the preparation of the skin biopsies from AD patients, Robin Ella, St. Joseph's Psychiatric Hospital Dessau for diagnostic advise, Siegfried Jorks, Institute for Transfusion Medicine Dessau for HLA typing analysis, and Anja Rabien for critical discussion of the manuscript. We thank Elisabeth Socha for providing support for the culture of ESCs and iPSCs. J.A. acknowledges partial support from the European Union funding/FP7 (FP7/2007-2013)/Grant Agreement n° 305299 (AgedBrainSYSBIO), BMBF grant (01GN1005) and from the Stiftung für Altersforschung of the Heinrich Heine University Düsseldorf.

Author details

[1]Departments of Dermatology, Venereology, Allergology and Immunology, Dessau Medical Center, 06847 Dessau, Germany. [2]Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University Duesseldorf, Moorenstr. 5, 40225 Duesseldorf, Germany. [3]Cologne Center for Genomics (CCG), Institute for Genetics, University of Cologne, 50931 Cologne, Germany. [4]Institute for Transfusion Medicine Dessau, Red Cross Blood Transfusion Service NSTOB, 06847 Dessau, Germany. [5]Molecular Embryology and Aging

Group, Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany. [6]Department of Biology, Chemistry and Pharmacy, Institute of Chemistry and Biochemistry, Freie Universität Berlin, Thielallee 63, 14195 Berlin, Germany. [7]Geriatrics Research Group, Department of Geriatric Medicine, Charité Universitätsmedizin Berlin, Reinickendorfer Str. 61, 13447 Berlin, Germany. [8]Current address: Max Delbrueck Center for Molecular Medicine (MDC), Robert Roessle Str. 10, D-13125 Berlin, Germany.

Received: 28 May 2014 Accepted: 22 January 2015
Published online: 14 February 2015

References

1. Blennow K, de Leon MJ, Zetterberg H. Alzheimer's disease. Lancet. 2006;368(9533):387–403.
2. Selkoe DJ. Alzheimer's disease results from the cerebral accumulation and cytotoxicity of amyloid beta-protein. J Alzheimers Dis. 2001;3(1):75–80.
3. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry. 2006;63(2):168–74.
4. Cruts M, van Duijn CM, Backhovens H, Van den Broeck M, Wehnert A, Serneels S, et al. Estimation of the genetic contribution of presenilin-1 and −2 mutations in a population-based study of presenile Alzheimer disease. Hum Mol Genet. 1998;7(1):43–51.
5. Selkoe DJ. Alzheimer's disease: genes, proteins, and therapy. Physiol Rev. 2001;81(2):741–66.
6. Friedhoff P, Schneider A, Mandelkow EM, Mandelkow E. Rapid assembly of Alzheimer-like paired helical filaments from microtubule-associated protein TAU monitored by fluorescence in solution. Biochemistry. 1998;37 (28):10223–30.
7. Haass C, Hung AY, Schlossmacher MG, Oltersdorf T, Teplow DB, Selkoe DJ. Normal cellular processing of the beta-amyloid precursor protein results in the secretion of the amyloid beta peptide and related molecules. Ann N Y Acad Sci. 1993;24(695):109–16.
8. Vassar R. The beta-secretase, BACE: a prime drug target for Alzheimer's disease. J Mol Neurosci. 2001;17(2):157–70.
9. Tanzi RE, Bertram L. Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. Cell. 2005;120(4):545–55.
10. Hardy J, Selkoe DJ. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. Science. 2002;297(5580):353–6.
11. Goedert M, Hasegawa M, Jakes R, Lawler S, Cuenda A, Cohen P. Phosphorylation of microtubule-associated protein tau by stress-activated protein kinases. FEBS Lett. 1997;409(1):57–62.
12. Gong CX, Iqbal K. Hyperphosphorylation of microtubule-associated protein tau: a promising therapeutic target for Alzheimer disease. Curr Med Chem. 2008;15(23):2321–8.
13. Takahashi M, Tomizawa K, Ishiguro K. Distribution of tau protein kinase I/glycogen synthase kinase–3beta, phosphatases 2A and 2B, and phosphorylated tau in the developing rat brain. Brain Res. 2000;857(1–2):193–206.
14. Grimes CA, Jope RS. The multifaceted roles of glycogen synthase kinase 3beta in cellular signaling. Prog Neurobiol. 2001;65(4):391–426.
15. Upadhya SC, Hegde AN. Role of the ubiquitin proteasome system in Alzheimer's disease. BMC Biochem. 2007;8 Suppl 1:S12.
16. Lehman NL. The ubiquitin proteasome system in neuropathology. Acta Neuropathol. 2009;118(3):329–47.
17. Hegde AN, Upadhya SC. Role of ubiquitin-proteasome-mediated proteolysis in nervous system disease. Biochim Biophys Acta. 2011;1809(2):128–40.
18. Riederer BM, Leuba G, Vernay A, Riederer IM. The role of the ubiquitin proteasome system in Alzheimer's disease. Exp Biol Med (Maywood). 2011;236(3):268–76.
19. Kloetzel PM, Soza A, Stohwasser R. The role of the proteasome system and the proteasome activator PA28 complex in the cellular immune response. Biol Chem. 1999;380(3):293–7.
20. Tanaka K. The proteasome: overview of structure and functions. Proc Jpn Acad Ser B Phys Biol Sci. 2009;85(1):12–36.
21. Ebstein F, Kloetzel P-M, Krüger E, Seifert U. Emerging roles of immunoproteasomes beyond MHC class I antigen processing. Cell Mol Life Sci. 2012;69(15):2543–58.
22. Israel MA, Yuan SH, Bardy C, Reyna SM, Mu Y, Herrera C, et al. Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. Nature. 2012;482(7384):216–20.
23. Park I-H, Arora N, Huo H, Maherali N, Ahfeldt T, Shimamura A, et al. Disease-specific induced pluripotent stem cells. Cell. 2008;134(5):877–86.
24. Shi Y, Kirwan P, Smith J, MacLean G, Orkin SH, Livesey FJ. A human stem cell model of early Alzheimer's disease pathology in Down syndrome. Sci Transl Med 2012, 4(124), doi: 10.1126/scitranslmed.3003771
25. Yagi T, Ito D, Okada Y, Akamatsu W, Nihei Y, Yoshizaki T, et al. Modeling familial Alzheimer's disease with induced pluripotent stem cells. Hum Mol Genet. 2011;20(23):4530–9.
26. Kondo T, Asai M, Tsukita K, Kutoku Y, Ohsawa Y, Sunada Y, et al. Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular Aβ and differential drug responsiveness. Cell Stem Cell. 2013;12(4):487–96.
27. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007;131(5):861–72.
28. Prigione A, Hossini AM, Lichtner B, Serin A, Fauler B, Megges M et. al. Mitochondrial-associated cell death mechanisms are reset to an embryonic-like state in aged donor-derived iPS cells harboring chromosomal aberrations. PLoS One 2011, 6(11), doi:101371/journal.pone.0027352
29. Müller F-J, Goldmann J, Löser P, Loring JF. A call to standardize teratoma assays used to define human pluripotent cell lines. Cell Stem Cell. 2010;6(5):412–4.
30. Greber B, Wu G, Bernemann C, Joo JY, Han DW, Ko K, et al. Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. Cell Stem Cell. 2010;6(3):215–26.
31. Al-Jaberi N, Lindsay S, Sarma S, Bayatti N, Clowry GJ. The Early Fetal Development of Human Neocortical GABAergic Interneurons. Cereb Cortex 2013, doi:10.1093/cercor/bht254
32. Epelbaum J, Guillou JL, Gastambide F, Hoyer D, Duron E, Viollet C. Somatostatin, Alzheimer's disease and cognition: an old story coming of age? Prog Neurobiol. 2009;89:153–61.
33. Jouhilahti E-M, Peltonen S, Peltonen J. Class III beta-tubulin is a component of the mitotic spindle in multiple cell types. J Histochem Cytochem. 2008;56(12):1113–9.
34. European Alzheimer's Disease I, Genetic, Environmental Risk in Alzheimer's D, Alzheimer's Disease Genetic C, Cohorts for H, Aging Research in Genomic E. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45(12):1452–8.
35. Matsumata M, Inada H, Osumi N. Fatty acid binding proteins and the nervous system: Their impact on mental conditions. Neurosci. Res. 2014, doi:10.1016/j.neures.2014.08.012
36. da Cruz e Silva OAB, Henriques AG, Domingues SCTS, da Cruz e Silva EF. Wnt signalling is a relevant pathway contributing to amyloid beta- peptide-mediated neuropathology in Alzheimer's disease. CNS Neurol Disord Drug Targets. 2010;9(6):720–6.
37. Li W, Tang Y, Fan Z, Meng Y, Yang G, Luo J et al.. Autophagy is involved in oligodendroglial precursor-mediated clearance of amyloid peptide. Mol Neurodegener 2013, 8, doi: 10.1186/1750-1326-8-27
38. Zhang C, Rodriguez C, Spaulding J, Aw TY, Feng J. Age-dependent and tissue-related glutathione redox status in a mouse model of Alzheimer's disease. J Alzheimers Dis. 2012;28(3):655–66.
39. Pérez-Palma E, Bustos BI, Villamán CF, Alarcón MA, Avila ME, Ugarte GD et al.. Overrepresentation of glutamate signaling in Alzheimer's disease: network-based pathway enrichment using meta-analysis of genome-wide association studies. PLoS One 2014, 9(4), doi:10.1371/journal.pone 0095413.
40. Swomley AM, Förster S, Keeney JT, Triplett J, Zhang Z, Sultana R et al. Abeta, oxidative stress in Alzheimer disease: Evidence based on proteomics studies. Biochim Biophys Acta 2013, doi: 10.1016/j.bbadis.2013.09.015
41. Wright S, Malinin NL, Powell KA, Yednock T, Rydel RE, Griswold-Prenner I. Alpha2beta1 and alphaVbeta1 integrin signaling pathways mediate amyloid-beta-induced neurotoxicity. Neurobiol Aging. 2007;28(2):226–37.
42. Guerreiro R, Brás J, Hardy J. SnapShot: genetics of Alzheimer's disease. Cell. 2013;155(4):968. e961.
43. Jiang T, Yu J-T, Tan M-S, Wang H-F, Wang Y-L, Zhu X-C, et al. Genetic variation in PICALM and Alzheimer's disease risk in Han Chinese. Neurobiol Aging. 2014;35(4):934. e931-933.
44. Morgen K, Ramirez A, Frölich L, Tost H, Plichta MM, Kölsch H et al. Genetic interaction of PICALM and APOE is associated with brain atrophy and cognitive impairment in Alzheimer's disease. Alzheimers Dement 2014, doi:10.1016/j.jalz.2013.11.001

45. Shang H, Fu J, Zhang X-M, Song R-R, Wang W-Z. Association between EXOC3L2 rs597668 polymorphism and Alzheimer's disease. CNS Neurosci Ther. 2013;19(10):834–9.
46. Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet. 2011;43(5):436–41.
47. Schmidt C, Wolff M, von Ahsen N, Zerr I. Alzheimer's disease: genetic polymorphisms and rate of decline. Dement Geriatr Cogn Disord. 2012;33(2–3):84–9.
48. Bertram L, Tanzi RE. Genome-wide association studies in Alzheimer's disease. Hum Mol Genet. 2009;18(R2):R137–45.
49. Mulligan VK, Chakrabartty A. Protein misfolding in the late-onset neurodegenerative diseases: common themes and the unique case of amyotrophic lateral sclerosis. Proteins. 2013;81(8):1285–303.
50. Lee GJ, Lu PH, Medina LD, Rodriguez-Agudelo Y, Melchor S, Coppola G, et al. Regional brain volume differences in symptomatic and presymptomatic carriers of familial Alzheimer's disease mutations. J Neurol Neurosurg Psychiatr. 2013;84(2):154–62.
51. Bryant C, Giovanello KS, Ibrahim JG, Chang J, Shen D, Peterson BS et al. Mapping the genetic variation of regional brain volumes as explained by all common SNPs from the ADNI study. PLoS One 2013, 8(8), doi:10.1371/journal.pone.0071723.
52. Hostage CA, Choudhury KR, Murali Doraiswamy P, Petrella JR, Alzheimer's Disease Neuroimaging Initiative. Mapping the effect of the apolipoprotein E genotype on 4-year atrophy rates in an Alzheimer disease-related brain network. Radiology. 2014;271(1):211–9.
53. Nelson PT, Abner EL, Schmitt FA, Kryscio RJ, Jicha GA, Santacruz K, et al. Brains with medial temporal lobe neurofibrillary tangles but no neuritic amyloid plaques are a diagnostic dilemma but may have pathogenetic aspects distinct from Alzheimer disease. J Neuropathol Exp Neurol. 2009;68(7):774–84.
54. Rhinn H, Fujita R, Qiang L, Cheng R, Lee JH, Abeliovich A. Integrative genomics identifies APOE ε4 effectors in Alzheimer's disease. Nature. 2013;500(7460):45–50.
55. Zhang B, Gaiteri C, Bodea L-G, Wang Z, McElwee J, Podtelezhnikov AA, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell. 2013;153(3):707–20.
56. Keller JN, Hanni KB, Markesbery WR. Impaired proteasome function in Alzheimer's disease. J Neurochem. 2000;75(1):436–9.
57. Mishto M, Bellavista E, Santoro A, Stolzing A, Ligorio C, Nacmias B, et al. Immunoproteasome and LMP2 polymorphism in aged and Alzheimer's disease brains. Neurobiol Aging. 2006;27(1):54–66.
58. Nijholt DAT, de Graaf TR, van Haastert ES, Oliveira AO, Berkers CR, Zwart R, et al. Endoplasmic reticulum stress activates autophagy but not the proteasome in neuronal cells: implications for Alzheimer's disease. Cell Death Differ. 2011;18(6):1071–81.
59. Dil Kuazi A, Kito K, Abe Y, Shin R-W, Kamitani T, Ueda N. NEDD8 protein is involved in ubiquitinated inclusion bodies. J Pathol. 2003;199(2):259–66.
60. Mori F, Nishie M, Piao YS, Kito K, Kamitani T, Takahashi H, et al. Accumulation of NEDD8 in neuronal and glial inclusions of neurodegenerative disorders. Neuropathol Appl Neurobiol. 2005;31(1):53–61.
61. Müller F-J, Schuldt BM, Williams R, Mason D, Altun G, Papapetrou EP, et al. A bioinformatic assay for pluripotency in human cells. Nat Methods. 2011;8(4):315–7.
62. Prigione A, Lichtner B, Kuhl H, Struys EA, Wamelink M, Lehrach H, et al. Human induced pluripotent stem cells harbor homoplasmic and heteroplasmic mitochondrial DNA mutations while maintaining human embryonic stem cell-like metabolic reprogramming. Stem Cells. 2011;29(9):1338–48.
63. Prigione A, Fauler B, Lurz R, Lehrach H, Adjaye J. The senescence-related mitochondrial/oxidative stress pathway is repressed in human induced pluripotent stem cells. Stem Cells. 2010;28(4):721–33.
64. Fecker LF, Schmude M, Jost S, Hossini AM, Picó AH, Wang X, et al. Efficient and selective tumor cell lysis and induction of apoptosis in melanoma cells by a conditional replication-competent CD95L adenovirus. Exp Dermatol. 2010;19(8):e56–66.
65. DAVID Bioinformatics Resources 6.7. [http://david.abcc.ncifcrf.gov/]
66. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.
67. STRING v9.1. [http://string-db.org/]
68. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013;41(D1):D808–15.

2.2.2 Footprint-free human fetal foreskin derived iPSCs: A tool for modeling hepatogenesis associated gene regulatory networks.

Sci Rep. 2017 Jul 24;7(1):6294. doi: 10.1038/s41598-017-06546-9.

Matz P(1)(2)(3), Wruck W(2), Fauler B(1), Herebian D(4), Mielke T(1), Adjaye J(5)(6).

Author information: (1)Max Planck Institute for Molecular Genetics, 14195, Berlin, Germany. (2)Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, 40225, Düsseldorf, Germany. (3)Institute of Biology, Humboldt University of Berlin, 10099, Berlin, Germany. (4)Department of General Pediatrics, Neonatology and Pediatric Cardiology, Heinrich Heine University, 40225, Düsseldorf, Germany. (5)Max Planck Institute for Molecular Genetics, 14195, Berlin, Germany. James.Adjaye@med.uni-duesseldorf.de. (6)Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, 40225, Düsseldorf, Germany. James.Adjaye@med.uni-duesseldorf.de.

Induced pluripotent stem cells (iPSCs) are similar to embryonic stem cells and can be generated from somatic cells. We have generated episomal plasmid-based and integration-free iPSCs (E-iPSCs) from human fetal foreskin fibroblast cells (HFF1). We used an E-iPSC-line to model hepatogenesis in vitro. The HLCs were characterized biochemically, i.e. glycogen storage, ICG uptake and release, UREA and bile acid production, as well as CYP3A4 activity. Ultra-structure analysis by electron microscopy revealed the presence of lipid and glycogen storage, tight junctions and bile canaliculi- all typical features of hepatocytes. Furthermore, the transcriptome of undifferentiated E-iPSC, DE, HE and HLCs were compared to that of fetal liver and primary human hepatocytes (PHH). K-means clustering identified 100 clusters which include developmental stage-specific groups of genes, e.g. OCT4 expression at the undifferentiated stage, SOX17 marking the DE stage, DLK and HNF6 the HE stage, HNF4α and Albumin is specific to HLCs, fetal liver and adult liver (PHH) stage. We use E-iPSCs for modeling gene regulatory networks associated with human hepatogenesis and gastrulation in general.

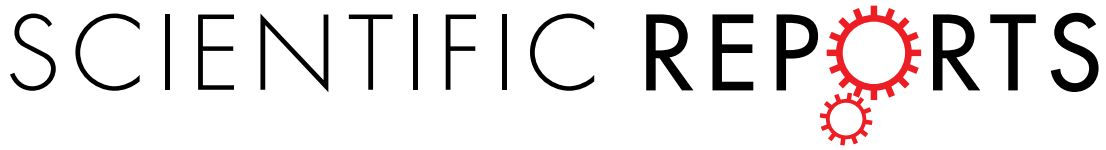
DOI: 10.1038/s41598-017-06546-9 PMCID: PMC5524812 PMID: 28740077 [Indexed for MEDLINE]

URL: https://www.nature.com/articles/s41598-017-06546-9

The data presented in this publication has been used also for the PhD thesis of Dr. Peggy Matz at the Humboldt University of Berlin.

Contribution of Wasco Wruck: Bioinformatic analysis, interpretation and manuscript writing.

www.nature.com/scientificreports



OPEN

Footprint-free human fetal foreskin derived iPSCs: A tool for modeling hepatogenesis associated gene regulatory networks

Received: 20 August 2015
Accepted: 13 June 2017
Published online: 24 July 2017

Peggy Matz[1,2,3], Wasco Wruck[2], Beatrix Fauler[1], Diran Herebian[4], Thorsten Mielke[1] & James Adjaye[1,2]

Induced pluripotent stem cells (iPSCs) are similar to embryonic stem cells and can be generated from somatic cells. We have generated episomal plasmid-based and integration-free iPSCs (E-iPSCs) from human fetal foreskin fibroblast cells (HFF1). We used an E-iPSC-line to model hepatogenesis *in vitro*. The HLCs were characterized biochemically, i.e. glycogen storage, ICG uptake and release, UREA and bile acid production, as well as CYP3A4 activity. Ultra-structure analysis by electron microscopy revealed the presence of lipid and glycogen storage, tight junctions and bile canaliculi- all typical features of hepatocytes. Furthermore, the transcriptome of undifferentiated E-iPSC, DE, HE and HLCs were compared to that of fetal liver and primary human hepatocytes (PHH). K-means clustering identified 100 clusters which include developmental stage-specific groups of genes, e.g. OCT4 expression at the undifferentiated stage, SOX17 marking the DE stage, DLK and HNF6 the HE stage, HNF4α and Albumin is specific to HLCs, fetal liver and adult liver (PHH) stage. We use E-iPSCs for modeling gene regulatory networks associated with human hepatogenesis and gastrulation in general.

Human embryonic stem cells (hESCs) derived from inner cell mass cells of the blastocyst undergo symmetric self-renewal and are pluripotent i.e. can give rise to all cells within the three embryonic germ layers-endoderm, ectoderm, mesoderm and also germ cells[1]. Human Induced pluripotent stem cells (iPSCs) were initially derived from dermal fibroblasts by viral transduction mediated over-expression of four embryonic transcription factors OCT4, SOX2, KLF4 and c-MYC or OCT4, SOX2, NANOG and LIN28[2,3]. Human iPSCs share similar properties with hESCs, however, the integration of pro-viruses into the host genome of viral-derived iPSC is a risk factor for clinical applications in the future[4,5]. To overcome these drawbacks, non-viral reprogramming methods have been described using non-integrating Sendai viruses, episomal-based plasmid vectors, *in vitro*-derived mRNA and miRNA[5–11].

The liver is the largest internal organ and hepatocytes are the main functional cells in the liver. Hepatocytes perform a number of complex functions which are essential for life e.g. production of plasma proteins, synthesis of bile acids, the uptake and storage of glucose as well as drug detoxification. The use of primary human hepatocytes (PHH) is problematic, first, they cannot be expanded *in vitro* and second, they are difficult to obtain routinely or in sufficient quantities[12,13]. Alternatives such as human hepatocarcinoma-derived and transformed, permanent cell lines, including HepG2, THLE and HepaRG, have phenotypes significantly diverged from normal primary hepatocytes[14–16]. A potential alternative could be the differentiation of iPSCs to hepatocyte-like cells. Hepatocyte-like cells (HLCs) generated from human iPSC have shown great promise as an inexhaustible source of cells that mirror the genotype of the donor to satisfy this need. Several groups have already shown how multi-functional applicable HLCs generated from iPSC can serve as cellular models for drug screening and toxicology studies, as a source for disease modeling[17–19].

[1]Max Planck Institute for Molecular Genetics, 14195, Berlin, Germany. [2]Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, 40225, Düsseldorf, Germany. [3]Institute of Biology, Humboldt University of Berlin, 10099, Berlin, Germany. [4]Department of General Pediatrics, Neonatology and Pediatric Cardiology, Heinrich Heine University, 40225, Düsseldorf, Germany. Correspondence and requests for materials should be addressed to J.A. (email: James.Adjaye@med.uni-duesseldorf.de)

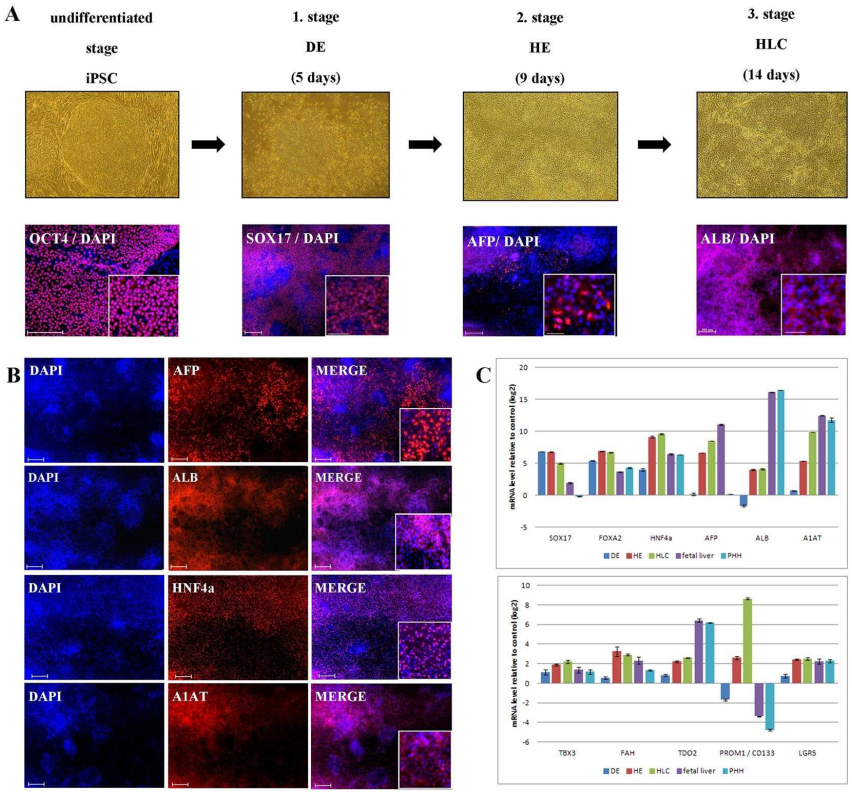


Figure 1. Derivation of hepatocyte-like cells (HLC) from E-iPSCs. (**A**) First row phase contrast images of the differentiation stages, from undifferentiated stage the episomal induced pluripotent stem cells (E-iPSCs) to definitive endoderm (DE), then hepatic endoderm (HE) and finally hepatocyte-like cells (HLCs). Second row immunofluorescence-based staining of stage specific proteins overlapped with DAPI (staining of nucleus). Scale bar: 100 μm Alexa Flour 594 (red). (**B**) Immunofluorescence-based staining of HLC specific proteins AFP, ALB, HNF4α and A1AT. Scale bar: 100 μm Alexa Flour 594 (red). (**C**) Expression patterns of liver specific marker genes during HLC differentiation compared to fetal liver and primary human hepatocytes (PHH) are shown by quantitative real-time PCR (qPCR). Three biological replicates in technical triplicates of each sample were analyzed. The data were normalized to E-iPSCs. The standard deviation is depicted by the error bars.

To date most studies use viral-derived iPSC to generate HLC, however these have drawbacks (e.g. genome integration). Additionally, most studies focus on one aspect of the multifunctional application of HLCs-derived from iPSCs such as the generation of HLCs from iPSCs, the maturation, modeling liver diseases, drug screening and toxicology[17–24].

In this study we used an integration-free, episomal-derived induced pluripotent stem cell line (E-iPSC) from human neonatal foreskin fibroblast (HFF1)[25,26] to derive and characterize hepatocyte-like cells (HLCs) as well as untangle human hepatogenesis-associated gene regulatory networks.

Results

Differentiation of E-iPSCs to hepatocyte-like cells (HLCs). We derived an integration-free, episomal-based induced pluripotent stem cell line (E-iPSC) from human neonatal foreskin fibroblast (HFF1)[25] to derive HLCs for this study. The derivation of HLCs was based on a slight modification of the protocol described by Sullivan *et al.*[18]. The differentiation to HLCs consists of three steps (Fig. 1A). First, the cells were differentiated towards definitive endoderm (DE) resulting in down-regulated expression of the pluripotent markers *OCT4*, *SOX2* and *NANOG* and activation of DE specific markers such as *SOX17*. The second step in the protocol resulted in the emergence of hepatic endoderm cells (HE) as defined by the expression of hepatoblasts markers such as *AFP*, *PROM1* and *LGR5* (Fig. 1B,C). Finally, the HE cells were forced into maturation resulting in HLCs expressing mature liver markers such as *ALB*, *A1AT*, *FOXA2*, *HNF4α TBX3*, *FAH* and *TDO2* but still maintaining *AFP*

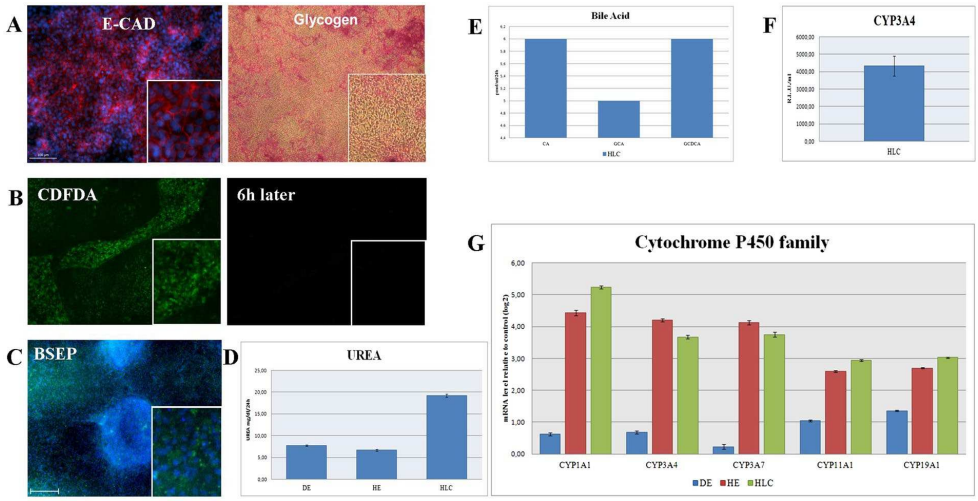


Figure 2. Functional analysis of hepatocyte-like cells (HLCs) derived from E-iPSCs. (**A**) E-Cadherin (E-CAD) antibody staining marking cell shape (left panel), Glycogen storage (right panel), Periodic Acid-Schiff (PAS) assay was used. Glycogen storage is indicated by pink or dark red-purple cytoplasm. (**B**) Visualization of 5 (and 6)-Carboxy-2′,7′-dichlorofluorescein diacetate (CDFDA), immunofluorescence image of HLCs direct after incubation with CDFDA (left panel), and immunofluorescence image of HLCs six hours later (right panel). (**C**) Immunofluorescence-based protein staining of bile salt export pump (BSEP). (**D**) Analysis of UREA production in E-iPSC-DE (DE), E-iPSC-HE (HE) and E-iPSC-HLCs (HLC). Three biological replicates in technical triplicates of each sample were analyzed. The levels of urea are presented as a percentage, considering measured levels of urea in mg/dL/24 h. The error bars indicate the standard errors of the mean. (**E**) Measurement of bile acid secretion of E-iPSC-HLCs (HLC). (**F**) Measurement of CYP3A4 secretion of HLC samples. Three biological replicates in technical triplicates were analyzed. The levels of CYP3A4 are presented as relative light units per milliliter (R.L.U./ml). The error bars indicate the standard errors of the mean. (**G**) Quantitative real-time PCR (qPCR) analysis of cytochrome P450 family member activity of all stages E-iPSC-DE (DE), E-iPSC-HE (HE) and E-iPSC-HLC (HLC) are shown. Three biological replicates in technical triplicates of each sample were analyzed. The data were normalized to E-iPSCs. The standard deviation is depicted by the error bars.

expression (Fig. 1B,C)[20,27–31]. Quantitative real-time PCR confirmed the results of the immunofluorescence-based detection of protein expression (Fig. 1C).

Functional analyses of E-iPSC derived HLCs. The HLCs bear hallmarks of primary hepatocytes, i.e. (i) cobblestone-shaped epithelial cells expressing E-Cadherin (E-CAD), (ii) the ability to store glycogen as confirmed by Periodic-Acid-Schiff (PAS) staining (Fig. 2A). (iii) Uptake and release of ICG (data not shown) and CDFDA were measured (Fig. 2B). (iv) BSEP was detactable in HLCs by immunofluorescence-based protein staining (Fig. 2C). (v) Urea secretion was measured in all three stages of the differentiation protocol, as anticipated the highest level of production was in HLCs (Fig. 2D). (vi) Bile acids (CA, GCA, GCDCA) were also produced by HLCs (Fig. 2E and Supplementary Figure S1D). (vii) CYP3A4 activity was measured (Fig. 2F). Quantitative real-time PCR and heatmap-based analysis confirmed the expression of CYP3A4 as well as other members of the cytochrome P450 super family of enzymes (Fig. 2G and Supplementary Figure S1A). (viii) Electron microscopy revealed the ultra-structure typical of hepatocytes such as bile canaliculi with microvilli, lipid storage and tight junctions (Fig. 3A). (ix) Bi-nucleated cells could be shown by bright field microscopy (Fig. 3B). Finally, the efficiency of HLC differentiation was scored by HNF4α expression, as well as a double staining of ALBUMIN and HNF4α (Fig. 3C,D).

Hepatogenesis associated transcriptional road map. A cluster dendrogram and accompanying correlation co-efficients demonstrates high similarities between replicates. Furthermore, fetal liver and PHH formed a cluster and iPSCs, HE and DE formed a cluster which is then extended by HLCs (Fig. 4A). K-means clustering identified 100 clusters which include developmental stage-specific groups of genes, e.g. OCT4 expression at the undifferentiated stage, SOX17 marking DE stage, HNF6 at the HE stage, PROX1 at the HLC stage, AFP marking the fetal liver stage and ALB marking the mature liver (PHH) stage (Fig. 4B and Supplementary Figure S1B). We further identified upstream regulators of genes within the six selected clusters shown in Fig. 4B by transcription factor over-representation analysis via the oPOSSUM data base[32]. These gene regulatory networks associated with these clusters are presented in Supplementary Figure 3. The network for iPSCs (Supplementary

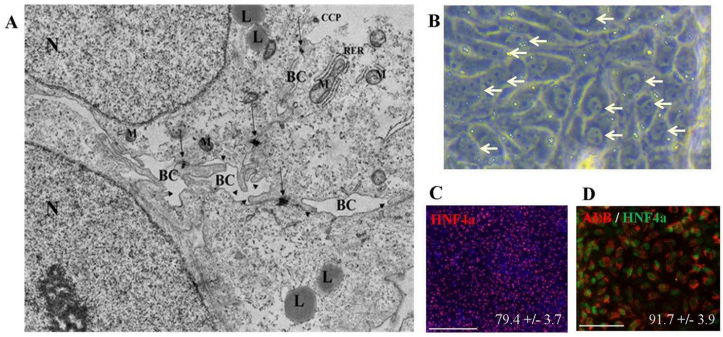


Figure 3. Ultra structure and maturation. (**A**) Electron microscopy image of E-iPSC-HLCs. BC = bile canaliculi; L = lipid; CCP = clathrin coated pits; M = mitochondrion; RER = rough endoplasmic reticulum; N = nucleus; arrow = tight junctions; arrowhead = microvilli. (**B**) Bright field microscopy of HLCs. Bi-nuleated cells are marked by arrow-heads. (**C**) Immunofluorescence-based staining of HNF4α in end-stage HLCs. 79.4% (+/−3.7%) of the cells counted positive for HNF4α. Scale bar: 200 μm Alexa Flour 594 (red). (**D**) Immunofluorescence-based double staining of Albumin (ALB) and HNF4α in end-stage HLCs. 91.7% (+/− 3.9) of HNF4α positive cells were also positive for ALB. Scale bar: 100 μm, Alexa Flour 594 (red) and Alexa Flour 488 (green).

Figure S4A) shows the well known regulatory relations between OCT4 (POU5F1), SOX2, NANOG, KLF4. Most significant factors from the oPOSSUM analysis were STAT1, MZF1 and KLF4 (Z-Score > 10). In the network for DE (Supplementary Figure S4B) SP1, INSM1, MZF1, KLF4, REST are most significant (Z-Score >= 10), in HE (Supplementary Figure S4C) LHX3, MIZF, CTCF, in HLC (Supplementary Figure S4D), PLAG1, EWSR1-FLI1, IRF2, in fetal liver (Supplementary Figure S4E) TAL1::GATA1, HNF1A, ZFN143, GATA1, HNF1B and in primary human hepatocytes (Supplementary Figure S4F) HNF1A, CTCF, ZFX, HNF4A, FOXA2, FOXA1, CEBPA.

The microarray data and quantitative real-time PCR confirmed the expression profile of the listed genes. KRT17 and CXCR4 marking the definitive endoderm stage (DE), ANXA1, TTR and TBX3 represents the hepatocyte-like cell stage (HLC), AFP marking the fetal liver stage and ALB represents the mature liver (PHH) stage (Fig. 4C). Figure 4D shows a heatmap of the top 30 most abundantly expressed genes in each differentiation stage (k-means cluster). In order to assign a tissue-type to the HLCs we applied a tissue prediction tool - KeyGenes[33] (Fig. 4E and Supplementary Table S1), which confirmed HLC (cluster 9) as liver, comparable to fetal liver and PHH. Furthermore, venn diagram analysis shows the numbers of significantly differentially expressed genes between HLCs, fetal liver and PHHs (Fig. 5A). The HLC-related genes *ANXA1*, *TTR* and *TBX3* as well as the fetal liver-related gene *AFP* and *ALB* representing the matured liver stage PHH are located in the intersection of all three samples and are included in the 11506 genes (Fig. 5A and Supplementary Table S2). A closer look into the exclusive expressed genes in HLCs (1808 genes) uncovered tight junction-specific genes such as *CLDN9*, *CLDN18*, *OCLN*, *PARD6A* and *PARD6B* (Fig. 5B and Supplementary Table S3). One step further, a venn diagram was generated from HLC vs. fetal liver, HLC vs. PHH and fetal liver vs. PHH (Fig. 5C and Supplementary Table S4). DAVID analysis of Hippo signaling related genes from the intersection of HLC vs. PHH and HLC vs. fetal liver dedicated the activity of cell-cell contact related pathways, adherent and tight junction pathway. A chart of these Hippo pathway related genes underlines the predominant expression of cell-cell contact related genes in HLCs (Fig. 5D and Supplementary Table S5). The Hippo pathway, which is responsible for maturation and stabilization of the tight junctions in hepatocytes, and ABC transporters, which are accountable for the uptake and efflux of e.g. bile acids and metabolites, are over represented in HLC, fetal liver and PHH. Bile acid related transporter genes such as *NTCP*, *MRP2*, *ASBT* and *MDR2/3* are highly expressed in HLC compared to the DE and HE stage (Fig. 5D and Supplementary Figure S1C–E).

A venn diagram of genes which were expressed in DE, HE and HLC shows the relation between the stages of the HLC differentiation (Fig. 6A and Supplementary Table S6). DAVID analysis based on these genes uncover genes which are related to pathways that define the functionality of the liver such as drug metabolism, metabolism of xenobiotics and fatty acid metabolism (Supplementary Figure S2).

Cell fate determination: hepatocytes or cholangiocytes? To date most studies on liver cell fate decision making have been conducted in mice[34,35]. This type of analyses can now be conducted in human using iPSCs. To demonstrate this we generated a heatmap consisting of key genes co- or differentially- expressed in undifferentiated iPSC, DE, HE and HLCs (Fig. 6B and Supplementary Figure S3 and Supplementary Table S7). Progenitor-related genes such as *HNF1A* and *HNF1B* are expressed in HE and HLCs whereas *PROX1* is expressed exclusively in HLCs. Hepatocyte-specific genes such as *ALB*, *AFP*, *ABCB4* and *CYP3A7* are expressed in the HLC samples. Additionally, there exist a group of hepatocyte-related genes which are expressed in HLCs and HE, e.g. *ABCC2*, *RARB* and *TRR*. *WNT3A* as a marker for cholangiocytes as well as *SOX9* and *KRT7* are expressed in both HE and HLCs, whereas *AQP1* and *DLK1* are expressed exclusively in HLCs (Fig. 6A,B). To analyze the

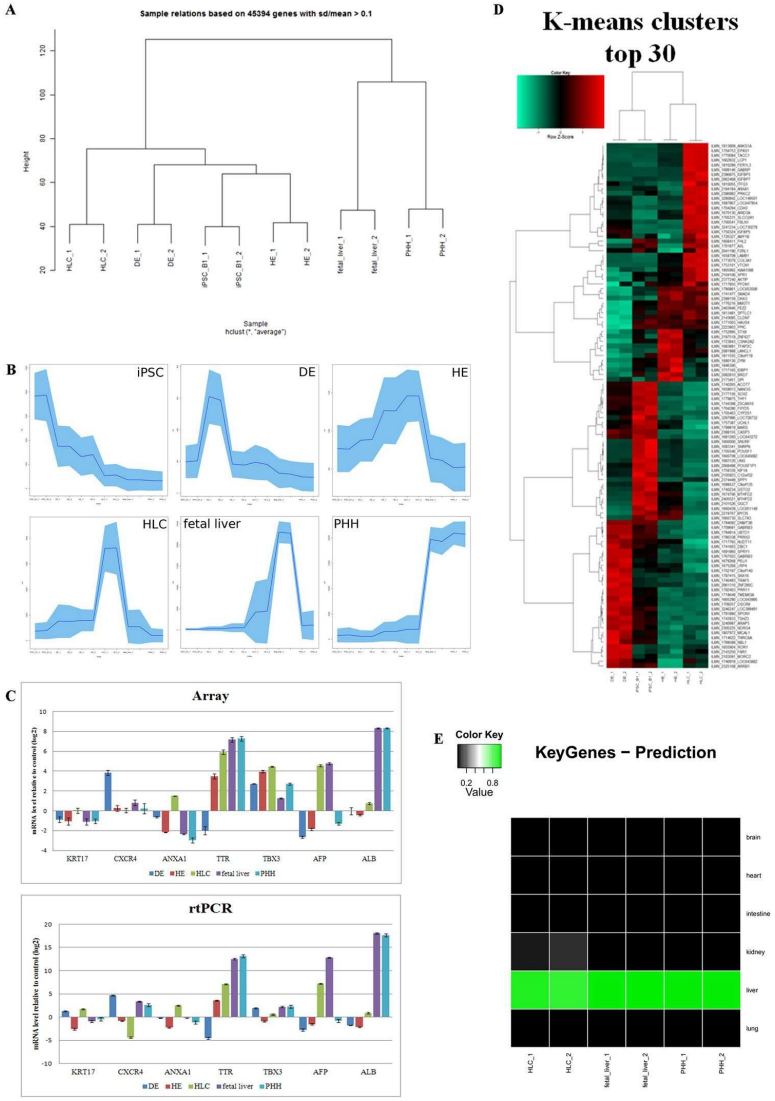


Figure 4. Transcriptional dynamic of hepatocyte-like cells (HLCs) derived from E-iPSCs. (**A**) A cluster dendrogram similarities between replicates and relationship of the samples E-iPSC (iPSC_B1), DE, HE, HLC, fetal liver and PHH. (**B**) K-means cluster 68 contained OCT4 marking undifferentiated stage (iPSC), cluster 81 (sub-cluster 1) contained SOX17 marking DE stage (DE), cluster 37 (sub-cluster 2) represents HE stage (HE), cluster 51 represents HLC stage (HLC), cluster 72 contained AFP marking fetal liver stage (fetal liver) and cluster 91 contained liver marker ALB (PHH). (**C**) Confirmation of microarray data by quantitative real-time (qPCR). On the left hand the array expression data and on the right hand the qPCR expression data of the following genes are shown: *KRT17* and *CXCR4* marking the definitive endoderm stage (DE), *ANXA1*, *TTR* and *TBX3* represents the hepatocyte-like cell stage (HLC), *AFP* marking the fetal liver stage and *ALB* represents the mature liver (PHH) stage. Three biological replicates in technical triplicates of each sample were analyzed. The data were normalized to E-iPSCs. The standard deviation is depicted by the error bars. (**D**) Heatmap of the top 30 genes from each K-means clusters in Fig. 4B. (**E**) KeyGenes prediction for k-means Hepatocyte-like-cell (HLC) cluster9. Data sets for human liver, brain, intestine, kidney, lung and heart were downloaded from NCBI GEO and the KeyGenes tool was employed to generate a training set for these Illumina microarray platform data. As the test set genes from the HLC cluster 9 and HLC, fetal liver and primary human hepatocyte samples were used.

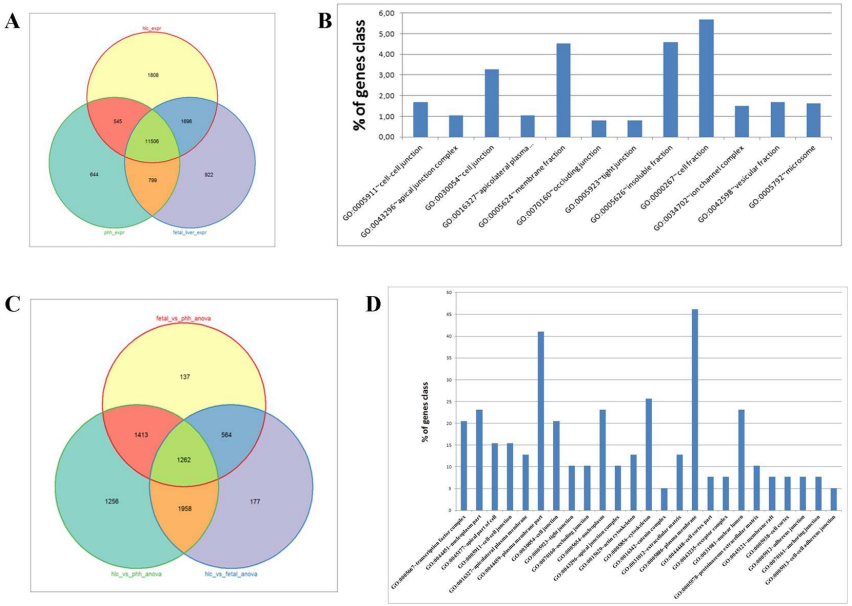


Figure 5. Comparative transcriptome profile analyses. (**A**) Venn diagram of HLCs, fetal liver and PHH. (**B**) GO cellular components of HLC exclusively expressed genes (Fisher extract $p < 0.01$). (**C**) Venn diagram of fetal liver vs. PHH anova, HLC vs. fetal liver anova and HLC vs. PHH anova. (**D**) GO cellular components of HIPPO pathway genes exclusively expressed genes (Fisher extract $p < 0.01$) in the intersection of HLC vs. fetal liver anova and HLC vs. PHH anova (intersection with 1958 genes).

relation between these genes for cell fate decision a transcription factor network was created by the use of the data base oPOSSUM[32] (Supplementary Figure S4). To make it more manageable transcription factors with a Z-score >10 (green circles) and downstream regulating transcription of the progenitor genes (red circles) are shown (Fig. 6C and Supplementary Figure S5). Myc, HNF1A, SP1, MZF1 5–13, HNF4A, Klf4 were the most significant upstream transcription factors (Z-score > 10, $p < 1e\text{-}15$). This demonstrates the dedicated functionalities of this core regulatory network: HNF1A and HNF4A determine the liver fate by regulating ALB and AFP while KLF4 regulates SOX17. Bipotential markers DLK1 and NOTCH3 are both regulated by MYC, Klf4, SP1 and MZF1 5–13. All cholangiocyte marker genes are regulated by KLF4, SP1 and MZF1. *SOX9* is only regulated by these three, while *ONECUT1*, *ONECUT2* and *SALL4* are regulated by these three and additionally by MYC and HNF4A. Furthermore, MYC was the most prominent transcription factor in this network (Z-score of 12.54) and regulates *PROX1* and the transcription factor HNF1A, which is also regulated by HNF4A, as well as the cholangiocyte-related gene *WNT3A* (Fig. 6C, Supplementary Table S8)[28,36,37].

Discussion

In this study we used an episomal-derived and integration-free iPSC line to model hepatogenesis *in vitro*. Yu *et al.*[6,26] showed that iPSCs generated using episomal-based plasmids are free of vector and transgene sequences as we have also shown[25]. A whole-genome sequencing of iPSCs, which were generated by episomal vectors based on the EBNA1/OriP episomal replicon, showed (i) lack of integration of the episomal vector DNA in the host genome, (ii) loss of the episomal vectors in the iPSCs and (iii) no visible changes in the genomes of the iPSCs[26,38,39]. The episomal approach is a reliable method for iPSC derivation[40].

Hepatocytes are the main cell type supporting the detoxification function of the liver and as such they are already extensively used for toxicology screens. Our episomal-derived, viral- and integration-free iPSC line is able to differentiate into hepatocyte-like cells (HLCs) which have similar functional properties as liver-biopsy derived primary human hepatocyes (Figs 1, 2, 3 and 4). These cells can be used for (i) toxicology and drug screening, (ii) future application in tissue replacement therapies, (iii) modeling human diseases *in vitro*.

An important pathway for hepatogenesis is the Hippo-signaling pathway which influences liver cell fate and size[41,42]. This signal transduction pathway is crucial for early embryonic development, embryonic and adult stem cells, cell proliferation, differentiation, apoptosis, organ size, specific functions in adult organs and tumorgenesis[41,43–46]. The Hippo pathway is responsible for maturation and stabilization of the tight junctions in hepatocytes[47] and is over represented in HLCs, fetal liver and PHH. The existence of tight junctions in E-iPSC-derived HLCs is shown by electron microscopy (Figs 3A and 5D). ABC transporters, which are accountable for the uptake and efflux of e.g. bile acids, are also over represented in HLCs, fetal liver and PHH (Fig. 4D). Bile acid measurement shows

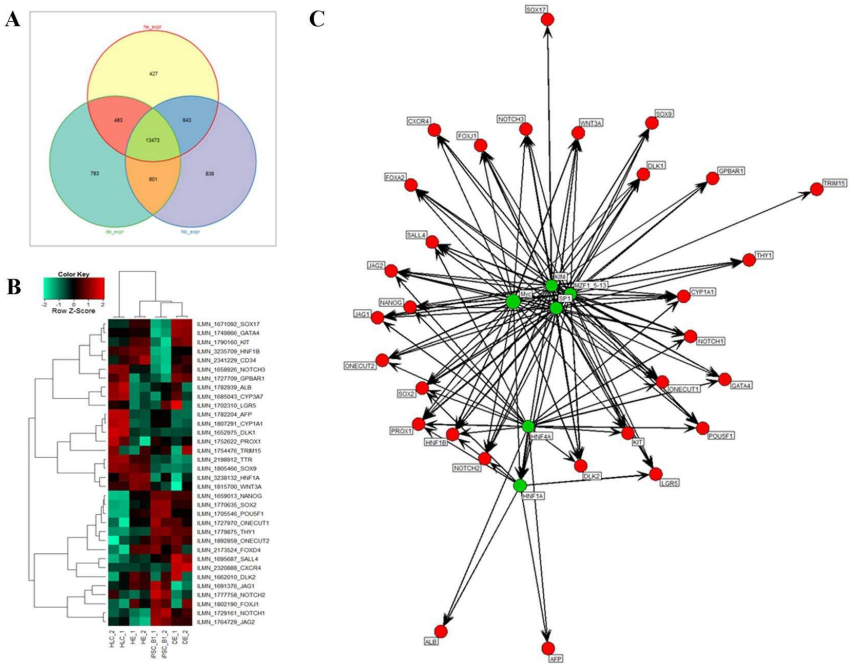


Figure 6. Cell fate decision. (**A**) Venn diagram of DE, HE and HLC. (**B**) Heatmap of bipotential progenitor-associated, hepatocyte-related and cholangiocyte-associated genes. (**C**) An induced network of transcription factors with a Z-score >10 (green, size of circle corresponds to Z-score) and associated genes (red).

the excretion of primary bile acids in HLCs. Primary bile acids are involved in drug metabolism and synthesis of cholesterol, steroids and other lipids[48]. Hepatocytes synthesize primary bile acids which carried out by the gut microbiota convert to secondary bile acids by several reactions including dehydroxylation, dehydrogenation and epimerization[49]. A heatmap of bile acid related transporter genes underlines the functionality of the HLCs and shows the highest expression of *NTCP*, *MRP2*, *ASBT* and *MDR2/3* in the HLC samples (Fig. 4D).

A cluster dendrogram shows the highest similarities between replicates. Furthermore, fetal liver and PHH formed a cluster and iPSCs, HE and DE formed a cluster which is then extended by HLCs (Fig. 4A). Differences and commonalities between HLCs, fetal liver and PHH are assessed by statistical tests vs. iPSCs and pairwise statistical tests of the three experiments vs. each other. Most genes (4953) are in the intersection set common to all three experiments. Pairwise intersections fetal liver/PHH have 2823 genes, HLCs/PHH 1699 genes and HLCs/fetal liver 988 genes (Supplementary Table S9).

To assess the differences between HLCs, fetal liver and PHH an ANOVA followed by pairwise t-tests was performed. Venn diagram shows the numbers of significantly differentially expressed genes in these pairwise t-tests after filtering for the most variable genes found in the ANOVA (Fig. 5A).

Transcription factor binding site analysis revealed transcription factors are over-represented in these clusters. Top transcription factors by z-score were STAT1, MZF1, KLF4, SP1 and IRF1 for iPSCs, ELF5, FEV, INSM1, FOXI1 and STAT1 for DE, LHX3, MIZF, CTCF, NR3C1 and PAX6 for HE, PLAG1, EWSR1-FLI1, IRF2, MEF2A and ELF5 for HLCs, TAL1:GATA1, HNF1A, ZNF143, GATA1 and HNF1B for fetal liver, HNF1A, CTCF, HNF4A and FOXA2 for PHH (data not shown).

The cluster of iPSCs, definite endoderm (DE) and hepatic endoderm (HE) is successively extended by hepatocyte-like cells (HLCs), fetal liver and primary human hepatocytes (PHH). In most cases fetal liver and PHH are located in one cluster as well as iPSCs, DE and HE are formed another cluster. HLCs mostly are located in between the fetal liver/PHH and the early differentiation stage cluster (Fig. 4A).

Liver specific analysis of genes which are exclusively expressed in HLCs by venn diagram, PaGenBase and DAVID uncover the functionality of the HLC derived from E-iPSCs by showing the activity of cell-cell contact related pathways as well as liver-specific metabolism pathways such as drug metabolism, metabolism of xenobiotics and fatty acid metabolism (Figs 5 and 6 and Supplementary Figure S2).

We demonstrated the feasibility of using iPSCs as *in vitro* models for studying liver cell fate decision making. The heatmap presented in Fig. 6B shows differential expression of key cell fate regulating genes specific to the hepatic endoderm and in some cases also in the HLCs. This implies that the HLCs also harbor cell populations with bipotential properties similar to hepatoblasts. For example, the hepatic endoderm cells express *DLK2* and

also a set of transcription factors specific to this stage (Onecut1/HNF-6) which are not expressed in HLCs. These set of transcription factors are putative candidates for directing biliary epithelial cells/cholangiocytes cell fate. However, HLCs express *DLK1* and progenitor-related genes such as *PROX1* and *LGR5* which are not expressed in HEs. Genes such as *WNT3A*, *NOTCH3*, *HNF1A*, *HNF1B* and *SOX9* are expressed in both stages (Fig. 6B). Our overall findings based on gene expression patterns supports the notion that hepatic endoderm (HE) and HLCs in our hepatocyte differentiation protocol are equivalent to the DLK, HNF6 and SOX9-positive bipotential hepatoblasts present in fetal liver and are common progenitors for hepatocytes and biliary epithelial cells/cholangiocytes (Fig. 6B)[34,35].

The Notch signaling, hepatocyte nuclear factor-6 (HNF-6) and PROX1 are amongst factors known to regulate lineage commitment in the bipotential hepatoblast progenitor cell population[35,50,51]. It should be possible to change the fate of HE cells by manipulating the expression levels of e.g. *PROX1*, *SOX9*, and *HNF6* or even by using small molecules targeting for instance Notch signaling.

Odom *et al.*[52] described a core transcriptional regulatory circle in human hepatocytes consisting of six transcription factors (ONECUT1/HNF-6, FOXA2, HNF1A, HNF4A, CREB1 and USF1). These transcription factors bound promoters that are central for liver development and function. For this they used a mixture of human hepatocytes from multiple healthy donors to maximize the diversity of gender and age[52]. Our study shows the expression of the transcription factors *ONECUT1/HNF-6*, *FOXA2*, *HNF1A*, *HNF4A* in human HLCs derived from human E-iPSCs. Furthermore, we have demonstrated that amongst other transcription factors, HNF1A and HNF4A are involved in orchestrating cell fate decision of bipotential hepatoblast cells to become either hepatocytes or biliary epithelial cells/cholangiocytes (Fig. 6C).

The generation of a transcription factor network via the oPOSSUM data base[32] uncovered transcription factors which are involved in cell fate decisions during hepatogenesis. The most prominent transcription factor in our network is MYC -transcription factors which regulates numerous biological processes such as glycolysis, cell proliferation and differentiation[53,54]. MYC also regulates the expression of bipotential hepatoblast-related genes (e.g. *DLK1*, *PROX1*), cholangicyte-related genes (e.g. *ONECUT1*, *WNT3A* and *SALL4*) as well as hepatocyte-related genes (e.g. *HNF1A*). The second prominent transcription factor is HNF4A- a central regulator of hepatocyte differentiation and function[55] and regulates hepatocyte-related genes (e.g. *ALB*, *AFP*), cholangiocyte-related genes (e.g. *ONECUT1*, *SALL4*) as well as the hepatoblast-related gene *PROX1* (Fig. 6C, Supplementary Table S8)[28,36,37]. his transcription network underlines the bipotential progenitor-related characteristics of the HE and HLC stage in our differentiation procedure. This implies that there are bipotential progenitors within HE and HLC cell populations.

K-means clustering identified developmental stage-specific groups of genes, e.g. OCT4 expression at the undifferentiated stage, SOX17 marking the DE stage, DLK and HNF6 the HE stage, HNF4A and Albumin is specific to HLCs, fetal liver and adult liver (PHH) stage gain an insight into hepatogenesis. Furthermore, gene regulatory networks generated by oPOSSUM data base uncovered the presence of bipotential progenitor populations in both stages in HE and HLC. This analysis should lay the foundation for future efforts to generate long-term cultures of cholangiocytes and HLCs.

The bipotential progenitor population in iPSC-derived HLCs and the presence of AFP expression underline their fetal status. However, we uncovered human hepatogenesis-associated gene regulatory networks which are involved in cell fate decision making during hepatogenesis.

Conclusion

In summary, we have demonstrated the derivation of integration-free E-iPSCs from somatic cells and differentiated these to hepatocyte-like cells (HLCs) capable of storing glycogen, ICG uptake and release, UREA and bile acid production, as well as CYP3A4 activity. Ultra-structure analysis by electron microscopy revealed the presence of lipid and glycogen storage, tight junctions and bile canaliculi- all typical features of biopsy-derived primary hepatocytes. Model organisms such as zebrafish or mouse are used in order to analyze developmental processes of hepatogenesis[34,35,56]. HLCs derived from human E-iPSCs can be used to analyze the human hepatogenesis in details. We uncovered a gene regulatory network which uncovered the presence of bipotential progenitor populations in HE and HLC stage. Additionally, MYC was identified as a prominent regulator of bipotential hepatoblast-related genes expression.

Methods

Cell Culture and Differentiation. Human neonatal foreskin fibroblasts HFF1 were purchased from ATCC (HFF1 #SCRC-1041) and were maintained in Dulbecco's modified Eagle medium (DMEM™, Gibco) containing 10% fetal bovine serum (FBS™, Invitrogen) and 0.5% penicillin and streptomycin (Invitrogen). Human ES and iPS cells were maintained on irradiated mouse embryonic fibroblast (MEF) cells in KnockOut™ DMEM (Invitrogen) supplemented with 20% KnockOut™ Serum Replacement (Invitrogen), 0.1 mM non-essential amino acids (Invitrogen), 0.1 mM L-glutamine (Invitrogen), 0.1 mM ß-Mercaptoethanol (Sigma), 0.5% penicillin and streptomycin and 8ng/ml basic fibroblast growth factor (bFGF, Invitrogen) as described by Wolfrum *et al.*[57]. The human ESC lines H1 and H9 were purchased from WiCell Research Institute (Madison, WI, USA, www.wicell.org, #WA01 and #WA09).

All used cells and cell lines were cultured at 37 °C and 5% CO2 in an incubator (INNOVA CO-170 Incubator, New Brunswick Scientific) under humidified atmosphere. All treatments and maintenance procedures were carried out using a clean bench type HeraSafe (Haereus Instruments).

For differentiation iPSC into HLC the protocol from Sullivan *et al.*[18] was used. We modified the last step of HLC generation. During the last step of differentiation we used 25ng/ml dexamethasone instead of 10 μM hydrocortisone 21-hemisuccinate.

Functional Assays for HLC. *PAS staining.* Glycogen storage was identified by Periodic Acid-Schiff (PAS) Staining System (Sigma-Aldrich). Cells were fixed with 4% paraformaldehyde for 15 min and stained according to the manufacturer's instructions.

Uptake and release. To detect the uptake and release of substances ICG (indocyanine green; Cardiogreen, ICG; Sigma) and CDFDA (5 (and 6)-Carboxy-2′,7′-dichlorofluorescein diacetate, CDFDA; Sigma) were used. The cells were incubated in culture medium with freshly diluted ICG (1 mg/ml) or CDFDA (1 μM) for 30 min. at 37 °C. The cells were washed with PBS, fresh culture medium was added and uptake of dye was documented. The release of ICG and CDFDA was examined after 6 h. The results of ICG and PAS assays were examined under an Olympus CK2 phase-contrast microscope and representative morphology was recorded at a magnification of ×50 using a Canon 300D digital camera. The fluorophore of CDFDA was visualized using a Zeiss, LSM 510 Meta confocal microscope with a connected camera for microscopy model AxioCam ICc3 and the software Axiovision 4.6 at a magnification of ×200.

Urea measurement. Urea secretion was quantified by a colorimetric assay QuantiChrom™ Urea Assay Kit (DIUR-500 BioAssay Systems) following the manufacturer's instructions. The assay detects urea directly by using substrates that specifically bind urea. Urea assays were carried out in 96-well plates, and concentrations were measured using a plate reader. Analysis of urea production in E-iPSC-DE (DE), E-iPSC-HE (HE) and E-iPSC-HLC (HLC) were performed. Three biological replicates in technical triplicates of each sample were analyzed. The levels of urea are presented as a percentage, considering measured levels of urea in mg/dL/24 h.

Bile acid measurement. Bile acids (cholic acid CA, chenodeoxycholic acid CDCA, deoxycholic acid DCA, ursodeoxycholic acid UDCA, lithocholic acid LCA) including their glycine- and taurine derivatives were analyzed by UPLC-MS/MS. The system consisted of an Acquity UPLC-H Class (Waters, UK) coupled to a Xevo-TQS tandem mass spectrometer (Waters, UK) which is equipped with an ESI source operating in the negative ion mode. Quantitative data were conducted in the multiple reaction monitoring (MRM) mode. The chromatographic separation was performed on Waters UPLC BEH C18 column (100 mm, 2.1 mm ID, 1.7 μm; Waters, UK) using acetonitrile and acidic water (0.1% formic acid) as mobile phases. Analytes were separated by a gradient elution. The injection volume was 5 μL and the column was maintained at 40 °C.

CYP3A4 measurement. CYP3A4 activity was measured by using the pGlo kit (Promega) according to manufacturer's instruction for nonlytic CYP450 activity estimation. The CYP3A4 production was measured in E-iPSC-HLC (HLC). Three biological replicates in technical triplicates of each sample were analyzed. The levels of CYP3A4 are presented as relative light units per milliliter (R.L.U./ml). The error bars indicate the standard errors of the mean.

Electron microscopy. Cells grown on Thermanox® plastic coverslips (Nunc), were fixed in a modified Karnofsky solution, 2%PFA/2,5%GA in 50 mM Cacodylate buffer, pH7.4 at 4 °C. Cells were washed in 50 mM Cacodylatpuffer/50 mM NaCl and post-fixed for 90 min at room temperature with 0,5% OsO_4 in the same buffer. After washing steps with water, cells were incubated for 40 min with 0,1% tannic acid in 250 mM Hepes pH7,4, washed with water and stained with 2% uranyl acetate, 90 min at room temperature. Cells were dehydrated in a graded series of ethanol and embedded in Spurr's resin (Low Viscosity Spurr Kit, Ted Pella, CA, USA). Ultra-thin sections (70 nm) were prepared with an ultramicrotome (Reichert Ultracut E, Leica) and mounted on pioloform-coated slot grids from copper. Sections were counterstained with uranyl acetate and lead citrate.

Ultrathin-sections were first examined using a Philips CM100 transmission electron microscope operated at 100 kV and finally imaged using a FEI Tecnai Spirit transmission electron microscope operated at 120 kV, which was equipped with a 2 k × 2 k Eagle CCD camera (FEI). The MSI-Raster application within the Leginon Software package[58] was used to automatically image selected regions of interest at a final nominal magnification of 15000× applying a defocus of −4 μm. Raw micrographs were stitched using the Trakem2 plugin implemented in the Fiji software platform[59,60].

Microarray -Based Gene Expression Analysis. Total RNA from iPSC, DE, HE, HLC, PHH and fetal liver in replicates were extracted using the MiniRNeasy Kit (Qiagen) according to the manufacturer's instructions and quality checked by Nanodrop analysis (Nanodrop Technologies, Wilmington, DE, USA, http://www.nanodrop.com). Approximately 500 ng of DNase treated RNA was sent to ATLAS Biolabs (http://www.atlas-biolabs.de) for whole transcriptome analysis by using microarray. All basic expression data analysis was carried out using the BeadStudio software 3.0. Raw data were background-subtracted and normalized using the "rank invariant" algorithm and then filtered for significant expression on the basis of negative control beads. For correlation coefficient analysis and the generation of Venn diagrams, detected gene expression was defined by a Detection P Value < 0.01 as output by BeadStudio. For differential gene expression analyses, genes had to be at least 1.5 fold up- or down-regulated in a group-wise comparison, to be considered significantly differentially expressed.

For more detailed information see supplementary Materials and Methods.

Data access. GEO Submission (GSE66282).

References

1. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science.* **282**, 1145–1147 (1998).
2. Takahashi, K., Okita, K., Nakagawa, M. & Yamanaka, S. Induction of pluripotent stem cells from fibroblast cultures. *Nat Protoc.* **2**, 3081–3089 (2007).
3. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science.* **318**, 1917–1920 (2007).

4. Chen, L. & Liu, L. Current progress and prospects of induced pluripotent stem cells. *Sci China C Life Sci.* **52**, 622–636 (2009).
5. Drews, K., Jozefczuk, J., Prigione, A. & Adjaye, J. Human induced pluripotent stem cells–from mechanisms to clinical applications. *J Mol Med (Berl).* **90**, 735–745 (2012).
6. Yu, J. *et al.* Human induced pluripotent stem cells free of vector and transgene sequences. *Science.* **324**, 797–801 (2009).
7. Zhou, H. *et al.* Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell.* **4**, 381–384 (2009).
8. Kiskinis, E. & Eggan, K. Progress toward the clinical application of patient-specific pluripotent stem cells. *J Clin Invest.* **120**, 51–59 (2010).
9. Plews, J. R. *et al.* Activation of pluripotency genes in human fibroblast cells by a novel mRNA based approach. *PLoS One.* **5**, e14397 (2010).
10. Warren, L. *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell.* **7**, 618–630 (2010).
11. Miyoshi, N. *et al.* Reprogramming of mouse and human cells to pluripotency using mature microRNAs. *Cell Stem Cell.* **8**, 633–638 (2011).
12. Elaut, G. *et al.* Molecular mechanisms underlying the dedifferentiation process of isolated hepatocytes and their cultures. *Curr Drug Metab.* **7**, 629–660 (2006).
13. Fraczek, J. *et al.* Primary hepatocyte cultures for pharmaco-toxicological studies: at the busy crossroad of various anti-dedifferentiation strategies. *Arch Toxicol.* **87**, 577–610 (2013).
14. Hart, S. N. *et al.* A comparison of whole genome gene expression profiles of HepaRG cells and HepG2 cells to primary human hepatocytes and human liver tissues. *Drug Metab Dispos.* **38**, 988–994 (2010).
15. Gerets, H. H. J. *et al.* Characterization of primary human hepatocytes, HepG2 cells, and HepaRG cells at the mRNA level and CYP activity in response to inducers and their predictivity for the detection of human hepatotoxins. *Cell Biol Toxicol.* **28**, 69–87 (2012).
16. Jetten, M. J. A. *et al.* Baseline and genotoxic compound induced gene expression profiles in HepG2 and HepaRG compared to primary human hepatocytes. *Toxicol In Vitro.* **27**, 2031–2040 (2013).
17. Rashid, S. T. *et al.* Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. *J Clin Invest.* **120**, 3127–3136 (2010).
18. Sullivan, G. J. *et al.* Generation of functional human hepatic endoderm from human induced pluripotent stem cells. *Hepatology.* **51**, 329–335 (2010).
19. Yusa, K. *et al.* Targeted gene correction of 1-antitrypsin deficiency in induced pluripotent stem cells. *Nature.* **478**, 391–394 (2011).
20. Jozefczuk, J., Prigione, A., Chavez, L. & Adjaye, J. Comparative analysis of human embryonic stem cell and induced pluripotent stem cell-derived hepatocyte-like cells reveals current drawbacks and possible strategies for improved differentiation. *Stem Cells Dev.* **20**, 1259–1275 (2011).
21. Hannan, N. R. F., Segeritz, C. P., Touboul, T. & Vallier, L. Production of hepatocyte-like cells from human pluripotent stem cells. *Nat Protoc.* **8**, 430–437 (2013).
22. Takebe, T. *et al.* Generation of a vascularized and functional human liver from an iPSC-derived organ bud transplant. *Nat Protoc.* **9**, 396–409 (2014).
23. Greenhough, S., Medine, C. N. & Hay, D. C. Pluripotent stem cell derived hepatocyte like cells and their potential in toxicity screening. *Toxicology.* **278**, 250–255 (2010).
24. Takayama, K. *et al.* Prediction of interindividual differences in hepatic functions and drug sensitivity by using human iPS-derived hepatocytes. *Proc Natl Acad Sci USA* **111**, 16772–16777 (2014).
25. Matz, P. & Adjaye, J. Episomal-based generation of an iPS cell line from human fetal foreskin fibroblasts. *Stem Cell Research.* **16**, 67–69 (2016).
26. Yu, J. *et al.* Efficient feeder-free episomal reprogramming with small molecules. *PLoS One.* **6**, e17557 (2011).
27. Agarwal, S., Holton, K. L. & Lanza, R. Efficient differentiation of functional hepatocytes from human embryonic stem cells. *Stem Cells.* **26**, 1117–1127 (2008).
28. Hay, D. C. *et al.* Efficient differentiation of hepatocytes from human embryonic stem cells exhibiting markers recapitulating liver development *in vivo*. *Stem Cells.* **26**, 894–902 (2008).
29. Schmidt, S. K. *et al.* Antimicrobial and immunoregulatory properties of human tryptophan 2,3-dioxygenase. *Eur J Immunol.* **39**, 2755–2764 (2009).
30. Lüdtke, T. H. W., Christoffels, V. M., Petry, M. & Kispert, A. Tbx3 promotes liver bud expansion during mouse development by suppression of cholangiocyte differentiation. *Hepatology.* **49**, 969–978 (2009).
31. Teckman, J. H. & Jain, A. Advances in alpha-1-antitrypsin deficiency liver disease. *Curr Gastroenterol Rep.* **16**, 367 (2014).
32. Kwon, A. T., Arenillas, D. J., Worsley Hunt, R. & Wasserman, W. W. oPOSSUM-3: advanced analysis of regulatory motif over-representation across genes or ChIP-Seq datasets. *G3 (Bethesda).* **2**, 987–1002 (2012).
33. Roost, M. S. *et al.* KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas. *Stem Cell Rep.* **4**, 1112–1124 (2015).
34. Tanimizu, N. *et al.* Isolation of hepatoblasts based on the expression of Dlk/Pref-1. *J Cell Sci.* **116**, 1775–1786 (2003).
35. Vanderpool, C. *et al.* Genetic interactions between hepatocyte nuclear factor-6 and Notch signaling regulate mouse intrahepatic bile duct development *in vivo*. *Hepatology.* **55**, 233–243 (2012).
36. Dianat, N. *et al.* Generation of functional cholangiocyte-like cells from human pluripotent stem cells and HepaRG cells. *Hepatology.* **60**, 700–714 (2014).
37. Seth, A. *et al.* Prox1 ablation in hepatic progenitors causes defective hepatocyte specification and increases biliary cell commitment. *Development.* **141**, 538–547 (2014).
38. Mayshar, Y. *et al.* Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell.* **7**, 521–531 (2010).
39. Cheng, X. *et al.* Self-renewing endodermal progenitor lines generated from human pluripotent stem cells. *Cell Stem Cell.* **10**, 371–384 (2012).
40. Schlaeger, T. M. *et al.* A comparison of non-integrating reprogramming methods. *Nat Biotechnol.* **33**, 58–63 (2015).
41. Camargo, F. D. *et al.* YAP1 increases organ size and expands undifferentiated progenitor cells. *Curr Biol.* **17**, 2054–2060 (2007).
42. Yimlamai, D. *et al.* Hippo pathway activity influences liver cell fate. *Cell.* **157**, 1324–1338 (2014).
43. Pan, D. Hippo signaling in organ size control. *Genes Dev.* **21**, 886–897 (2007).
44. Zhao, B. *et al.* Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. *Genes Dev.* **21**, 2747–2761 (2007).
45. Tremblay, A. M. & Camargo, F. D. Hippo signaling in mammalian stem cells. *Semin Cell Dev Biol.* **23**, 818–826 (2012).
46. Li, P. *et al.* Functional role of Mst1/Mst2 in embryonic stem cell differentiation. *PLoS One.* **8**, e79867 (2013).
47. Varelas, X. *et al.* The Crumbs complex couples cell density sensing to Hippo-dependent control of the TGF-β-SMAD pathway. *Dev Cell.* **19**, 831–844 (2010).
48. Lamba, V., Ghodke, Y., Guan, W. & Tracy, T. S. microRNA-34a is associated with expression of key hepatic transcription factors and cytochromes P450. *Biochem Biophys Res Commun. Mar* **445**, 404–411 (2014).
49. Hofmann, A. F. & Hagey, L. R. Key discoveries in bile acid chemistry and biology and their clinical applications: history of the last eight decades. *J Lipid Res.* **55**, 1553–1595 (2014).

50. Clotman, F. *et al.* The onecut transcription factor HNF6 is required for normal development of the biliary tract. *Development.* **129**, 1819–1828 (2002).
51. Dudas, J. *et al.* The homeobox transcription factor Prox1 is highly conserved in embryonic hepatoblasts and in adult and transformed hepatocytes, but is absent from bile duct epithelium. *Anat Embryol (Berl).* **208**, 359–366 (2004).
52. Odom, D. T. *et al.* Core transcriptional regulatory circuitry in human hepatocytes. *Mol Syst Biol.* **2**, 2006–2017 (2006).
53. Cai, Q., Lin, T., Kamarajugadda, S. & Lu, J. Regulation of glycolysis and the Warburg effect by estrogen-related receptors. *Oncogene.* **32**, 2079–2086 (2013).
54. Qu, A. *et al.* Role of Myc in hepatocellular proliferation and hepatocarcinogenesis. *J Hepatol.* **60**, 331–338 (2014).
55. Watt, A. J., Garrison, W. D. & Duncan, S. A. HNF4: a central regulator of hepatocyte differentiation and function. *Hepatology.* **37**, 1249–1253 (2003).
56. Gao, M. *et al.* Hepassocin is required for hepatic outgrowth during zebrafish hepatogenesis. *Biochem Biophys Res Commun.* **463**, 466–471 (2015).
57. Wolfrum, K. *et al.* The LARGE principle of cellular reprogramming: lost, acquired and retained gene expression in foreskin and amniotic fluid-derived human iPS cells. *PLoS One.* **5**, e13703 (2010).
58. Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J Struct Biol.* **151**, 41–60 (2005).
59. Saalfeld, S., Cardona, A., Hartenstein, V. & Tomančák, P. As-rigid-as-possible mosaicking and serial section registration of large ssTEM datasets. *Bioinformatics.* **26**, i57–i63 (2010).
60. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods.* **9**, 676–682 (2012).

Acknowledgements

J.A. acknowledges support from the German Federal Ministry of Education and Research (BMBF) grants (01GN0807) and the medical faculty of Heinrich Heine University, Düsseldorf.

Author Contributions

Peggy Matz: Conception and design, Collecting and assembly of data, Data analysis and interpretation, Manuscript writing. Wasco Wruck: Bioinformatic analysis, interpretation and manuscript writing. Beatrix Fauler: Electron microscopy. Diran Herebian: Bile acid measurements. Thorsten Mielke: Electron microscopy. James Adjaye: Conception, interpretation, editing and final approval of the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-06546-9

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2.2.3 Modeling Nonalcoholic Fatty Liver Disease with Human Pluripotent Stem Cell-Derived Immature Hepatocyte-Like Cells Reveals Activation of PLIN2 and Confirms Regulatory Functions of Peroxisome Proliferator-Activated Receptor Alpha.

Stem Cells Dev. 2016 Aug 1;25(15):1119-33. doi: 10.1089/scd.2015.0383. Epub 2016 Jul 15.

Graffmann N(1), Ring S(1), Kawala MA(1), Wruck W(1), Ncube A(1), Trompeter HI(2), Adjaye J(1).

Author information: (1)1 Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University Düsseldorf , Düsseldorf, Germany . (2)2 Institute for Transplantation Diagnostics and Cell Therapeutics, Heinrich Heine University Düsseldorf , Düsseldorf, Germany .

Nonalcoholic fatty liver disease (NAFLD/steatosis) is a metabolic disease characterized by the incorporation of fat into hepatocytes. In this study, we developed an in vitro model for NAFLD based on hepatocyte-like cells (HLCs) differentiated from human pluripotent stem cells. We induced fat storage in these HLCs and detected major expression changes of metabolism-associated genes, as well as an overall reduction of liver-related microRNAs. We observed an upregulation of the lipid droplet coating protein Perilipin 2 (PLIN2), as well as of numerous genes of the peroxisome proliferator-activated receptor (PPAR) pathway, which constitutes a regulatory hub for metabolic processes. Interference with PLIN2 and PPARα resulted in major alterations in gene expression, especially affecting lipid, glucose, and purine metabolism. Our model recapitulates many metabolic changes that are characteristic for NAFLD. It permits the dissection of disease-promoting molecular pathways and allows us to investigate the influences of distinct genetic backgrounds on disease progression.

DOI: 10.1089/scd.2015.0383 PMCID: PMC4971413 PMID: 27308945 [Indexed for MEDLINE]

URL: https://www.liebertpub.com/doi/full/10.1089/scd.2015.0383

Contribution of Wasco Wruck: Bioinformatic analysis, interpretation and manuscript writing.

STEM CELLS AND DEVELOPMENT
Volume 25, Number 15, 2016
Mary Ann Liebert, Inc.
DOI: 10.1089/scd.2015.0383

Modeling Nonalcoholic Fatty Liver Disease with Human Pluripotent Stem Cell-Derived Immature Hepatocyte-Like Cells Reveals Activation of PLIN2 and Confirms Regulatory Functions of Peroxisome Proliferator-Activated Receptor Alpha

Nina Graffmann,[1] Sarah Ring,[1] Marie-Ann Kawala,[1] Wasco Wruck,[1] Audrey Ncube,[1] Hans-Ingo Trompeter,[2] and James Adjaye[1]

Nonalcoholic fatty liver disease (NAFLD/steatosis) is a metabolic disease characterized by the incorporation of fat into hepatocytes. In this study, we developed an in vitro model for NAFLD based on hepatocyte-like cells (HLCs) differentiated from human pluripotent stem cells. We induced fat storage in these HLCs and detected major expression changes of metabolism-associated genes, as well as an overall reduction of liver-related microRNAs. We observed an upregulation of the lipid droplet coating protein Perilipin 2 (PLIN2), as well as of numerous genes of the peroxisome proliferator-activated receptor (PPAR) pathway, which constitutes a regulatory hub for metabolic processes. Interference with PLIN2 and PPARα resulted in major alterations in gene expression, especially affecting lipid, glucose, and purine metabolism. Our model recapitulates many metabolic changes that are characteristic for NAFLD. It permits the dissection of disease-promoting molecular pathways and allows us to investigate the influences of distinct genetic backgrounds on disease progression.

Introduction

Nonalcoholic fatty liver disease (NAFLD) is a widespread disease in the western hemisphere. Due to a high-fat diet and a lack of exercise, hepatocytes of NAFLD patients accumulate fat in the form of lipid droplets (LDs) [1]. This is often associated with type 2 diabetes and considered part of the metabolic syndrome [1]. Insulin resistance and obesity-associated chronic inflammation of adipose tissue are critical factors for the development and progression of NAFLD [2,3]. This is often seen as a ''first hit'' manifesting in the rather benign accumulation of LDs, called steatosis. A ''second hit'', frequently due to an increase of reactive oxygen species-mediated stress, induces the progression toward nonalcoholic steatohepatitis (NASH), which is accompanied by liver inflammation and fibrosis [3]. Approximately, 29% of patients with NASH develop cirrhosis. Up to 27% of these further develop hepatocellular carcinoma [1].

Hepatocytes store triacylglycerides (TAGs) in LDs as a reaction to an overload with free fatty acids. These are either derived directly from the diet or result from inflammation induced lipolysis in adipose tissues [2]. The occurrence of LDs in >5% of hepatocytes is the main diagnostic criterion for NAFLD [1].

In LDs, TAGs are enclosed by a lipid monolayer, which is encapsulated by distinct proteins, predominantly from the PAT (Perilipin/ADRP/TIP47) family [4–6]. Perilipins regulate hydrolysis of TAGs by controlling the activity of lipases and their access to LDs [7–9]. Perilipin 2 (PLIN2 or Adipophilin, ADRP) is ubiquitously expressed and plays a major role in the formation of LDs [10–12]. PLIN2 expression correlates with LD content in hepatocytes [13]. A reduction of PLIN2 expression with antisense oligonucleotides reduced liver TAG content and decreased the expression of genes involved in fatty acid and steroid metabolism in mice [14,15]. In addition, PLIN2 knockout mice develop neither obesity nor NAFLD when fed a high-fat diet because they have a higher energy turnover compared to their wild-type counterparts [16].

Nutrition and energy uptake are important factors for the development of NAFLD. However, there exist major differences between humans and mice. Various established diets reproduce effects of NAFLD/NASH in mice. Unfortunately, they fail to mirror the whole spectrum of symptoms observed in humans. While high-fat diets induce obesity and NAFLD,

[1]Institute for Stem Cell Research and Regenerative Medicine and [2]Institute for Transplantation Diagnostics and Cell Therapeutics, Heinrich Heine University Düsseldorf, Düsseldorf, Germany.

mice generally do not proceed toward NASH even if the diet is supplemented with fructose. To induce NASH, mice are usually fed with a methionine–choline-deficient diet. A major drawback of this diet, however, is the fact that mice do not become obese, which is a major risk-factor for NAFLD in humans [17,18]. In addition, there exist several knockout mouse models, none of which is capable of reflecting all aspects of the disease [17].

Several groups have used human hepatocarcinoma cell lines or immortalized primary hepatocytes to model NAFLD [19,20]. However, cancer-derived cell lines are of limited use for dissecting the molecular basis of NAFLD as they harbor genomic and hence functional aberrations compared to healthy primary liver cells [21,22]. The use of liver biopsy-derived primary human hepatocytes for modeling NAFLD is also limited because they can only be cultivated for a few days before the onset of dedifferentiation [23] or have to be immortalized by virus-mediated transduction with SV40. In addition, liver biopsies, especially those from the early stages, are very rare.

To overcome these limitations, we in this study aimed at dissecting the molecular basis of NAFLD using hepatocyte-like cells (HLCs), which were in vitro derived from human pluripotent stem cells (hPSCs). We used the human embryonic stem cell (ESC) line H1, as well as induced pluripotent stem cells (iPSCs), derived from fetal foreskin fibroblasts of a healthy individual [24,25]. We were able to monitor the accumulation of fat in the HLCs, as well as major biochemical alterations concerning lipid, glucose, and purine metabolism. Our new model system is suitable for the analysis of disease triggering factors, as well as new therapeutics.

Material and Methods

Cell culture

HepG2 cells (ATCC®HB-8065™) were cultured in DMEM low glucose with 10% FCS, 1% Penicillin/Streptomycin, and 1% GlutaMAX (Gibco). For fat induction, cells were induced with 50 μM oleic acid (OA) (Stock solution 100 mM in ethanol). As control, cells were treated with the corresponding amount of ethanol. Fat induction was performed 24 h after passaging.

Differentiation of hPSCs into HLCs

hPSCs were cultured on Matrigel (Corning) coated plates in TeSR E8 medium (STEMCELL Technologies). Medium was changed daily and spontaneously differentiated cells were removed manually before splitting the cells. One or two days after passaging, differentiation into definitive endoderm was induced with definitive endoderm medium: 96% RPMI 1640, 2% B27 (without retinoic acid), 1% GlutaMAX (Glx), 1% Penicillin/Streptomycin (P/S) (all Gibco), 100 ng/mL Activin A (Peprotech), and for the first 3 days, 50 ng/mL WNT3A (R&D). After 5 days, medium was changed toward hepatic endoderm medium as follows: 78% Knockout DMEM, 20% Knockout serum replacement, 0.5% Glx, 1% P/S, 0.01% 2-Mercaptoethanol (all Gibco), and 1% DMSO (Sigma).

After 5 days, differentiation was continued with HLC medium as follows: 82% Leibovitz 15 medium, 8% fetal calf serum, 8% Tryptose Phosphate Broth, 1% Glx, 1% P/S (all Gibco) with 1 μM Insulin (Sigma), 10 ng/mL hepatocyte growth factor (HGF) (Peprotech), 20 ng/mL Oncostatin M (OSM) 209 a.a. (Immunotools), and 25 ng/mL Dexamethasone (DEX) (Sigma) (Fig. 1A). During the whole differentiation period, medium was changed daily.

Fat induction in HLCs was performed on day 12 of the differentiation process. Interference with peroxisome proliferator-activated receptor alpha (PPARα) activity was performed by treatment with 50 μM Fenofibrate (agonist) or 2 μM GW6471 (antagonist, both from Cayman Chemical) in parallel with OA induction (Fig. 1C).

Liver-specific biochemical assays of HLCs

The amount of urea produced by the cells over a period of 24 h was determined from the cell culture supernatant using the QuantiChrom™ Urea Assay Kit (BioAssay Systems) according to the manufacturer's recommendations. Cytochrome p450 3A4, 3A5, and 3A7 activity was measured with the P450-Glo™ CYP3A4 Assay Luciferin-PFBE (Promega) using a luminometer (Lumat LB 9507; Berthold Technologies). The presence of active transporter proteins was assessed by the uptake and release of Indocyanine Green dye. Cells were incubated for 30 min with 1 mg/mL Cardiogreen (Santa Cruz Biotechnology, Inc.). Afterward, they were washed with PBS and images were captured with a light microscope (Primo Vert; Zeiss). Subsequently, cells were cultured in their usual medium for 6 h and images were again captured.

Staining of LDs

Paraformaldehyde-fixed cells were incubated for 20 min with either a 60% working solution of Oil Red O (Sigma) or with BODIPY 493/503 (1 μg/mL; Life technologies) in PBS/0.05% Tween. After washing, images were captured with a light microscope (Primo Vert; Zeiss) or a fluorescence microscope, respectively (LSM700; Zeiss).

Immunocytochemistry

For intracellular antibody staining, paraformaldehyde-fixed cells were permeabilized and unspecific binding sites were blocked by incubating for 2 h at room temperature with blocking buffer (1×PBS with 10% normal goat or donkey serum, 1% BSA, 0.5% Triton, and 0.05% Tween). Afterward, blocking buffer was diluted 1:2 with 1×PBS and cells were incubated with the primary antibody overnight at 4°C. Cells were washed thrice with 1×PBS/0.05% Tween and incubated with the secondary antibody, for 2 h at room temperature. Cells were washed as above and images captured using a fluorescence microscope (LSM700; Zeiss).

Extracellular stainings were performed in the same manner without detergents. The following primary antibodies were used: Alpha Fetoprotein, Albumin (Sigma) E Cadherin (CST), HNF4α (Abcam), SOX17 (R&D), and PLIN2 (Proteintech). For details on antibodies, see Supplementary Table S1; Supplementary Data are available online at www.liebertpub.com/scd. DNA was stained with Hoechst 33342. Individual channel images were processed and merged with Photoshop CS6.

Western blot

Cells were lysed in 1×RIPA buffer (50 mM Tris HCl, pH 8, 150 mM NaCl, 1% IGEPAL (NP-40), 0.1% SDS, 1 mM EDTA, and 0.5% Na-Deoxycholate) with protease inhibitors. Twenty microgram of protein was analyzed by western

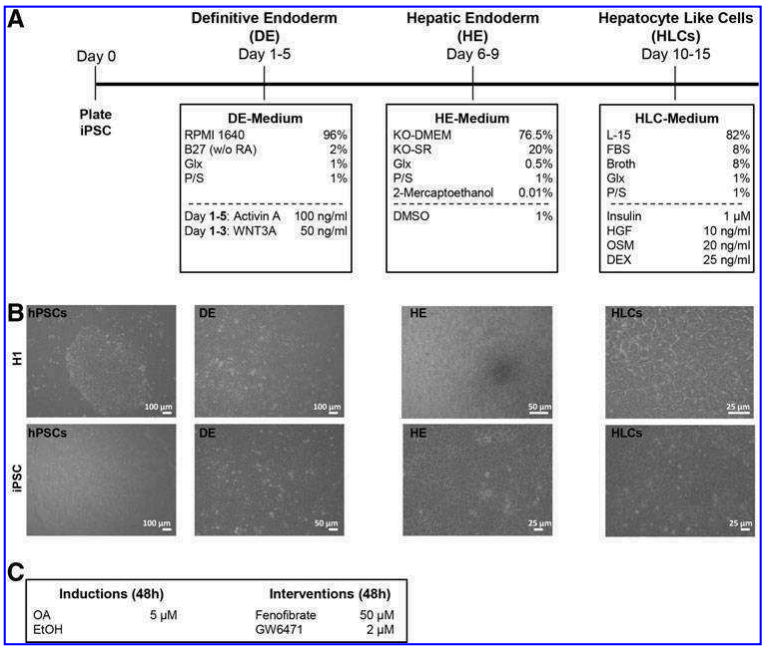


FIG. 1. Differentiation of hPSCs into HLCs. hPSCs were differentiated into HLCs using a three-step protocol (see also Materials and Methods section) **(A)**. Morphological changes of the differentiating cells are documented for each stage. At the definitive endoderm (DE) stage, the dense cell–cell contact is lost, and cells acquire the typical endodermal morphology. The characteristic polygonal shape of hepatocytes is already evident at the hepatic endoderm (HE) stage. This morphology is even more pronounced at the HLC stage **(B)**. Two days before the end of differentiation, steatosis was induced with 50 µM OA, and control cells were treated with an equal volume of ethanol. To interfere with LD formation, PPARα modulators were also added on the same day **(C)**. HLCs, hepatocyte-like cells; hPSCs, human pluripotent stem cells; LD, lipid droplet; OA, oleic acid; PPARα, peroxisome proliferator-activated receptor alpha.

blot with antibodies against PLIN2 (Proteintech) and ACTIN (CST). HRP coupled secondary antibodies were obtained from Abcam. Chemiluminescence was detected on a Fusion FX instrument (PeqLab) and analyzed with Fusion Capt Advance software (PeqLab) using rolling ball background correction.

RNA isolation and quantitative real-time polymerase chain reaction

For RNA isolation, up to 500,000 cells were lysed in 500 µL TRIzol. RNA was isolated with the Direct-zol™ RNA Isolation Kit (Zymo Research) according to the user's manual, including the DNase digestion step. Reverse transcription of up to 1 µg RNA was performed with the TaqMan Reverse Transcription (RT) Kit (Applied Biosystems). Primers were purchased from MWG; sequences are provided in Supplementary Table S2. Real-time PCR was performed in technical triplicates with Power SYBR Green Master Mix (Life technologies) on a VIIA7 or StepOnePlus (both Life technologies) machine. Mean values were normalized to actin and, subsequently, to the ethanol control. In case of siRNA experiments, data were normalized to the ethanol-treated nt siRNA sample. Experiments were carried out in biological duplicates and are depicted as mean values (log2) with standard error of the mean. Unpaired student's *t*-tests were performed for calculating significances.

Transcriptome and bioinformatics analysis

Microarray experiments were performed using the Affymetrix PrimeView chip (BMFZ, Düsseldorf). Details of data analysis are given in Supplementary Methods.

Liver-specific microRNA array

cDNA synthesis was carried out using the miScript II RT Kit (QIAGEN) according to the user's manual. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) and data analysis were performed according to the user's manual (miScript miRNA PCR Array; QIAGEN). All values were normalized to six different housekeeping genes (snoRNA/snRNA).

microRNA target gene validation

microRNA target gene validation was performed in HEK293T cells as previously described [26,27]. In brief, 3′ untranslated region (UTR) fragments of putative target

genes were cloned at the 3′ end of the *Firefly* luciferase open reading frame (ORF) in dual-luciferase reporter vector pmirGLO (Promega). Ds-oligonucleotides spanning the predicted microRNA binding sites for ATL3 (hsa-miR-106b) and CPAMD8 (hsa-miR-122) were used (Supplementary Table S3), whereas the 3,406-bp EPHA7-3′ UTR was represented by a 1,384-bp PCR fragment covering two predicted hsa-miR-106b binding sites (Supplementary Table S4). Normalization for effects of endogenous HEK293T microRNAs on the given 3′ UTR was achieved by transfection of both empty pmirGLO, as well as pmirGLO/3′ UTR, into HEK293T cells. Pairwise cotransfections of empty pmirGLO or pmirGLO/3′ UTR with the microRNA mimic of interest (Supplementary Table S5) (Dharmacon) were performed. *Firefly* and *Renilla* activities were determined 24 h after transfection. All transfections were performed in at least two independent biological experiments with quadruple transfections each. Mean values with standard deviations are shown. Significances were calculated with unpaired Student's *t*-test, $***P \leq 0.001$.

PLIN2 knockdown

PLIN2 knockdown was performed by transfecting 5 pmol PLIN2 siRNA or a nontarget (nt) control siRNA (Thermo Scientific) into 100,000 HepG2 cells using lipofectamine RNAiMAX (Life Technologies). Forty-eight hours post transfection, the cells were harvested for further analyses.

Results

Pluripotent stem cells differentiate into HLCs and show characteristic activities of liver cells

We differentiated hPSCs into HLCs using a three-step protocol, which is based on published protocols [28–30], but has been adapted in our laboratory to optimize the individual differentiation steps (Fig. 1A). We used the established ESC line H1 and an iPSC line derived from fetal foreskin fibroblasts [25] to compare two distinct hPSC lines. Morphological changes were monitored at every step of the differentiation process (Fig. 1B). ESCs and iPSCs behaved similarly during differentiation. At the definitive endoderm stage, they adopted a typical flat and elongated shape; then at the hepatic endoderm stage, cells from both sources started to adopt the polygonal shape that is characteristic for hepatocytes. At the maturation step, that is, the HLC stage, the vast majority of the cells had acquired the polygonal morphology.

Stainings with antibodies for the hepatocyte markers albumin, HNF4α, and E-Cadherin were all positive for HLCs derived from ESCs and iPSCs (Fig. 2A). In addition, several activity tests indicated that HLCs behave like hepatocytes (Fig. 2B–D). They were able to synthesize urea (Fig. 2B) and expressed active transporters as demonstrated by their ability to take up and release indocyanine green dye (Fig. 2C). Another important characteristic of hepatocytes is the activity of phase I enzymes, for example, members of the Cytochrome p450 family, which was strongly increased in HLCs compared to undifferentiated hPSCs (Fig. 2D). Thus, we can reliably generate functional HLCs from ESCs, as well as from iPSCs, which are suitable for disease modeling.

HLCs can be induced to accumulate LDs

To analyze early steatosis in cell culture, we first established a protocol for the induction of LDs in HLCs, which is based on the addition of 50 μM OA for 48 h into the medium. For both cell types, we observed similar increases in LD accumulation with Oil Red O and BODIPY 493/503 staining (Fig. 3A, B). *PLIN2* expression increased consistently after induction with OA (Fig. 3C). In addition, we analyzed the expression of several other genes involved in lipid metabolism to get a first impression of the immediate impact of fatty acid overload (Fig. 3C and Table 1). All factors were selected, because they were significantly regulated in liver cells of patients with high levels of steatosis compared to low-level steatosis [31]. In all cases, except for *ACADSB*, the expression changes in HLCs mirrored those observed in patient liver biopsies [31].

We decided to further focus on increased *PLIN2* expression as an indicator and marker for the induction of steatosis, because its role during progression of the disease has been analyzed extensively [13–16].

Transcriptome and associated pathway analysis of HLCs after steatosis induction

We next wanted to know how the induction of steatosis with OA affects the transcriptomes 48 h post treatment. To achieve this, we induced steatosis in HLCs derived from H1 ESCs and iPSCs as described above. Ethanol-treated cells were used as a solvent control. We found that ~13,000 genes were significantly expressed in H1- or iPSC-derived HLCs, respectively (Fig. 4A, B). In both cases, about 200 genes were exclusively expressed in the control cells, while 129 (H1) and 186 (iPSC) genes were exclusively expressed in OA-treated cells (Fig. 4A, B).

Analysis of the genes that were higher or exclusively expressed in HLCs treated with OA revealed significant enrichment of gene ontology (GO)-terms related to lipid metabolism and transport (Fig. 4C and Supplementary Table S6). As OA is dissolved in ethanol, it is not possible to completely rule out an ethanol-mediated influence on the expression of metabolically relevant genes. However, GO analysis for genes, which were exclusively expressed in ethanol-treated control cells, did not reveal any nonethanol-related metabolic pathways. Instead, pathways connected to signaling and nonhepatic development were enriched (Fig. 4C and Supplementary Table S7).

Overall, the individual regulated genes, as well as the associated GO categories and pathways, differed between both cell lines. This probably reflects innate discrepancies of the differentiation propensity of ES and iPS cells, as well as of course their different genetic background. Nonetheless, numerous genes associated with the PPAR pathway, which plays a major role in the regulation of lipid metabolism, were upregulated after OA treatment of HLCs derived from H1 or iPSCs (Supplementary Fig. S1).

Heatmap analysis of several factors involved in lipid and glucose metabolism, as well as in insulin signaling, revealed that although the absolute transcription levels of many factors differed between H1- and iPSC-derived HLCs, the OA-induced changes were frequently qualitatively similar (Fig. 4D). To get a more detailed insight into the transcriptional

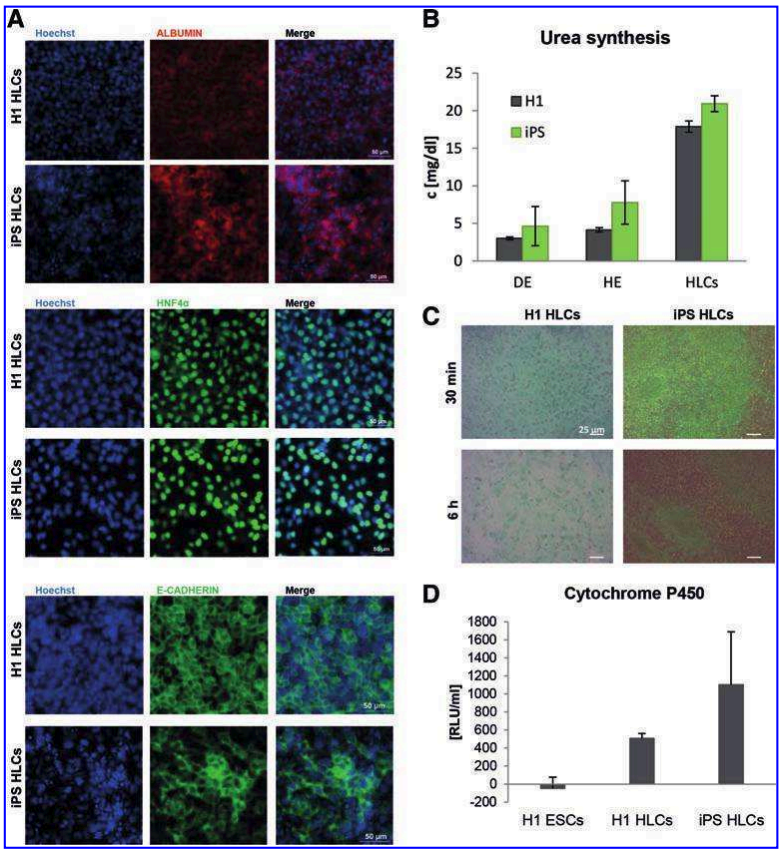


FIG. 2. Characterization of HLCs. The ES cell line H1 and iPS cells were differentiated into HLCs as described in Fig. 1. **(A)** HLCs derived from H1 (*upper row*) and iPS cells (*lower row*) express the liver-specific proteins Albumin, HNF4α (scale bar 50 µm), and E-Cadherin (scale bar 20 µm). **(B–D)** HLCs have liver-specific activity. **(B)** Urea production increases during differentiation and reaches its maximum at the HLC. **(C)** HLCs take up indocyanine green dye (*upper row*) within 30 min and release it within 6 h (*lower row*). **(D)** HLCs have phase I enzyme activity as measured by a luminometric assay detecting Cytochrome P450 (CYP) 3A4, 3A5, and 3A7 activity. In all cases, representative experiments are shown. In **(B)** and **(D)**, data represent mean value ± standard deviation. HLCs, hepatocyte-like cells. Color images available online at www.liebertpub.com/scd

changes that occur early in steatosis, we expanded our panel of directly investigated genes (Table 1) and included more factors relevant for lipid metabolism (*CPT1A*, *CPT2*, and *HADH*). We also added *APOC2* as another lipid binding protein and *GSK3A*, which regulates glucose metabolism (Fig. 4E). Interestingly, only *CPT1A* and *APOC2* were consistently upregulated in both samples, while we observed only minor and opposed expression changes for the other factors.

Many liver-specific microRNAs are downregulated in HLCs after induction of steatosis

Next, we analyzed whether induction of steatosis in H1- and iPSC-derived HLCs alters expression of liver-specific microRNAs. To this end, we used the liver finder microRNA array composed of 84 liver-related microRNAs. Interestingly, we found most microRNAs downregulated upon steatosis induction and only a few microRNAs upregulated (Fig. 5A). This finding implies that an altered microRNA expression profile is part and parcel of the early events in steatosis-induced cells, especially since the liver-specific microRNA hsa-miR-122 was among the most strongly downregulated miRNAs, together with hsa-miR-106b (Fig. 5A).

Bioinformatic target gene predictions revealed several thousand putative target genes for hsa-miRs-106b and -122 (Supplementary Table S8). Among these, we further analyzed liver-related genes *ATL3*, *EPHA7* (putative miR-106b targets), and *CPAMD8* (putative miR-122 target), which were upregulated upon steatosis induction in HLCs derived from H1 and

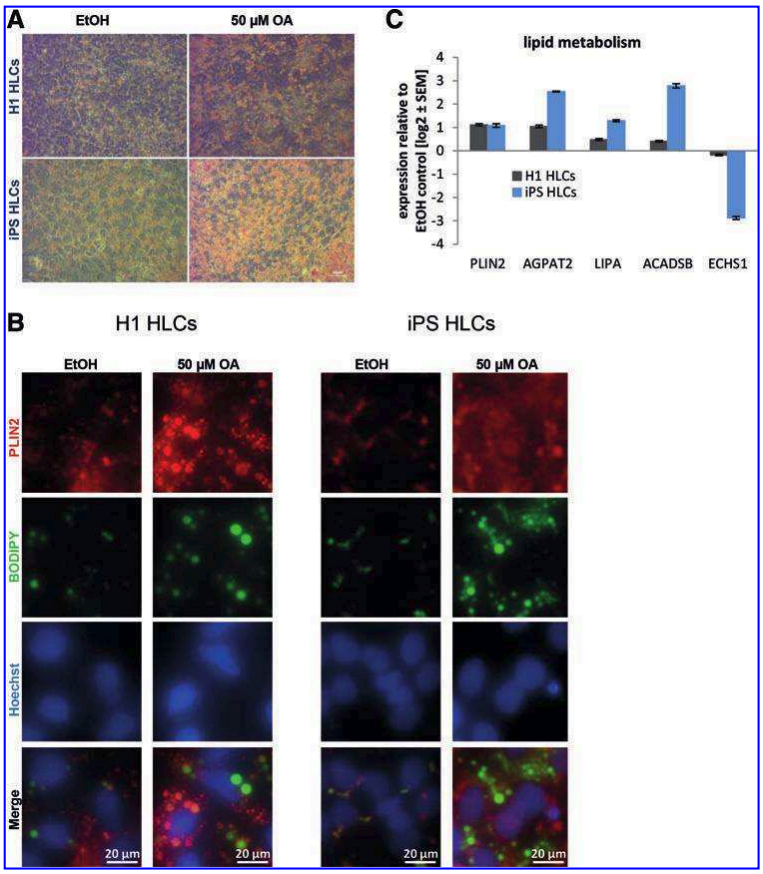


FIG. 3. Induction of steatosis in HLCs. hPSCs were differentiated into HLCs. Steatosis was induced by treatment with 50 µM OA for 48 h. Ethanol (EtOH)-treated cells served as controls. Steatosis was monitored by Oil Red O (*red*) **(A)** or BODIPY 493/503 (*green*) **(B)** staining of LDs. The control cells treated with ethanol show limited LD accumulation, while those treated with 50 µM OA for 48 h have abundant LDs. Scale bars: 20/50 µm. **(C)** Expression of genes involved in lipid metabolism was analyzed using qRT-PCR ($n=2$). Gene expression was normalized to β-actin and, subsequently, to the control samples. For each sample, the mean ± standard error is shown as log2 scale. HLCs, hepatocyte-like cells; hPSCs, human pluripotent stem cells; LDs, lipid droplets; OA, oleic acid; qRT-PCR, quantitative reverse transcription polymerase chain reaction. Color images available online at www.liebertpub.com/scd

iPSC (Fig. 5B). Remarkably, a significant number of miRNAs from the liver finder array were found among the predictions for these three genes (CPAMD8: 15%, ATL3: 12%, EPHA7: 44% from 84 miRNAs, see Supplementary Table S4). Luciferase reporter-based target gene validation indeed confirmed that hsa-miR-106b downregulates *ATL3* and *EPHA7*, while miR-122 targets *CPAMD8* (Fig. 5C–E).

siRNA-based suppression of PLIN2 level and function does not impair LD formation

Having characterized the impact of steatosis induction on the global gene expression, we next wanted to interfere with LD accumulation to find indications for putative treatments for NAFLD. First, we reduced PLIN2 expression using siRNA, because it is known that PLIN2 plays a major role during the development of steatosis [13–16]. As the transfection efficiency of in vitro derived HLCs is very low, we decided to focus on HepG2, a hepatocellular carcinoma line, for these experiments. Forty-eight hour treatment with a PLIN2 siRNA reduced its expression on protein level to about 33% (Fig. 6A and Supplementary Fig. S2A). Next, we induced the siRNA-treated HepG2 cells with OA for an additional 48 h. *PLIN2* siRNA reduced the expression of *PLIN2* on mRNA level to ~20% in the control and the OA-treated samples (Fig. 6B). However, even under siRNA knockdown conditions, *PLIN2* expression was still significantly higher in the induced sample compared to the control (Fig. 6B).

Table 1. Names and Functions of the Investigated Genes

Gene symbol	*Name*	*Function*	*Classification*
CPT1A	Carnitine palmitoyltransferase 1A	Transport acyl group of fatty acid-CoA conjugates across the mitochondrial membranes for beta-oxidation	Fatty acid catabolism
CPT2	Carnitine palmitoyltransferase 2		
ECHS1	Enoyl coenzyme A hydratase, short chain, 1, mitochondrial	Mitochondrial fatty acid beta-oxidation	
HADH	Hydroxyacyl-coenzyme A dehydrogenase	Mitochondrial fatty acid beta-oxidation	
LIPA	Lipase A, lysosomal acid, cholesterol esterase	Hydrolase of cholesteryl esters and triglycerides	
ACADSB	Acyl-coenzyme A dehydrogenase, short/branched chain	Dehydrogenase of acyl-CoA derivatives	
ACAT1	Acetyl-coenzyme A acetyltransferase 1	Ketone body metabolism	
PRKAA2	Protein kinase, AMP-activated, alpha 2 catalytic subunit	Regulator of FA and cholesterol biosynthesis	Regulators of metabolism
PPARα	Peroxisome proliferator-activated receptor alpha	Regulator of lipid and glucose metabolism	
GSK3A	glycogen synthase kinase 3 alpha	Regulator of glucose homeostasis	
AGPAT2	1-Acylglycerol-3-phosphate O-acyltransferase 2 (lysophosphatidic acid acyltransferase, beta)	Phospholipid biosynthesis	Biosynthesis
HMGCR	3-Hydroxy-3-methylglutaryl-coenzyme A reductase	Cholesterol synthesis	
PLIN2	Perilipin 2	Coats LDs	LD coating
APOC2	Apolipoprotein C-II	Lipid-binding protein (very low-density lipoprotein)	

LD, lipid droplet.

Contrary to our expectations, PLIN2 knockdown did not alter the accumulation of LDs in the cells after incubation with OA (Fig. 6C). Nonetheless, microarray analysis revealed that transcriptomes of cells treated with *PLIN2* siRNA cluster away from those incubated with the nt control siRNA (Supplementary Fig. S2B, left branches). In both clusters, OA treated and control cells form separate subclusters. A detailed view on the expression of lipid and glucose metabolism related genes also confirmed that *PLIN2* siRNA treatment induced major transcriptional changes (Fig. 6E).

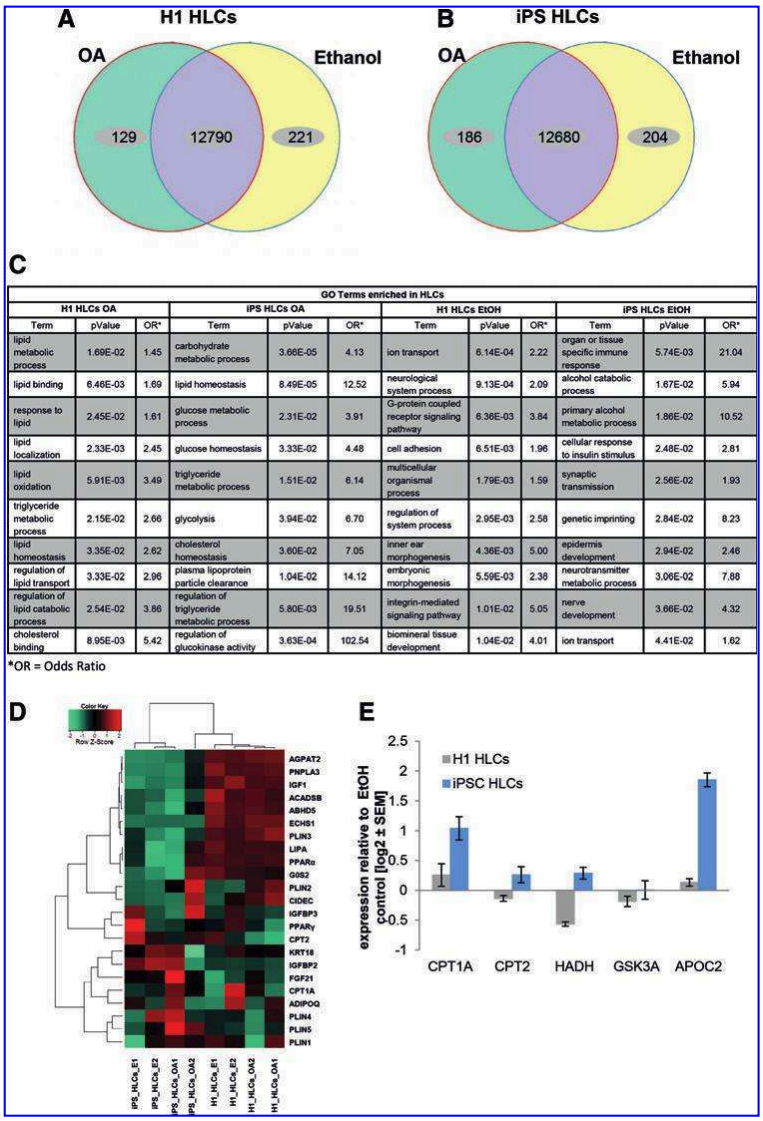
Numerous factors involved in lipid catabolism such as *ACADSB*, *LIPA*, and *CPT1A* were downregulated after *PLIN2* siRNA treatment. These tendencies were confirmed by qRT-PCR (Fig. 6D). However, in this study, the differences between the OA and EtOH-treated cells became more obvious. Interestingly, also *PPARα* and γ, which regulate lipid metabolism, were downregulated after PLIN2 knockdown (Fig. 6E). This argues in favor of an active PPAR gene regulatory network associated with steatosis. Overall, PLIN2 knockdown modulates the expression of steatosis-related genes, but does not reduce lipid accumulation, at least in the early stage investigated.

Modulation of the lipid metabolism regulating factor PPARα has major impact on numerous metabolism-related pathways

The hepatic nuclear receptor PPARα is activated by a variety of ligands [32,33]. It regulates lipid and glucose metabolism in the liver and PLIN2 is one of its known targets [34,35]. Transcriptome analysis revealed that many members of the PPARα signaling pathway were upregulated either in H1- or iPSC-derived HLCs after treatment with OA (Supplementary Fig. S1). Its expression was also influenced by PLIN2 knockdown (Fig. 6D, E). As we could not transfect HLCs with siRNA against *PLIN2* to a significant level, we decided to interfere with PPARα signaling by activating or inhibiting PPARα action with Fenofibrate or GW6471, respectively [36].

iPSC-derived HLCs were treated for 48 h with OA and either Fenofibrate or GW6471. BODIPY staining revealed that HLCs incorporated LDs regardless of treatment with the PPARα modulators (Fig. 7A). This is in line with results from Rogue et al. who showed that PPARα agonist treatment reduced fat load in HepaRG cells only after a prolonged incubation of 14 days [37]. However, microarray analysis revealed that even short-term treatment with Fenofibrate and GW6471 had an impact on gene expression. The transcriptomes of HLCs incubated with either Fenofibrate or GW6471 clearly clustered away from each other (Fig. 7B).

Heatmap-based analysis of the PPAR pathway revealed that cells treated with Fenofibrate or GW6471 behave differently (Fig. 7C). In both cases, there were two subclusters detectable that correspond to the control cells and the OA-treated cells. We monitored the expression levels of genes important for lipid and glucose metabolism in more detail by qRT-PCR. In most cases, treatment with the agonist and the antagonist resulted in opposing changes in gene expression, as expected (Fig. 7D). Inhibition of PPARα with GW6471 resulted in downregulation of genes involved in lipid catabolism, while activation using Fenofibrate reduced expression of *AGPAT2* and *HMGCR*, which are involved in biosynthesis of phospholipids and cholesterol, respectively (Fig. 7D).



C

GO Terms enriched in HLCs											
H1 HLCs OA			iPS HLCs OA			H1 HLCs EtOH			iPS HLCs EtOH		
Term	pValue	OR*	Term	pValue	OR*	Term	pValue	OR*	Term	pValue	OR*
lipid metabolic process	1.69E-02	1.45	carbohydrate metabolic process	3.66E-05	4.13	ion transport	6.14E-04	2.22	organ or tissue specific immune response	5.74E-03	21.04
lipid binding	6.46E-03	1.69	lipid homeostasis	8.49E-05	12.52	neurological system process	9.13E-04	2.09	alcohol catabolic process	1.67E-02	5.94
response to lipid	2.45E-02	1.61	glucose metabolic process	2.31E-02	3.91	G-protein coupled receptor signaling pathway	6.36E-03	3.84	primary alcohol metabolic process	1.86E-02	10.52
lipid localization	2.33E-03	2.45	glucose homeostasis	3.33E-02	4.48	cell adhesion	6.51E-03	1.96	cellular response to insulin stimulus	2.48E-02	2.81
lipid oxidation	5.91E-03	3.49	triglyceride metabolic process	1.51E-02	6.14	multicellular organismal process	1.79E-03	1.59	synaptic transmission	2.56E-02	1.93
triglyceride metabolic process	2.15E-02	2.66	glycolysis	3.94E-02	6.70	regulation of system process	2.95E-03	2.58	genetic imprinting	2.84E-02	8.23
lipid homeostasis	3.35E-02	2.62	cholesterol homeostasis	3.60E-02	7.05	inner ear morphogenesis	4.36E-03	5.00	epidermis development	2.94E-02	2.46
regulation of lipid transport	3.33E-02	2.96	plasma lipoprotein particle clearance	1.04E-02	14.12	embryonic morphogenesis	5.59E-03	2.38	neurotransmitter metabolic process	3.06E-02	7.88
regulation of lipid catabolic process	2.54E-02	3.86	regulation of triglyceride metabolic process	5.80E-03	19.51	integrin-mediated signaling pathway	1.01E-02	5.05	nerve development	3.66E-02	4.32
cholesterol binding	8.95E-03	5.42	regulation of glucokinase activity	3.63E-04	102.54	biomineral tissue development	1.04E-02	4.01	ion transport	4.41E-02	1.62

*OR = Odds Ratio

FIG. 4. Gene expression profiles of HLCs challenged with OA enable characterization of the early steps in NAFLD. Venn diagrams of genes expressed (detection *P* value <0.05) in H1- **(A)** or iPSC- **(B)** derived HLCs ($n = 2$). *Green* and *yellow* segments represent genes exclusively expressed in OA treated or control cells, respectively. *Purple* segments comprise genes, which are expressed under both conditions, but not necessarily with the same intensity. **(C)** GO-analysis of genes upregulated or exclusively expressed under OA treatment reveals an enrichment of NAFLD-related categories, while ethanol control cells predominantly express genes mapping to divergent categories. Shown are preselected significant GO-Terms; for full data set, see Supplementary Tables S6 and S7. **(D)** Heatmap of genes involved in insulin signaling and lipid or glucose metabolism reveals differences between H1- and iPSC-derived HLCs, which might, in part, be due to their distinct genetic background. However, qualitative changes are similar between both groups, which are reflected by the formation of subclusters connected to OA treatment. **(E)** Expression of genes involved in lipid and glucose metabolism was analyzed using qRT-PCR ($n = 2$). Gene expression was normalized to β-actin and, subsequently, to the control samples. For each sample, the mean ± standard error of duplicate experiments is shown as log2 scale. GO, gene ontology; HLCs, hepatocyte-like cells; iPSC, induced pluripotent stem cell; NAFLD, nonalcoholic fatty liver disease; OA, oleic acid; qRT-PCR, quantitative reverse transcription polymerase chain reaction. Color images available online at www.liebertpub.com/scd

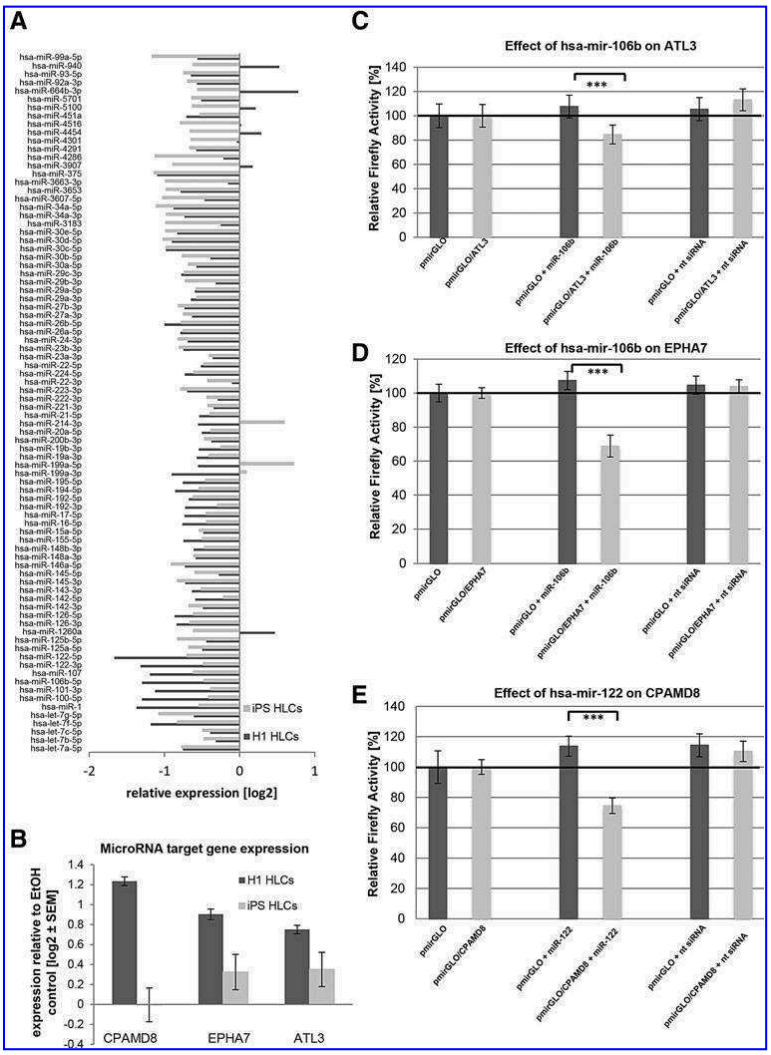


FIG. 5. Regulated expression of microRNAs upon induction of steatosis. **(A)** Expression of 84 liver-specific microRNAs from H1-HLCs (*dark gray*) and iPSC-HLCs (*light gray*) was analyzed after the induction of steatosis. MicroRNA expression was normalized to six different genes of small housekeeping RNAs. For H1-derived HLCs, biological and for iPSC-derived HLCs, technical duplicates were performed. The mean value is shown as log2 expression. With a few exceptions, all miRNAs were downregulated in steatosis-induced cells from both sources. **(B)** Expression of selected microRNA target genes was analyzed through qRT-PCR ($n = 2$). Gene expression was normalized to β-actin and, subsequently, to the control samples. For each sample, the mean ± standard error is shown as log2 scale. **(C–E)** Validation of predicted target genes for hsa-miRNA-106b [ATL3 **(C)**, EPHA7 **(D)**] and hsa-miRNA-122 [CPAMD8 **(E)**]. To test the influence of endogenous microRNAs, empty *Firefly*/*Renilla* dual-reporter vector pmirGLO and pmirGLO/3′ UTR (containing the 3′UTR of interest at the 3′ end of the Firefly open reading frame) were each transfected into HEK293T cells ($n = 4$). Normalized *Firefly* activities were compared with those of pairwise cotransfections of these vectors with the microRNA mimic of interest (miR-106b, -122, and an unspecific siRNA negative control) to test for unspecific effects of the given microRNA-mimic on *Firefly*/*Renilla* per se and for validation of the particular target prediction. *Dark gray* columns show normalized *Firefly* activities from pmirGLO (mimic-co)-transfections; *light gray* columns are those from pmirGLO/3′ UTR cotransfections. Percentage reductions of *Firefly* activities of pmirGLO/3′ UTR compared with pmirGLO are given, as well as their statistical significances (Student's *t*-test, unpaired, $^{***}P \leq 0.001$). All three predicted interactions were tested positive. HLCs, hepatocyte-like cells; iPSC, induced pluripotent stem cell; qRT-PCR, quantitative reverse transcription polymerase chain reaction; UTR, untranslated region.

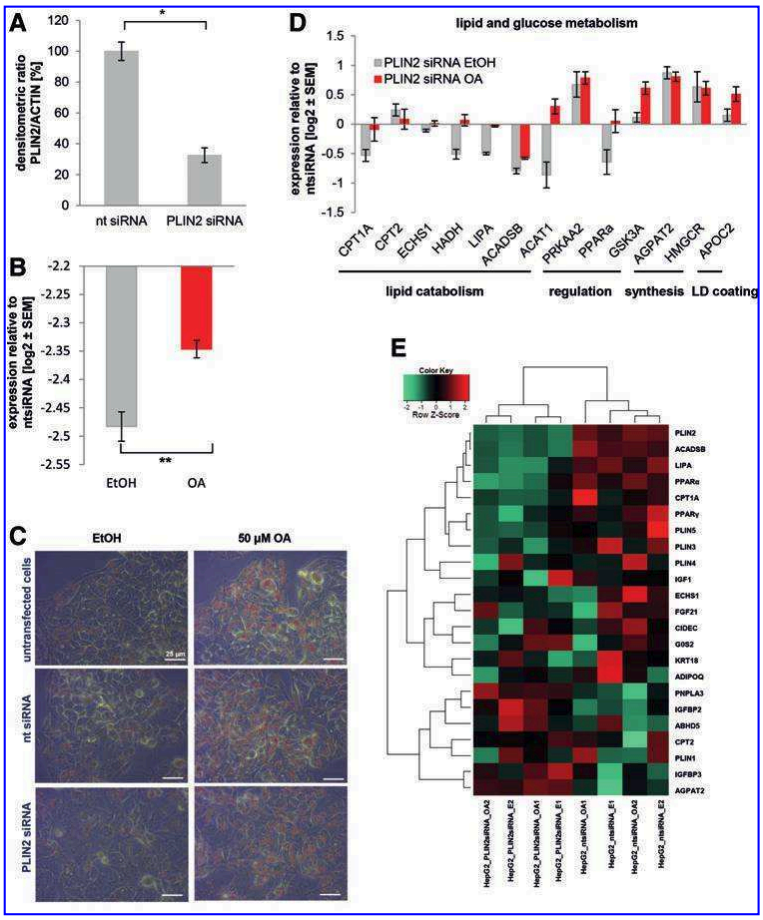


FIG. 6. Knockdown of PLIN2 in HepG2 cells does not affect LD formation, but alters gene expression. PLIN2 was knocked down in HepG2 cells by lipofectamine-mediated siRNA transfection ($n=2$). Cells were either transfected with nt siRNA or with PLIN2 siRNA. **(A)** After 48 h, PLIN2 expression was assessed by western blot and normalized to ACTIN levels. **(B)** After 48 h, steatosis was induced with 50 µM OA for an additional 48 h. As control, cells were treated with ethanol (EtOH). Expression of PLIN2 was analyzed on mRNA level using qRT-PCR. Although levels were reduced by ~80% after transfection, we observed significantly increased levels after OA treatment. Gene expression was normalized to β-ACTIN and, subsequently, to the control samples. For each sample, the mean ± standard error of duplicate experiments is shown as log2 scale. Significances were calculated with an unpaired Student's *t*-test. **(C)** LD formation was monitored by Oil Red O staining. In the *upper row*, untransfected control cells are shown, followed by cells transfected with nt siRNA (*middle row*) and PLIN2 siRNA, *last row*. Ethanol-treated control cells show only minor LD accumulation (*left column*), while those treated with 50 µM OA for 48 h have abundant LDs that are stained by Oil Red O. Scale bar: 25 µm. Expression of genes involved and lipid or glucose metabolism **(D)** were analyzed using qRT-PCR. **(E)** Heatmap representation of genes involved in insulin signaling and lipid or glucose metabolism reveals two clusters related to treatment with either PLIN2 or the nt siRNA. LD, lipid droplet; OA, oleic acid; PLIN, Perilipin 2; qRT-PCR, quantitative reverse transcription polymerase chain reaction. Color images available online at www.liebertpub.com/scd

When we compared the number of GO-terms associated with distinct metabolic categories that were significantly regulated after PPARα modulation, it became obvious that many were connected with glucose, lipid, and purine metabolism (Fig. 7E and Supplementary Table S9). Interestingly, genes exclusively expressed in GW6471-treated HLCs were predominantly mapped to lipid metabolism or transport, while this category was only marginally upregulated in Fenofibrate-treated cells, indicating that inhibition of PPARα strongly affects lipid metabolism (Fig. 7E). In contrast, expression of genes that are associated with purine metabolism was upregulated after PPARα activation, while

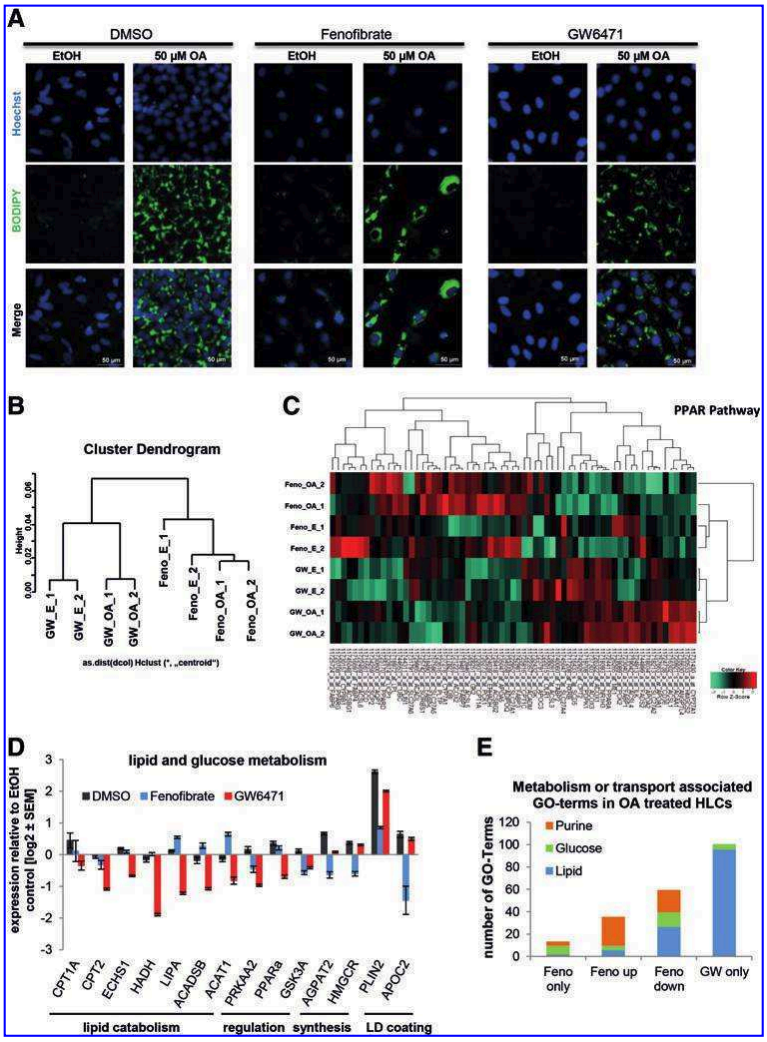


FIG. 7. PPARα signaling has major impact on the induction of steatosis. IPSC-derived HLCs were incubated in parallel with either Fenofibrate (Feno, PPARα agonist) or GW6471 (PPARα antagonist) during the 48 h OA induction. **(A)** Steatosis induction was monitored by BODIPY 493/503 (*green*) staining of LDs. In every case, LDs increased after OA induction (*right columns*) compared to the control (*left columns*), but no differences between the different PPARα treatments are visible. **(B)** Microarray-based transcriptome analysis revealed two distinct clusters representing Fenofibrate and GW6471 treatment and two subclusters related to OA treatment ($n = 2$). **(C)** Heatmap representation of PPAR pathway genes shows distinct gene expression profiles for Fenofibrate (''F'') and GW6471 (''GW'') treated cells with qualitative changes after OA induction. Expression of genes involved in lipid or glucose metabolism **(D)** was analyzed using qRT-PCR ($n = 2$). As expected, in most cases, Fenofibrate and GW6471 treatment had opposing effects on gene expression. Gene expression was normalized to β-actin and, subsequently, to the ethanol-treated control samples. For each sample, the mean ± standard error of duplicate experiments is shown as log2 scale. **(E)** Significantly expressed genes from the global analysis were subdivided into genes only expressed after Fenofibrate treatment, up- or downregulated after Fenofibrate treatment, compared to GW6471 treatment or only expressed after GW6471 treatment. Then they were assigned to GOs (Supplementary Table S9). The numbers of GO-terms that were associated with lipid, glucose, or purine metabolism and transport are displayed. GO, gene ontology; HLCs, hepatocyte-like cells; LDs, lipid droplets; OA, oleic acid; PPARα, peroxisome proliferator-activated receptor alpha; qRT-PCR, quantitative reverse transcription polymerase chain reaction. Color images available online at www.liebertpub.com/scd

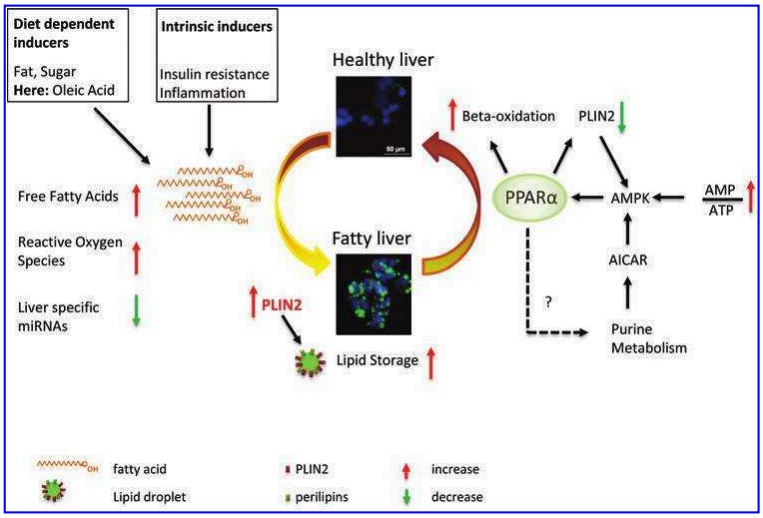


FIG. 8. Schematic overview of inducers of NAFLD and potential intervention points. The transition from a healthy liver (*upper panel*) toward a fatty liver (*lower panel*) is shown. Factors that promote NAFLD are shown on the *left*. Besides diet-related inducers that are in this study mimicked by OA addition to the medium, intrinsic factors like insulin resistance promote NAFLD. From a global perspective, the triggers result in increased free fatty acids and reactive oxygen species. We also observed a global downregulation of hepatic miRNAs. Eventually, every trigger results in an upregulation of PLIN2 expression and an increase in lipid storage within hepatocytes. Beneficial processes that might reduce fat content of hepatocytes are shown on the *right side*. In general, PPARα activation (in our case by application of Fenofibrate) decreases the amount of stored fat by inducing β-oxidation. Besides fibrates, which activate PPARα, it is also regulated by AMPK, which senses the low energy state of a cell as measured by the ratio between AMP and ATP. AMPK expression increases in response to PLIN2 knockdown conditions and its activity might be regulated by intermediates of purine metabolism, which is in turn regulated by PPARα. This indicates a possible positive feedback loop that has to be verified by additional experiments. Besides PPARα modulation, a direct reduction of PLIN2 expression could also be beneficial. AMP, adenosine monophosphate; AMPK, adenosine monophosphate-activated protein kinase; ATP, adenosine triphosphate; NAFLD, nonalcoholic fatty liver disease; OA, oleic acid; PLIN, Perilipin 2; PPARα, peroxisome proliferator-activated receptor alpha. Color images available online at www.liebertpub.com/scd

none of the genes that were exclusively expressed in GW6471-treated cells mapped onto this GO-category.

A detailed analysis of the purine metabolic pathway demonstrated that genes associated with the early steps of purine metabolism tended to be downregulated under Fenofibrate treatment, while genes for the later steps were either upregulated or exclusively expressed during Fenofibrate treatment (Supplementary Fig. S3). This implies that PPARα activation regulates purine metabolism.

Taken together, we have generated a robust in vitro model for NAFLD based on human PSC-derived HLCs. Upon OA induction, the cells recapitulate many features observed in NAFLD patients (LD accumulation, PLIN2 overexpression, and dysregulation of metabolic pathways) and they react to variation of PPARα activity.

Discussion

In this study, we established a cell culture model for deciphering the molecular basis underlying early steps of LD accumulation in NAFLD, based on HLCs derived from hPSCs. We showed that ESCs and iPSCs both differentiate into HLCs that express hepatocyte-specific proteins and have characteristic hepatocyte-like biochemical functions. HLCs could be induced to accumulate fat in LDs by incubating them with OA. Global gene expression analysis after OA induction revealed upregulation of many GO-categories associated with lipid or glucose metabolism. This effect was qualitatively similar between ESC- and iPSC-derived HLCs, but differed quantitatively due to the different genetic backgrounds of the two cell lines and the usual interdifferentiation variations. Importantly, ethanol-treated control cells had a completely different panel of genes upregulated, which mapped to GO-categories related to signaling or nonhepatic development. This indicates that expression changes observed in the OA group are not induced by the solvent ethanol, but rather by OA itself. However, for the future, it is important to compare effects of different steatosis inducing agents, for example, palmitic acid or steatogenic drugs.

After OA induction, many members of the PPAR signaling pathway, which controls lipid metabolism, were upregulated. Of the different PPAR isoforms, PPARα is most abundant in the liver. It regulates lipid metabolism as a reaction to nutritional status and partly depends on insulin signaling [34]. In general, activation of PPARα with fibrates is a recognized treatment for the metabolic syndrome [38] and its expression levels negatively correlate with the severity of NASH [39].

Upon induction of steatosis, expression of a panel of genes important for lipid metabolism changed in the same manner as previously observed by comparing liver biopsies from patients with high versus low levels of steatosis [31]. Among these factors, we identified *PLIN2* upregulation as a suitable marker for successful induction of steatosis. PLIN2 is involved in the synthesis of LDs and plays an important role in the onset of steatosis. It has been shown that PLIN2 knockout mice develop neither obesity nor steatosis [16] and the knockdown of PLIN2 expression with an antisense oligonucleotide protected mice against the development of fatty liver [14].

To analyze the influence of PLIN2 expression on the development of steatosis, we knocked down PLIN2 in HepG2 cells using siRNAs. Although the knockdown was highly efficient with only 32% of protein expression remaining, we could not observe any changes in LD accumulation after induction with OA. This might be due to the short time period of only 48 h. There might still be enough PLIN2 left for coverage of some LDs, while the shortage of PLIN2 limits the amount of LDs that can be assembled over the long term. In addition, several other proteins are involved in LD formation. During steatosis, LDs increase in size relative to the abundance of TAGs [36]. The largest LDs are predominantly covered by PLIN1, which competes on the LD surface with PLIN2 [36]. Global gene expression analysis of PLIN2 knockdown cells revealed indeed that *PLIN1* and *PLIN4* expression was enhanced in at least two of the four samples compared to the nt siRNA samples.

Overall, we observed profound genome-wide transcriptional changes, which are reflected by the fact that the transcriptomes of *PLIN2* and nt siRNA-treated cells cluster separately from each other. Interestingly, expression levels of several factors known to have important regulatory functions on metabolism were altered after PLIN2 knockdown. For example, *PRKAA2*, which is the catalytic subunit of adenosine monophosphate-activated protein kinase (AMPK), a sensor of nutritional level that becomes activated during fasting [40] was strongly upregulated in PLIN2 knockdown cells, but reduced again upon OA induction. This shows that the cells still react to nutritional abundance. The lack of PLIN2, although it does not have a detectable influence on LDs, seems to transfer some kind of fasting signal that activates *PRKAA2* transcription. However, our data do not provide any information on PRKAA2 activity, which has to be analyzed in the future.

In addition, *PPARα* and γ, which regulate lipid metabolism, were downregulated in *PLIN2* siRNA-treated cells. This fits well with the predominant downregulation of lipid catabolism genes that are at least, in part, directly regulated by PPARα. Our data point at the existence of a PPAR-gene regulatory network that depends on *PLIN2* levels. In addition, expression levels of genes important for insulin signaling were altered after *PLIN2* siRNA treatment. Interestingly, the expression of *IGFBP2* and *3*, two important regulators of IGF1 signaling [41], were induced, while IGF itself was reduced. Overall, these data indicate that although PLIN2 knockdown has no detectable effect on LD formation, it has a major impact on metabolic activities of the cell.

A knockdown of PLIN2 was not possible in HLCs because of the low transfection efficiency in primary cells. As PLIN2 is a target of PPARα signaling [39], we interfered with PPARα activity using two small molecules, Fenofibrate (agonist) and GW6471 (antagonist). Again, we could not observe any changes in LD formation after treating induced HLCs with either Fenofibrate or GW6471. However, we investigated only short-term effects and it has been shown that PPARα activation over a longer time period reduces steatosis in HepaRG cells [37]. Nonetheless, we observed major gene expression changes that occur as an immediate reaction toward PPARα modulation accompanying OA induction. Expressed genes could be mapped to pathways related to lipid, glucose, and purine metabolism. Interestingly, the latter was not present in OA-induced HLCs, where PPARα activity was reduced by GW6471. In this study, the most prominently upregulated GOs belonged to lipid metabolism. This indicates that the PPARα agonist and antagonist do not simply influence gene expression in opposing directions, although we observed this for a panel of key metabolic genes, but they also modify metabolic pathways from different angles.

Genes of the purine metabolism were generally regulated in Fenofibrate-treated cells. A recent study by Asby et al. demonstrated that Aminoimidazole carboxamide ribonucleotide (AICAR), an intermediate product in de-novo purine synthesis, activates AMPK [42], which is a key regulator of metabolism that senses the nutritional state of the cell and induces PPARα activity and catabolic pathways [40]. We observed that genes, which directly precede the synthesis of AICAR, were either upregulated or exclusively expressed during Fenofibrate treatment. Thus, it is possible that in addition to a direct influence on fat and glucose metabolism, PPARα also indirectly enhances these pathways by regulating purine metabolism, and thus, AMPK activity.

For most of the genes that were analyzed in more detail, treatment with the PPARα agonist and antagonist resulted as expected in opposing changes in expression. While GW6471-mediated inhibition of PPARα resulted in downregulation of genes involved in lipid catabolism, activation using Fenofibrate reduced expression of *AGPAT2* and *HMGCR*, which are involved in biosynthesis of phospholipids and cholesterol, respectively. Thus, our HLC model can, in part, reproduce the beneficial role of PPARα enhancement in patients with metabolic syndrome at least with regard to the reduced expression of lipid and cholesterol synthesizing enzymes.

In addition to major gene expression changes, we observed that the expression of most liver-specific microRNAs was downregulated as an early reaction to steatosis induction. Among the most strongly downregulated miRNAs, miR-122 is a key factor in liver development, differentiation, and homeostasis. It is elevated in the serum of NAFLD patients, while its expression levels in hepatocytes are concomitantly reduced, which is corroborated by our observations [43,44]. MiR-122a knockout mice develop steatosis and have altered levels of enzymes important for lipogenesis, LD formation, and lipid transport [45,46]. MiR-106b has not yet been associated with NAFLD, but its overexpression has been reported during development of cirrhosis and hepatocellular carcinoma [47].

Among the putative miR-106b and -122 targets predicted by at least four of the algorithms implemented in miRWalk, we confirmed *CPAMD8*, *ATL3*, and *EPHA7* as new miR-122 and miR-106b targets, respectively. None of these three proteins have been functionally associated with NAFLD. However, the validated, as well as the predicted, impact of

downregulated liver-related miRNAs on their expression together with their concordant upregulation in induced HLCs point to a functional role of these factors, but mechanisms remain to be elucidated.

We are aware of the fact that the two analyzed hPSC lines can only give a first impression of the general mechanisms of NAFLD development. To dissect these in more detail, we have generated several patient-specific iPSC lines, which are currently being characterized and which will be used in the future for more detailed analyses of the disease [48].

To summarize, our results show that hPSC-derived HLCs are a valuable in vitro model for investigating the molecular basis of the early steps of NAFLD. They accumulate LDs, and expression changes of metabolically relevant genes mirror those observed in liver biopsies of steatosis patients [31]. In addition, lipid metabolism can be regulated by modulating the activity of PPARα. A short overview of the different processes is given in Fig. 8.

Acknowledgments

J.A. acknowledges support from the Medical faculty of the Heinrich Heine University Düsseldorf. The authors thank M. Bohndorf and S. Wehrmeyer for technical support.

Prior Conference Presentation: Part of the work was presented as a poster at the 8th International Meeting of the Stem Cell Network North Rhine-Westphalia, April 21–22, 2015, Bonn, Germany.

Author Disclosure Statement

No competing financial interests exist.

References

1. Cohen JC, JD Horton and HH Hobbs. (2011). Human fatty liver disease: old questions and new insights. Science 332: 1519–1523.
2. Perry RJ, JP Camporez, R Kursawe, PM Titchenell, D Zhang, CJ Perry, MJ Jurczak, A Abudukadier, MS Han, et al. (2015). Hepatic acetyl CoA links adipose tissue inflammation to hepatic insulin resistance and type 2 diabetes. Cell 160:745–758.
3. Basaranoglu M, G Basaranoglu and H Senturk. (2013). From fatty liver to fibrosis: a tale of ''second hit''. World J Gastroenterol 19:1158–1165.
4. Okumura T. (2011). Role of lipid droplet proteins in liver steatosis. J Physiol Biochem 67:629–636.
5. Thiele C and J Spandl. (2008). Cell biology of lipid droplets. Curr Opin Cell Biol 20:378–385.
6. Greenberg AS, RA Coleman, FB Kraemer, JL McManaman, MS Obin, V Puri, QW Yan, H Miyoshi and DG Mashek. (2011). The role of lipid droplets in metabolic disease in rodents and humans. J Clin Invest 121:2102–2110.
7. Listenberger LL, AG Ostermeyer-Fay, EB Goldberg, WJ Brown and DA Brown. (2007). Adipocyte differentiation-related protein reduces the lipid droplet association of adipose triglyceride lipase and slows triacylglycerol turnover. J Lipid Res 48:2751–2761.
8. Sztalryd C, G Xu, H Dorward, JT Tansey, JA Contreras, AR Kimmel and C Londos. (2003). Perilipin A is essential for the translocation of hormone-sensitive lipase during lipolytic activation. J Cell Biol 161:1093–1103.
9. Miyoshi H, SC Souza, HH Zhang, KJ Strissel, MA Christoffolete, J Kovsan, A Rudich, FB Kraemer, AC Bianco, MS Obin and AS Greenberg. (2006). Perilipin promotes hormone-sensitive lipase-mediated adipocyte lipolysis via phosphorylation-dependent and -independent mechanisms. J Biol Chem 281:15837–15844.
10. Brasaemle DL, T Barber, NE Wolins, G Serrero, EJ Blanchette-Mackie and C Londos. (1997). Adipose differentiation-related protein is an ubiquitously expressed lipid storage droplet-associated protein. J Lipid Res 38:2249–2263.
11. Targett-Adams P, D Chambers, S Gledhill, RG Hope, JF Coy, A Girod and J McLauchlan. (2003). Live cell analysis and targeting of the lipid droplet-binding adipocyte differentiation-related protein. J Biol Chem 278:15998–16007.
12. Imamura M, T Inoguchi, S Ikuyama, S Taniguchi, K Kobayashi, N Nakashima and H Nawata. (2002). ADRP stimulates lipid accumulation and lipid droplet formation in murine fibroblasts. Am J Physiol Endocrinol Metab 283: E775–E783.
13. Motomura W, M Inoue, T Ohtake, N Takahashi, M Nagamine, S Tanno, Y Kohgo and T Okumura. (2006). Up-regulation of ADRP in fatty liver in human and liver steatosis in mice fed with high fat diet. Biochem Biophys Res Commun 340:1111–1118.
14. Imai Y, S Boyle, GM Varela, E Caron, X Yin, R Dhir, R Dhir, MJ Graham and RS Ahima. (2012). Effects of perilipin 2 antisense oligonucleotide treatment on hepatic lipid metabolism and gene expression. Physiol Genomics 44:1125–1131.
15. Imai Y, GM Varela, MB Jackson, MJ Graham, RM Crooke and RS Ahima. (2007). Reduction of hepatosteatosis and lipid levels by an adipose differentiation-related protein antisense oligonucleotide. Gastroenterology 132:1947–1954.
16. McManaman JL, ES Bales, DJ Orlicky, M Jackman, PS MacLean, S Cain, AE Crunk, A Mansur, CE Graham, TA Bowman and AS Greenberg. (2013). Perilipin-2-null mice are protected against diet-induced obesity, adipose inflammation, and fatty liver disease. J Lipid Res 54:1346–1359.
17. Takahashi Y, Y Soejima and T Fukusato. (2012). Animal models of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. World J Gastroenterol 18:2300–2308.
18. Machado MV, GA Michelotti, G Xie, TP de Almeida, J Boursier, B Bohnic, CD Guy and AM Diehl. (2015). Mouse models of diet-induced nonalcoholic steatohepatitis reproduce the heterogeneity of the human disease. PLoS One 10:e0127991.
19. Ricchi M, MR Odoardi, L Carulli, C Anzivino, S Ballestri, A Pinetti, LI Fantoni, F Marra, M Bertolotti, et al. (2009). Differential effect of oleic and palmitic acid on lipid accumulation and apoptosis in cultured hepatocytes. J Gastroenterol Hepatol 24:830–840.
20. De Gottardi A, M Vinciguerra, A Sgroi, M Moukil, F Ravier-Dall'Antonia, V Pazienza, P Pugnale, M Foti and A Hadengue. (2007). Microarray analyses and molecular profiling of steatosis induction in immortalized human hepatocytes. Lab Invest 87:792–806.
21. Currie E, A Schulze, R Zechner, TC Walther and RV Farese, Jr. (2013). Cellular fatty acid metabolism and cancer. Cell Metab 18:153–161.
22. Cantor JR and DM Sabatini. (2012). Cancer cell metabolism: one hallmark, many faces. Cancer Discov 2:881–898.
23. Godoy P, NJ Hewitt, U Albrecht, ME Andersen, N Ansari, S Bhattacharya, JG Bode, J Bolleyn, C Borner, et al. (2013). Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal liver cells and their use in investigating mech-

anisms of hepatotoxicity, cell signaling and ADME. Arch Toxicol 87:1315–1530.
24. Thomson JA, J Itskovitz-Eldor, SS Shapiro, MA Waknitz, JJ Swiergiel, VS Marshall and JM Jones. (1998). Embryonic stem cell lines derived from human blastocysts. Science 282:1145–1147.
25. Wang Y and J Adjaye. (2011). A cyclic AMP analog, 8-Br-cAMP, enhances the induction of pluripotency in human fibroblast cells. Stem Cell Rev 7:331–341.
26. Iwaniuk KM, J Schira, S Weinhold, M Jung, J Adjaye, HW Muller, P Wernet and HI Trompeter. (2011). Network-like impact of MicroRNAs on neuronal lineage differentiation of unrestricted somatic stem cells from human cord blood. Stem Cells Dev 20:1383–1394.
27. Trompeter HI, H Abbad, KM Iwaniuk, M Hafner, N Renwick, T Tuschl, J Schira, HW Muller and P Wernet. (2011). MicroRNAs MiR-17, MiR-20a, and MiR-106b act in concert to modulate E2F activity on cell cycle arrest during neuronal lineage differentiation of USSC. PLoS One 6:e16138.
28. Hay DC, D Zhao, J Fletcher, ZA Hewitt, D McLean, A Urruticoechea-Uriguen, JR Black, C Elcombe, JA Ross, R Wolf and W Cui. (2008). Efficient differentiation of hepatocytes from human embryonic stem cells exhibiting markers recapitulating liver development in vivo. Stem Cells 26:894–902.
29. Hay DC, J Fletcher, C Payne, JD Terrace, RC Gallagher, J Snoeys, JR Black, D Wojtacha, K Samuel, et al. (2008). Highly efficient differentiation of hESCs to functional hepatic endoderm requires ActivinA and Wnt3a signaling. Proc Natl Acad Sci U S A 105:12301–12306.
30. Jozefczuk J, A Prigione, L Chavez and J Adjaye. (2011). Comparative analysis of human embryonic stem cell and induced pluripotent stem cell-derived hepatocyte-like cells reveals current drawbacks and possible strategies for improved differentiation. Stem Cells Dev 20:1259–1275.
31. Wruck W, K Kashofer, S Rehman, A Daskalaki, D Berg, E Gralka, J Jozefczuk, K Drews, V Pandey, et al. (2015). Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes. Sci Data 2:150068.
32. McMullen PD, S Bhattacharya, CG Woods, B Sun, K Yarborough, SM Ross, ME Miller, MT McBride, EL LeCluyse, RA Clewell and ME Andersen. (2014). A map of the PPARalpha transcription regulatory network for primary human hepatocytes. Chem Biol Interact 209:14–24.
33. Grygiel-Gorniak B. (2014). Peroxisome proliferator-activated receptors and their ligands: nutritional and clinical implications—a review. Nutr J 13:17.
34. Pawlak M, P Lefebvre and B Staels. (2015). Molecular mechanism of PPARalpha action and its impact on lipid metabolism, inflammation and fibrosis in non-alcoholic fatty liver disease. J Hepatol 62:720–733.
35. Targett-Adams P, MJ McElwee, E Ehrenborg, MC Gustafsson, CN Palmer and J McLauchlan. (2005). A PPAR response element regulates transcription of the gene for human adipose differentiation-related protein. Biochim Biophys Acta 1728:95–104.
36. Pawella LM, M Hashani, E Eiteneuer, M Renner, R Bartenschlager, P Schirmacher and BK Straub. (2014). Perilipin discerns chronic from acute hepatocellular steatosis. J Hepatol 60:633–642.
37. Rogue A, S Antherieu, A Vluggens, T Umbdenstock, N Claude, C de la Moureyre-Spire, RJ Weaver and A Guillouzo. (2014). PPAR agonists reduce steatosis in oleic acid-overloaded HepaRG cells. Toxicol Appl Pharmacol 276:73–81.
38. Michalik L, J Auwerx, JP Berger, VK Chatterjee, CK Glass, FJ Gonzalez, PA Grimaldi, T Kadowaki, MA Lazar, et al. (2006). International Union of Pharmacology. LXI. Peroxisome proliferator-activated receptors. Pharmacol Rev 58:726–741.
39. Francque S, A Verrijken, S Caron, J Prawitt, R Paumelle, B Derudas, P Lefebvre, MR Taskinen, W Van Hul, et al. (2015). PPARalpha gene expression correlates with severity and histological treatment response in patients with non-alcoholic steatohepatitis. J Hepatol 63:164–173.
40. Fogarty S and DG Hardie. (2010). Development of protein kinase activators: AMPK as a target in metabolic disorders and cancer. Biochim Biophys Acta 1804:581–591.
41. Yamada PM, HH Mehta, D Hwang, KP Roos, AL Hevener and KW Lee. (2010). Evidence of a role for insulin-like growth factor binding protein (IGFBP)-3 in metabolic regulation. Endocrinology 151:5741–5750.
42. Asby DJ, F Cuda, M Beyaert, FD Houghton, FR Cagampang and A Tavassoli. (2015). AMPK activation via modulation of de novo purine biosynthesis with an inhibitor of ATIC homodimerization. Chem Biol 22:838–848.
43. Pirola CJ, T Fernandez Gianotti, GO Castano, P Mallardi, J San Martino, M Mora Gonzalez Lopez Ledesma, D Flichman, F Mirshahi, AJ Sanyal and S Sookoian. (2015). Circulating microRNA signature in non-alcoholic fatty liver disease: from serum non-coding RNAs to liver histology and disease pathogenesis. Gut 64:800–812.
44. Cermelli S, A Ruggieri, JA Marrero, GN Ioannou and L Beretta. (2011). Circulating microRNAs in patients with chronic hepatitis C and non-alcoholic fatty liver disease. PLoS One 6:e23937.
45. Hsu SH, B Wang, J Kota, J Yu, S Costinean, H Kutay, L Yu, S Bai, K La Perle, et al. (2012). Essential metabolic, anti-inflammatory, and anti-tumorigenic functions of miR-122 in liver. J Clin Invest 122:2871–2883.
46. Tsai WC, SD Hsu, CS Hsu, TC Lai, SJ Chen, R Shen, Y Huang, HC Chen, CH Lee, et al. (2012). MicroRNA-122 plays a critical role in liver homeostasis and hepatocarcinogenesis. J Clin Invest 122:2884–2897.
47. Tan W, Y Li, SG Lim and TM Tan. (2014). miR-106b-25/miR-17–92 clusters: polycistrons with oncogenic roles in hepatocellular carcinoma. World J Gastroenterol 20:5962–5972.
48. Jozefczuk J, K Kashofer, R Ummanni, F Henjes, S Rehman, S Geenen, W Wruck, C Regenbrecht, A Daskalaki, et al. (2012). A systems biology approach to deciphering the etiology of steatosis employing patient-derived dermal fibroblasts and iPS cells. Front Physiol 3:339.

Address correspondence to:
Prof. James Adjaye
Institute for Stem Cell Research
and Regenerative Medicine
Heinrich Heine University Düsseldorf
Moorenstrasse 5
Düsseldorf 40225
Germany

E-mail: james.adjaye@med.uni-duesseldorf.de

Received for publication December 16, 2015
Accepted after revision June 15, 2016
Prepublished on Liebert Instant Online June 16, 2016

2.2.4 IPSC-Derived Neuronal Cultures Carrying the Alzheimer's Disease Associated TREM2 R47H Variant Enables the Construction of an Aβ-Induced Gene Regulatory Network.

Int J Mol Sci. 2020 Jun 25;21(12). pii: E4516. doi: 10.3390/ijms21124516.
Martins S(1), Müller-Schiffmann A(2), Erichsen L(1), Bohndorf M(1), Wruck W(1), Sleegers K(3)(4), Van Broeckhoven C(3)(4), Korth C(2), Adjaye J(1).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine University, 40225 Düsseldorf, Germany. (2)Department of Neuropathology, Heinrich-Heine University, 40225 Düsseldorf, Germany. (3)Neurodegenerative Brain Diseases Group, VIB-Center for Molecular Neurology, University of Antwerp, 20610 Antwerp, Belgium. (4)Department of Biomedical Sciences, University of Antwerp, 20610 Antwerp, Belgium.

Genes associated with immune response and inflammation have been identified as genetic risk factors for late-onset Alzheimer's disease (LOAD). The rare R47H variant within triggering receptor expressed on myeloid cells 2 (TREM2) has been shown to increase the risk for developing Alzheimer's disease (AD) 2-3-fold. Here, we report the generation and characterization of a model of late-onset Alzheimer's disease (LOAD) using lymphoblast-derived induced pluripotent stem cells (iPSCs) from patients carrying the TREM2 R47H mutation, as well as from control individuals without dementia. All iPSCs efficiently differentiated into mature neuronal cultures, however AD neuronal cultures showed a distinct gene expression profile. Furthermore, manipulation of the iPSC-derived neuronal cultures with an Aβ-S8C dimer highlighted metabolic pathways, phagosome and immune response as the most perturbed pathways in AD neuronal cultures. Through the construction of an Aβ-induced gene regulatory network, we were able to identify an Aβ signature linked to protein processing in the endoplasmic reticulum (ER), which emphasized ER-stress, as a potential causal role in LOAD. Overall, this study has shown that our AD-iPSC based model can be used for in-depth studies to better understand the molecular mechanisms underlying the etiology of LOAD and provides new opportunities for screening of potential therapeutic targets.

DOI: 10.3390/ijms21124516 PMCID: PMC7350255 PMID: 32630447

URL: https://www.mdpi.com/1422-0067/21/12/4516

The data presented in this publication has been used also for the PhD thesis of Dr. Soraia Martins at the Heinrich Heine University Düsseldorf.

Contribution of Wasco Wruck: Formal analysis, Data curation, Writing—review and editing.

International Journal of
Molecular Sciences

MDPI

Article

IPSC-Derived Neuronal Cultures Carrying the Alzheimer's Disease Associated *TREM2* R47H Variant Enables the Construction of an Aβ-Induced Gene Regulatory Network

Soraia Martins [1], Andreas Müller-Schiffmann [2], Lars Erichsen [1], Martina Bohndorf [1], Wasco Wruck [1], Kristel Sleegers [3,4], Christine Van Broeckhoven [3,4], Carsten Korth [2] and James Adjaye [1,*]

1 Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine University, 40225 Düsseldorf, Germany; Soraia.Martins@med.uni-duesseldorf.de (S.M.); Lars.Erichsen@med.uni-duesseldorf.de (L.E.); Martina.Bohndorf@med.uni-duesseldorf.de (M.B.); Wasco.Wruck@med.uni-duesseldorf.de (W.W.)
2 Department of Neuropathology, Heinrich-Heine University, 40225 Düsseldorf, Germany; Andreas.Mueller@med.uni-duesseldorf.de (A.M.-S.); ckorth@uni-duesseldorf.de (C.K.)
3 Neurodegenerative Brain Diseases Group, VIB-Center for Molecular Neurology, University of Antwerp, 20610 Antwerp, Belgium; kristel.sleegers@uantwerpen.vib.be (K.S.); christine.vanbroeckhoven@uantwerpen.vib.be (C.V.B.)
4 Department of Biomedical Sciences, University of Antwerp, 20610 Antwerp, Belgium
* Correspondence: james.adjaye@med.uni-duesseldorf.de

Received: 27 May 2020; Accepted: 23 June 2020; Published: 25 June 2020

check for updates

Abstract: Genes associated with immune response and inflammation have been identified as genetic risk factors for late-onset Alzheimer´s disease (LOAD). The rare R47H variant within triggering receptor expressed on myeloid cells 2 (TREM2) has been shown to increase the risk for developing Alzheimer's disease (AD) 2–3-fold. Here, we report the generation and characterization of a model of late-onset Alzheimer's disease (LOAD) using lymphoblast-derived induced pluripotent stem cells (iPSCs) from patients carrying the TREM2 R47H mutation, as well as from control individuals without dementia. All iPSCs efficiently differentiated into mature neuronal cultures, however AD neuronal cultures showed a distinct gene expression profile. Furthermore, manipulation of the iPSC-derived neuronal cultures with an Aβ-S8C dimer highlighted metabolic pathways, phagosome and immune response as the most perturbed pathways in AD neuronal cultures. Through the construction of an Aβ-induced gene regulatory network, we were able to identify an Aβ signature linked to protein processing in the endoplasmic reticulum (ER), which emphasized ER-stress, as a potential causal role in LOAD. Overall, this study has shown that our AD-iPSC based model can be used for in-depth studies to better understand the molecular mechanisms underlying the etiology of LOAD and provides new opportunities for screening of potential therapeutic targets.

Keywords: late onset Alzheimer's disease; iPSC-derived neuronal cultures; TREM2 R47H; AβS8C dimer

1. Introduction

Currently, there are 47 million people worldwide living with dementia, a number that is estimated to increase to more than 131 million in 2050 [1]. Alzheimer's disease (AD) is a neurodegenerative disease and the most common and devastating cause of dementia, contributing to 60–70% of all cases [2]. AD is clinically characterized by a progressive decline of cognitive functions and, according to the classical

amyloid hypothesis two key molecules have been implicated in AD neuropathology: amyloid-beta (Aß) and the protein TAU [3]. Aß peptides are derived from sequential proteolytic cleavages of the amyloid precursor protein (APP). They form extracellular aggregated deposits known as amyloid plaques. Intracellularly, hyper-phosphorylated TAU forms aggregates composed of twisted filaments known as neurofibrillary tangles (NFTs). As a consequence of the imbalanced crosstalk between Aß and TAU, multiple neuropathological mechanism ensue, such as, synaptic toxicity, mitochondrial dysregulation and microglia-derived inflammatory responses, finally leading to neuronal death [4,5]. Age is the greatest risk factor for AD and it can divided it into early-onset AD (EOAD) when the patients are younger than 65, and late-onset AD (LOAD) when the patients manifest symptoms after the of age 65 [6]. Despite EOAD being responsible for a small minority of all cases, the studies of familiar AD patients (fAD) have revealed important aspects of the genetic factors implicated in the disease, such as the causal mutations in *APP*, *PSEN1* and *PSEN2*. On the other hand, LOAD is a very complex and multifactorial disease where most cases are sporadic with no clear familiar pattern of disease [7,8]. Many genetic risk factors have been implicated in increasing the susceptibility for LOAD, among which is the well establish apolipoprotein E (*APOE).* Individuals carrying one ε4 allele have a 3-fold increased risk of AD while individuals carrying the two ε4 alleles face an approximately 12-fold increased risk of AD [9,10]. More recently genome-wide association studies (GWAS) and large scale sequencing projects have led to the discovery of other genetic variants in more than 40 loci that influence the risk for LOAD [11–16]. These genes are known to be involved in biological pathways such as cholesterol metabolism, APP metabolism, MAPT metabolism, cytoskeleton and axon development, immune response and endocytosis/vesicle-mediated transport and epigenetics [17,18]. As a more direct link between immune responses and AD, especially microglia-related genes with an increased risk for developing LOAD were identified by high-throughput sequencing technologies [19,20]. One of multiple genetic risk variants identified in these studies is the rare p.Arg47His (R47H) variant within triggering receptor expressed on myeloid cells 2 (*TREM2*), which has been shown to increase the risk of developing AD by 2–3-fold in several European and North American populations [19–24].

TREM2 is a cell surface receptor of the immunoglobulin superfamily expressed on various cells of the myeloid linage including CNS microglia, bone osteoclasts, alveolar and peritoneal macrophages [25] According to neuropathology studies in AD patients, animal models and in vitro studies, the *TREM2* R47H variant induces a partial loss of function of TREM2, compromising microglia function and thus contributing to the development of AD. TREM2 deficiency in AD mouse models and patients carrying the R47H variant showed decreased clustering of microglia around the plaques, thereby facilitating the build-up of Aβ plaques and injury to adjacent neurons [26–29]. Recent data have shown that cells expressing the R47H variant displayed impaired TREM2-Aβ binding and altered TREM2 intracellular distribution and degradation, thus providing a potential mechanism by which *TREM2* R47H mutation increases the risk for LOAD [30,31]. The adoption of induced pluripotent stem cells (iPSCs) technology provides a platform to derive a reliable human disease model for better understanding the effect of risk factors in neurons derived from primary cells of affected patients. iPSC modeling of AD has provided an important proof-of-principle regarding the utility of such cells for a better understanding of the molecular mechanisms associated with the etiology of AD. So far, a number of the human iPSC-based AD models have concentrated on using iPSCs derived from EOAD or LOAD patients with unidentified mutations [32–38].

Here, we report for the first time the generation and characterization of a model of LOAD using lymphoblast-derived iPSCs from patients harboring the R47H mutation in *TREM2*, as well as from control individuals without dementia. To date gene regulatory networks governing LOAD have been generated using human AD brain biopsies. In our current study, we have shown the feasibility of using an iPSC-based approach to derive biologically meaningful pathways and an Aβ-induced regulatory network from neuronal cultures that mirrors some of the pathways that have been identified by the LOAD brain biopsies, namely immune response, phagocytosis and unfolded protein response pathways [39]. Our study thus demonstrates that AD iPSC-derived neuronal cultures can be used for

in depth studies to understand the molecular mechanisms underlying the onset of Alzheimer's disease and for screening of potential therapeutic targets.

2. Results

2.1. Ipscs Efficiently Differentiate into a Functional Neuronal Culture

iPSCs derived from lymphoblasts from two LOAD patients carrying the *TREM2* R47H risk variant (AD2-2 and AD2-4), as well as aged-matched control individuals without dementia (CON8 and CON9) were used for this study [40–43]. The summary of the characteristics of the iPSC lines used in this study as well as their APOE status are shown in Table 1.

Table 1. Summary of the healthy controls and Alzheimer's disease (AD) induced pluripotent stem cell (iPSC) lines used in this study.

iPSCs Name	Status	AD Risk Variant	Age	Age at Onset	Gender	APOE Genotype	Reference
CON8	Control individual	Control	69	-	M	3/4	[43]
CON9	Control individual	Control	75	-	F	3/3	[40]
AD-2-2	AD patient	TREM2 p.R47H heterozygous	65	60	M	4/4	[42]
AD-2-4	AD patient	TREM2 p.R47H heterozygous	67	64	F	2/4	[41]

It has been suggested that GABAergic neurotransmission plays a very important role in AD pathogenesis such as Aβ toxicity, hyperphosphorylation of TAU and the APOE effect [44–46]. In light of this information, we modified a previously established embryoid body-based protocol [47] to generate iPSC-derived neuronal cultures enriched in GABAergic interneurons. Figure 1A shows the timeline schematic for the protocol in which all iPSC lines were successfully differentiated into neuronal networks enriched in GABAergic interneurons within a course of 80 days (Figure S1). To qualitatively characterize the progression of differentiation, we performed immunostaining for various markers during the differentiation process. Neural rosettes expressed the progenitor markers PAX6 and Nestin (Figure 1B) and after being selected and grown as neurospheres for 7 days, the progenitor cells ($SOX1^+$) acquired predominantly a forebrain identity due to the expression of the medial ganglionic eminence (MGE) transcription factor NKX2.1 (Figure 1C), in addition to the telencephalic transcription factor FOXG1 (Figure 1D). After maturation, the neural cultures were composed of $GFAP^+$ glia cells and neurons expressing the pan-neuronal markers Tubulin Beta-III and MAP2 (Figure 1E). Neurons differentiated for 80 days expressed the maturation markers Synapsin I (SYN1) and neurofilaments (SMI-32; Figure 1F), as well as the neurotransmitter, GABA (Figure 1G). In order to assess the maturation status of the neuronal cultures, we performed RNA sequencing to analyze the transcriptome profile at day 80. Figure 1H shows a heat map of Pearson correlation analysis for key maturation neuronal markers together with the glia markers OLIG2 and GFAP in the iPSC-derived neuronal cultures compared to commercially bought RNA from fetal, adult and AD brain. All iPSC-derived neuronal cultures expressed similar levels of dopaminergic and serotonergic markers and higher levels of GABAergic interneuron markers. To complement and independently confirm these expression data, quantitative real-time PCR (qRT-PCR) analysis was carried out to evaluate the expression levels of GABAergic interneuron markers *PV, SOM, CALB2, GAD67* and *GAD65* (Figure S1). Despite the variability of expression levels of the different markers, we observed that the iPSC-derived neuronal cultures might be composed mostly of somatostain (SST) and calretinin (CALB2) subtypes of GABA

interneurons. Moreover, due to the low expression level observed for *TREM2* when compared with the commercially bought fetal, adult and AD brain RNA, qRT-PCR was performed for all iPSC-derived neuronal cultures. *TREM2* is expressed in all lines but however significantly upregulated in AD2-2 (Figure 1I). Taken together, we proposed (i) that the presence of the *TREM2* R47H variant in the AD2-2 and AD2-4 lines has no significant effect on the neuronal differentiation capacity when compared to the control lines CON8 and CON9, (ii) though we did not analyze our neuronal cultures for the presence of microglia, the mixed neuronal culture might probably harbor these.

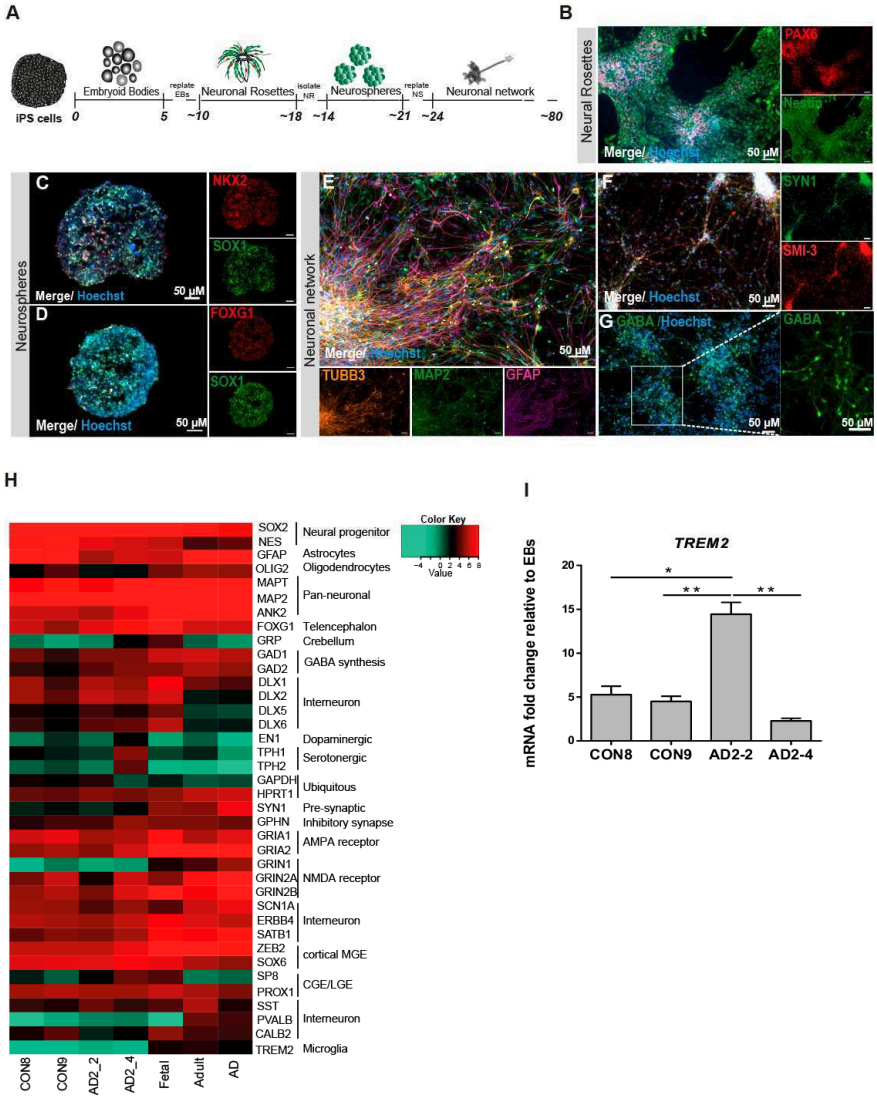


Figure 1. Differentiation and characterization of iPSC-derived neuronal cultures. (**A**) Scheme illustrating the main stages of the differentiation protocol for generating iPSC-derived neuronal network enriched in GABAergic interneurons. (B-G) Representative immunocytochemistry images of (**B**) neural rosettes

expressing the progenitor markers PAX6 (red) and Nestin (green), (**C**) neurosphere expressing the progenitor marker SOX1 (green) and the MGE marker NKX2.1 (red), (**D**) neurosphere expressing the progenitor marker SOX1 (green) and the forebrain marker FOXG1 (red), (**E**) neuronal network expressing the pan-neuronal markers TUBB3 (orange) and MAP2 (green) as well as GFAP (magenta), (**F**) neural maturation markers SYN1 (green) and SMI-3 (red) and (**G**) interneurons expressing the neurotransmitter GABA (green). Nuclei are stained with Hoechst. Scale bar, 50 µM. (**H**) Heatmap of Pearson correlation analysis of RNA-seq data from neural differentiation of control (CON8 and CON9) and AD lines (AD2_2 and AD2 _4) and commercially bought RNA from fetal, adult and AD brain for neural progenitor, early neuronal and mature dopaminergic, serotonergic, GABAergic interneuronal markers and glia markers. (**I**) Relative gene expression of TREM2 in iPSC-derived GABAergic interneurons network from control and AD lines shown as fold change relative to embryoid bodies (EBs). * $p < 0.05$, ** $p < 0.01$, one-way ANOVA, followed by Tukey´s multiple comparisons test. Data are presented as mean ± SEM from three independent experiments.

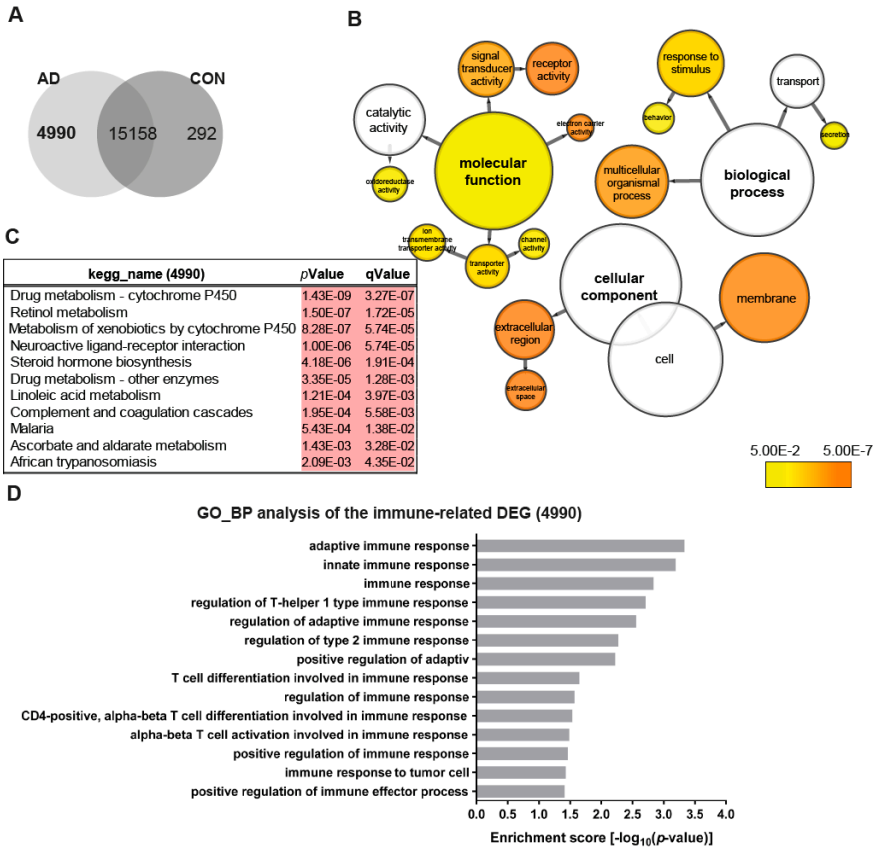
2.2. The AD Neuronal Network Shows a Distinct Gene Expression Associated with Metabolism and Immune-Related Pathways

To obtain an overview of the transcriptome changes between the AD (AD2-2 and AD2-4) and the control (CON8 and CON9) neuronal cultures, we screened for differentially expressed genes (DEGs). Employing RNA-seq, we identified 4990 genes exclusively expressed in the AD neuronal cultures (Figure 2A). BiNGO was used to perform gene ontology (GO) term enrichment analysis of the 4990 genes, the results are illustrated as a tree-like structure (Figure 2B, Table S1). In depth analyses of the cellular component identified significant enrichment associated with membrane and extracellular space. Regarding biological processes, these genes were significantly enriched in processes related to the response to the stimulus and transport. Moreover, molecular functions such as signal transducer activity, receptor activity and transporter activity, including ion membrane transporter activity and channel activity were significantly enriched. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis revealed metabolic pathways, which include drug metabolism—cytochrome P450, retinol metabolism and steroid hormone biosynthesis together with a neuroactive ligand–receptor interactor (Figure 1C). As it has been shown that TREM2 regulates innate immunity in AD [48], we additionally analyzed GO terms for biological processes of immune-related genes within the 4990 gene set. Remarkably, 14 significantly enriched terms associated with the regulation of innate and adaptive immune response were identified (Figure 2D). Overall, these data may suggest that AD neuronal cultures exhibit alterations in key signaling pathways related to metabolism and the immune system.

2.3. Characterization of AD Hallmarks in CON and AD Neuronal Cultures

Numerous evidence support the notion that the small oligomers of $A\beta_{42}$ are intricately associated with the amyloid cascade [49,50]. However, recent studies have shown that Aβ dimers, abundantly detected in brains of AD patients, are sufficient to account for neurotoxicity and initiating the amyloid cascade [51–54]. Here, we aimed at investigating the effects of the *TREM2* R47H mutation in Aβ production as well as the response of CON and AD iPSC-derived neuronal cultures to stimulations with the well described Aβ-S8C dimer [55–57]. After 4 months of differentiation, neurospheres were dissociated into single cells and differentiated for a further 6 weeks. Aβ levels were measured and the neuronal cultures were stimulated with 500 nM of the Aβ-S8C dimer for 72 h (Figure 3A). Conditioned media from the non-stimulated CON and AD lines were analyzed for comparative $A\beta_{40}$ and $A\beta_{42}$ levels employing ELISA. Interestingly, neurons derived from the AD iPSCs lines (AD2-2 and AD2-4) and the CON iPSCs lines (CON8 and CON9) secreted Aβ with similar $A\beta_{42}$ ratio (Figure 3B-D). We further performed Western blot analysis to evaluate the levels of TAU phosphorylation at Ser202/Thr205 (AT8 epitope), total TAU and total APP after stimulation with the Aβ-S8C dimer (Figure 3E). Although phosphorylation of TAU was found in all neuronal cultures, no significant differences in the expression levels of total TAU (Figure 3F) and phosphorylated TAU (Figure 3G) were observed between AD and

CON neuronal cultures after treatment with the Aβ-S8C dimer. Surprisingly, the results revealed that stimulation with the Aβ-S8C dimer induced a modest and uniform increase in the expression levels of APP in all CON and AD neuronal cultures (Figure 3H). To focus on the effect of the Aβ-S8C dimer, we quantified APP levels in pooled samples, and this revealed significantly increased APP expression (Figure 3I). Taken together, these results confirm that the neuronal cultures (CON8 and AD) secrete Aβ and although no significant differences in the expression of total and phosphorylated TAU were observed, APP expression was significantly elevated after Aβ-S8C dimer stimulation. We therefore conclude that the CON and the AD iPSC-derived neuronal cultures were capable of recapitulating in vitro the hallmarks of AD-like cellular pathology.



kegg_name (4990)	*p*Value	qValue
Drug metabolism - cytochrome P450	1.43E-09	3.27E-07
Retinol metabolism	1.50E-07	1.72E-05
Metabolism of xenobiotics by cytochrome P450	8.28E-07	5.74E-05
Neuroactive ligand-receptor interaction	1.00E-06	5.74E-05
Steroid hormone biosynthesis	4.18E-06	1.91E-04
Drug metabolism - other enzymes	3.35E-05	1.28E-03
Linoleic acid metabolism	1.21E-04	3.97E-03
Complement and coagulation cascades	1.95E-04	5.58E-03
Malaria	5.43E-04	1.38E-02
Ascorbate and aldarate metabolism	1.43E-03	3.28E-02
African trypanosomiasis	2.09E-03	4.35E-02

Figure 2. Distinct gene expression profiles associated with AD neuronal networks. (**A**) Venn diagram illustrating genes exclusively expressed in the AD neural network (4990), the control (CON) network (292) or common between both (intersection -15158) (detection *p*-value < 0.05). (**B**) BiNGO analysis of the differentially expressed genes (DEGs; 4990) exclusively expressed in the TREM2 neuronal network (4990). The orange color of the circles correspond to the level of significance of the over-represented gene ontology (GO) category and the size of the circles is proportional to the number of genes in the category (*p*-value < 0.05). (**C**) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the genes exclusively expressed in the AD neuronal network (4990). (**D**) Significantly enriched gene ontology (GO) terms (biological processes) of the genes exclusively expressed in the AD neuronal network (4990) associated with immune system processes (*p*-value < 0.05).

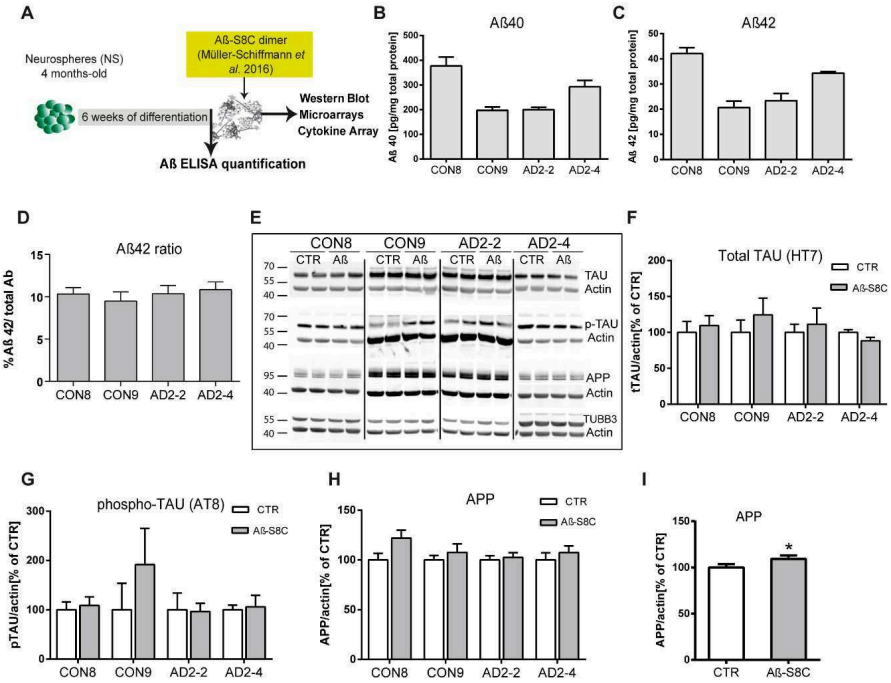


Figure 3. Stimulation of iPSC-derived neuronal cultures with the Aβ-S8C dimer. (**A**) Scheme illustrating the approach. Neurospheres were maintained for 4 months in culture, dissociated into single cells, differentiated for 6 weeks then stimulated with 500 nM of Aβ-S8C dimer for 72 h. Western blotting, microarrays and cytokine arrays were performed. (B-D) ELISA quantification of (**B**) total Aβ40, (**C**) total Aβ42 levels and (**D**) Aβ42/Aβ40+42 ratio from media collected from the interneuronal network and normalized to the total protein content. All data are presented as mean ± SEM from six independent experiments. (**E**) Representative Western blot images of endogenous TAU, phosphorylated TAU (Ser 202 and Thr 205), APP and the neural differentiation marker βIII-Tubulin after stimulation with 500 nM of the Aβ-S8C dimer. β-ACTIN was used as a loading control. (F-H) Quantification of (**F**) total TAU, (**G**) phosphorylated TAU and (**H**) APP levels. Results are normalized against β-ACTIN and shown as a percentage of control (CTR). All data are presented as mean ± SEM from 3 independent experiments. (**I**). Effect of the Aβ-S8C dimer on APP levels in iPSCs derived neuronal network (CON8, CON9, AD2-2 and AD2-4) compared to control. Data are presented as mean ± SEM from 3 independent experiments from 4 biological replicates. * $p < 0.05$, one-tail t-test versus control.

2.4. The Aβ-S8C Dimer Induces Metabolic Dysregulation in AD Neuronal Cultures

To assess the impact of the Aβ-S8C dimer on the gene expression profiles of CON and AD iPSC-derived neuronal cultures, we performed transcriptome analysis of CON8 and AD2-4 iPSC-derived neuronal cultures before and post stimulation with the Aβ-S8C dimer. This analysis identified differential expressed genes (DEGs) between the control and Aβ-S8C dimer treatment. Hierarchical cluster analysis revealed a clear separation of CON8 and AD-TREM2-4 iPSC-derived neuronal cultures (Figure 4A). Remarkably, CON8_Aβ clustered separately from AD2-4_Aβ, therefore implying that genetic background effects were more pronounced than the response elicited by the Aβ-S8C dimer.

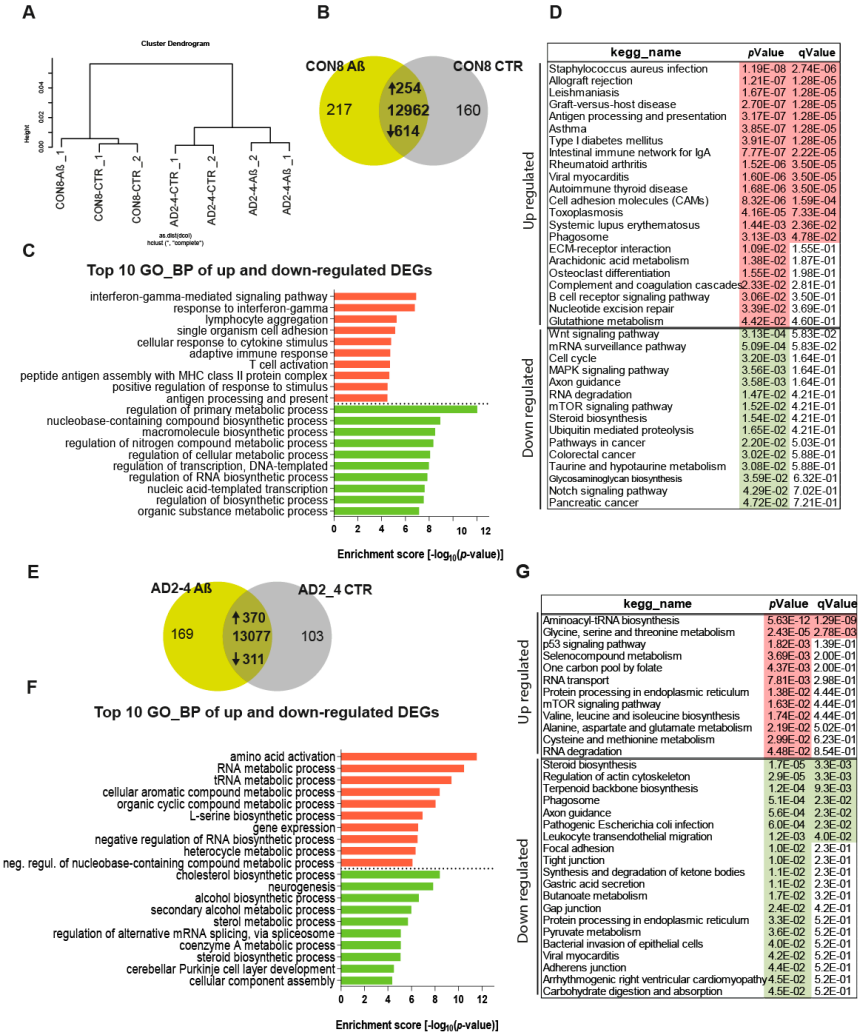


Figure 4. Gene expression profiles of the CON8 and AD neural cultures stimulated with the Aβ-S8C dimer. (**A**) Dendrogram obtained by hierarchical cluster analysis of microarray-based (Affymetrix) gene expression data for CON8 and AD2-4 stimulated with the Aβ-S8C peptide. Transcriptomes of CON8_CTR cluster with CON8_ Aβ while those of the AD2-4_CTR cluster separately with TREM2-4_ Aβ. (**B**) Venn diagram showing genes expressed only in the CON8 neural network subjected to Aβ-S8C peptide stimulation (green), the control condition (grey) and common to both conditions (intersection; detection p value < 0.05). (**C**) Top 10 significantly enriched gene ontology (GO) terms for biological processes of DEGs upregulated (254-red) and downregulated (614-green) subjected to Aβ-S8C peptide stimulation in CON8 neuronal cultures (p-value < 0.05). (**D**) KEGG enrichment analysis of up- and downregulated DEGs (p-value < 0.05). (**E**) Venn diagram showing genes expressed only in the AD2_4 neuronal culture when stimulated with the Aβ-S8C peptide (green), the control condition (grey) and common to both conditions (intersection; detection p value < 0.05). (**F**) Top 10 significantly enriched gene ontology (GO) terms for biological processes of DEGs upregulated (370-red) and downregulated (311-green) subjected to Aβ-S8C peptide stimulation in the AD2_4 neuronal cultures (p-value < 0.05). (**G**) KEGG enrichment analysis of up- and downregulated DEGs (p-value < 0.05).

Evaluation of DEGs in CON8 neuronal cultures before and after stimulation with the Aβ-S8C dimer identified 868 genes (Figure 4B), 254 were upregulated and 614 downregulated (Table S1). Figure 4C shows the related Top10 GO BP (biological processes) terms. Upregulated genes in CON8_Aβ were significantly enriched for GO terms such as interferon-gamma-mediated signaling pathway and cellular response to cytokine stimulus. In contrast, the downregulated genes in CON8_Aβ in comparison to control were associated with the GO terms, regulation of primary metabolic processes and regulation of RNA biosynthetic process. In agreement with the GO analysis, KEGG pathway analysis for the same set of genes revealed the upregulated genes in CON8_ Aβ to be associated in pathways related to inflammatory responses, for example, *Staphylococcus aureus* infection and antigen processing and presentation. In addition, CON8_Aβ also showed upregulation of the phagosome pathway, while Wnt signaling pathway and axon guidance were among the downregulated KEGG pathways (Figure 4D, Table S1).

Focusing on AD neuronal cultures, 681 DEGs were identified when comparing Aβ-S8C dimer stimulated and non-stimulated AD2-4 neuronal cultures (Figure 4F), of these 370 were upregulated and 311 downregulated (Table S2). Figure 4F lists the Top 10 GO BP terms. The upregulated genes in AD2-4_Aβ were significantly enriched for amino acid activation and RNA metabolic process. In contrast, the downregulated genes were associated amongst others with cholesterol biosynthetic process and neurogenesis. KEGG pathway analysis revealed upregulation of pathways such as glycine, serine and threonine metabolism, p53 signaling pathway and mTOR signaling pathway. Surprisingly, in contrast to CON8_Aβ, AD2-4 neuronal cultures stimulated with Aβ-S8C dimers showed down-regulation of the phagosome pathway (Figure 4G, Table S2). Taken together, these results imply that the AD2-4 neuronal cultures respond in a unique way to Aβ-S8C dimer stimulation, namely a metabolic dysregulation in contrast to an inflammatory response, which could be observed in the CON8 neuronal cultures.

2.5. Aβ-S8C Dimer Stimulation of the AD Neuronal Culture Revealed Indications of Impaired Phagocytosis-Related Pathway

TREM2 is crucial for regulating phagocytosis in microglia and the effect in phagocytosis by the AD-associated *TREM2* mutations have recently been a focus of studies [58–63]. As described above, phagocytosis appeared as a significantly upregulated pathway in CON8 but was downregulated in AD2-4 neuronal cultures after Aβ stimulation. We then analyzed differential expression of genes associated with this pathway. Figure 5A depicts the KEGG annotated phagosome pathway with upregulated genes in CON8_Aβ (red) and those downregulated in AD2-4_Aβ (green). After stimulation with the Aβ-S8C dimer, CON8 induced upregulation *of HLA-DMA, HLA-DMB. HLA-DOA, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRB1* and *HLA-F*, all genes associated with the Major Histocompatibility complex II (MHCII). In contrast to the AD2-4 non-stimulated cultures, stimulation with the Aβ-S8C dimer induced down-regulation of *TUBB4A, TUBB4B, DYNC1H1, LAMP2, ATP6V1A, ACTB, THBS1, CALR* and *TUBBA1C*. Table S3 shows the relative mRNA expression, from which the expression of *CALR, DYNC1H, LAMP2, HLA-DOA* and *HLA-DQB1* was confirmed by RT-PCR (Figure 5B). Taken together, these results suggest that neuronal cultures harboring the *TREM2* R47H variant but not controls likely undergo an impaired phagocytosis response in the presence of the Aβ-S8C dimer.

2.6. AD Neuronal Cultures Show a Compromised Inflammatory Response-Related Gene Expression Pattern upon Stimulation with the Aβ-S8C Dimer

Based on the fact that dysregulated cytokine production from microglia, astrocytes and neurons are associated with the development of AD [64], we analyzed the cytokine expression profile as well as the secretion profile from the AD neuronal cultures after stimulation with the Aβ-S8C dimer. Employing microarray-based global gene expression data, a heatmap-based analysis of 100 key cytokines (extracted from the Proteome Profiler Human XL Cytokine Array, R&D systems) revealed that stimulation with the Aβ-S8C dimer induced transcriptional changes in a subset of these genes in AD-TREM2-4 (Figure 6A). Interestingly, the AD2-2 neuronal culture showed down-regulation of

cytokines, chemokines and acute phase genes such as *IL1RL1*, *IL13*, *IL15*, *IL16*, *IL27*, *IL32*, *CXCL10*, *CXCL11*, *TFRC*, *SERPINE1*, *C5*, *THBS1*, *RLN2*, *SPP1*, *EGF*, *LIF*, *GC*, *BSG*, *MPO*, *CST3*, *FLT3LG* and *CCL20*. Surprisingly, only *IGFBP2*, *RBP4*, *VEGFA*, *CXCL5*, *IL19* and *TDGF1* had higher expression levels after Aβ-S8C dimer stimulation when compared to the control samples. We next aimed at determining if stimulation with the Aβ-S8C dimer could also alter the secretion of cytokines and chemokines in the AD neuronal cultures. To this end, we collected the cell culture supernatants from the AD2-2 and AD2-4 neuronal cultures 72 h post stimulation with the Aβ-S8C dimer and from non-stimulated controls. Thereafter, we carried out secretome analysis employing the proteome profiler cytokine array (Figure 6B). In agreement with the previous results, the level of secretion of all cytokines and chemokines decreased after Aβ-S8C dimer stimulation when compared to control (Figure 6C), with the exception of ICAM-1, MIF and SerpinE1. Taken together, these results might imply that AD neuronal cultures compromise the efficient activation of the inflammasome pathway in response to Aβ-S8C dimer stimulation.

2.7. A Protein–Protein Interaction (PPI) Network Identifies an AD-Depended Aβ-S8C Signature

To gain insights into a probable gene expression signature triggered by the Aβ-S8C dimer in LOAD, we focused on genes exclusively expressed in the AD neuronal culture after stimulation with the Aβ-S8C dimer. A Venn diagram analysis revealed that most (12687) genes were expressed in common between CON and AD with and without Aβ-S8C dimer stimulation (Figure 7A, Table S4). However, 95 genes were exclusively expressed in AD neuronal cultures stimulated with Aβ-S8C dimer. GO analysis (Figure 7B) unveiled several terms related to neuron and immune-system related processes including stimulatory C-type lectin receptor signaling pathway as most significant. Pathway analysis of the 95 AD Aβ-S8C genes (Figure 7C) revealed neuroactive ligand-receptor interaction as the most significant pathway and metabolic pathways with the higher number of genes. The 95 genes were further analyzed in a protein–protein interaction network (PPI) based on interactions from the BioGrid database resulting in a network containing APP and a big hub centered around *HSPA5*, which encodes the endoplasmic reticulum chaperone BiP (Figure 7D). HSPA5 has been reported to control the activation of the unfolding protein response (UPR), a pro-survival pathway in response to ER stress caused by misfolded proteins. Since there is evidence that the ER stress response, namely the UPR plays a role in the pathogenesis of AD [65], we took a deeper look into the GO terms related to ER after Aβ-S8C dimer stimulation (Table S2, highlighted in yellow). We observed that the Aβ-S8C dimer triggered an ER stress response, which elevated the expression of *ATF3* and *DDIT* in both CON and AD. Interestingly, the ER stress response was more prominent in the AD neuronal cultures, where several genes from the UPR were upregulated (*XBP1*, *AT4*, *PUMA* and *HERPUD1*) in contrast to *HSPA5* and *CALR*, which were downregulated (Figure 8, Table S5). These results highlighted the unique response triggered by the Aβ-S8C dimer in the AD neuronal cultures. By generating a PPI network we were able to link the Aβ-S8C signature genes to ER-stress, namely the activation of UPR.

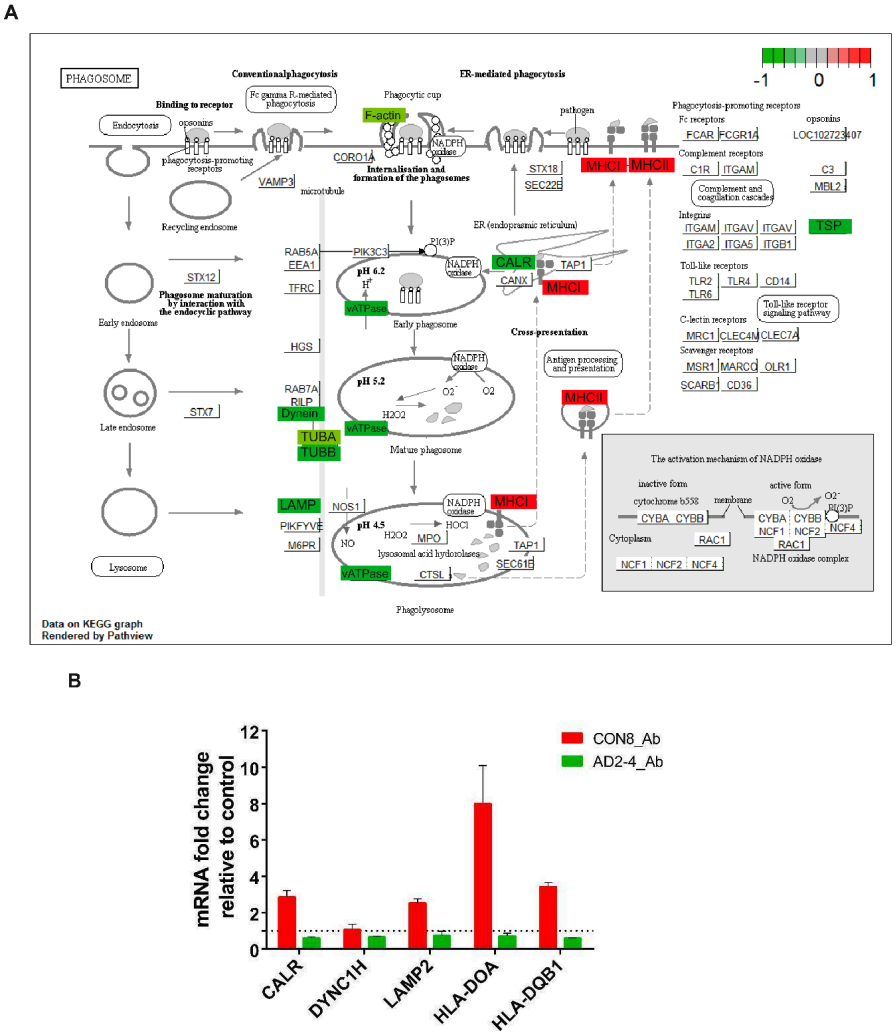


Figure 5. Representation of the KEGG phagosome pathway. (**A**) Upregulated DEGs genes in response to Aβ-S8C peptide stimulation in CON8 neuronal cultures are shown as red boxes and downregulated DEGs in AD2-4 neuronal cultures are shown as green boxes. (**B**) Relative gene expression of *CALR, DYNC1H, LAMP2, HLA-DOA* and *HLA-DQB1* analyzed by RT-PCR. Data are presented as mean ± SEM from two independent experiments.

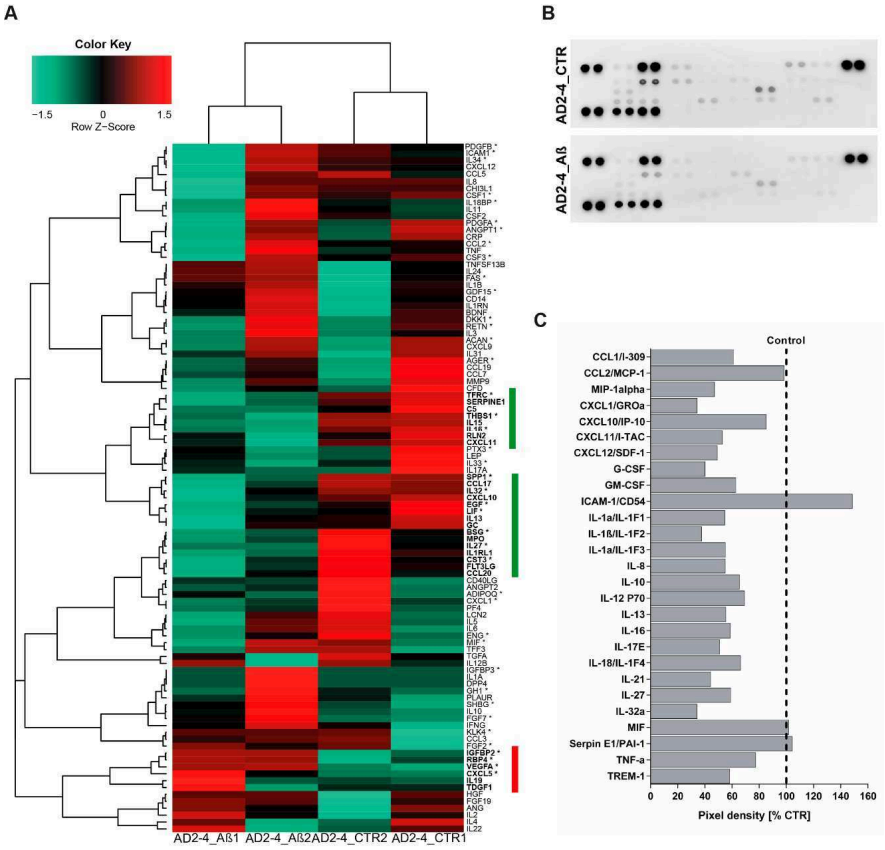


Figure 6. Analysis of cytokine expression and secreted factors in AD neuronal cultures upon Aβ-S8C stimulation. (**A**) Heatmap of Pearson correlation analysis of microarrays data from AD neural differentiation under Aβ-S8C stimulation (AD2-4_ Aβ1 and AD2-4_ Aβ2) or control (AD2-4_ CTR1 and AD2-4_ CTR2) showing the differential expression of cytokines. The highlighted genes in green represent a cluster of cytokines downregulated upon Aβ-S8C stimulation whereas the highlighted genes in red represent an upregulated cluster of cytokines. (**B**) Human cytokine array showing the effect of the Aβ-S8C peptide on the secreted factors of neuronal cultures from pooled AD2-2 and AD2-4 culture supernatants of control condition and 72 h of Aβ-S8C stimulation. (**C**) Quantitative analysis of the secreted factors shows that Aβ-S8C treatment decreases the amount of secreted cytokines in AD neuronal cultures.

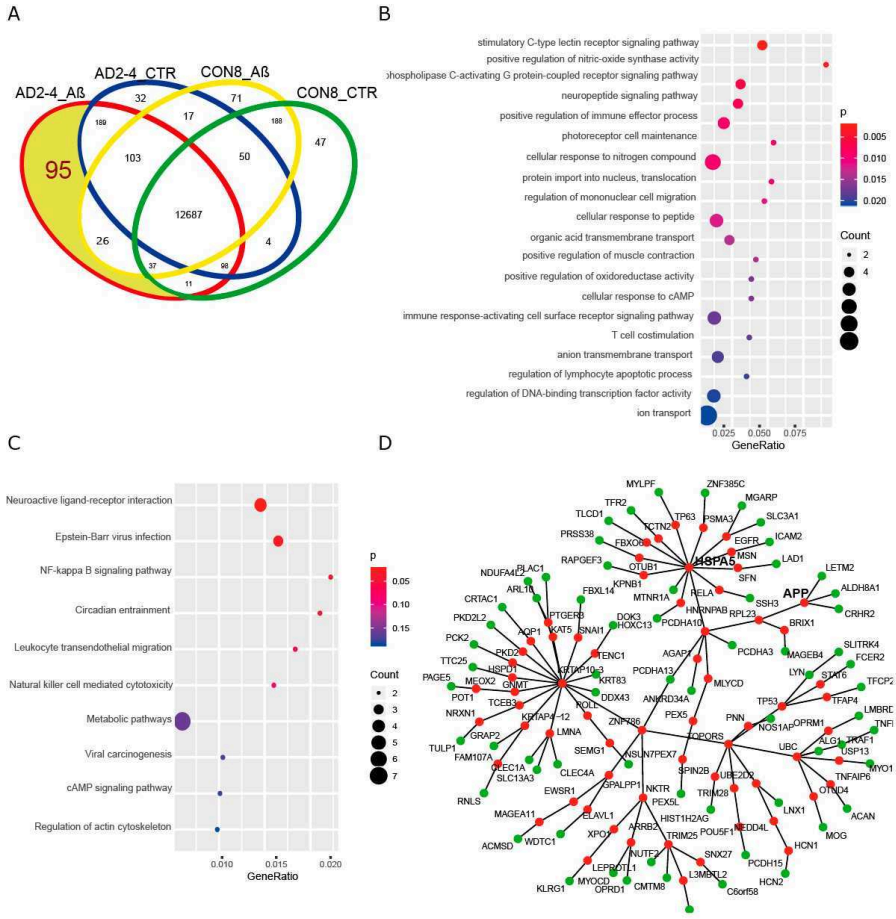


Figure 7. Aβ-S8C stimulated AD neuronal cultures activate a protein-protein interaction network, which includes APP and HSPA5. (**A**) Venn diagram dissecting 95 genes expressed in Aβ-S8C stimulated AD neuronal cultures from genes expressed in AD control and healthy neuronal cultures with and without Aβ-S8C stimulation. (**B**) Dot plot of gene ontologies (biological process) overrepresented in the 95 AD_Aβ genes. (**C**) Dot plot of KEGG pathways overrepresented in the 95 AD_Aβ genes. (**D**) Protein–protein interaction network derived from the 95 AD_Aβ genes with APP and HSPA5. Nodes from the 95 genes are colored green and the nodes added using the Biogrid database to connect the network are colored red.

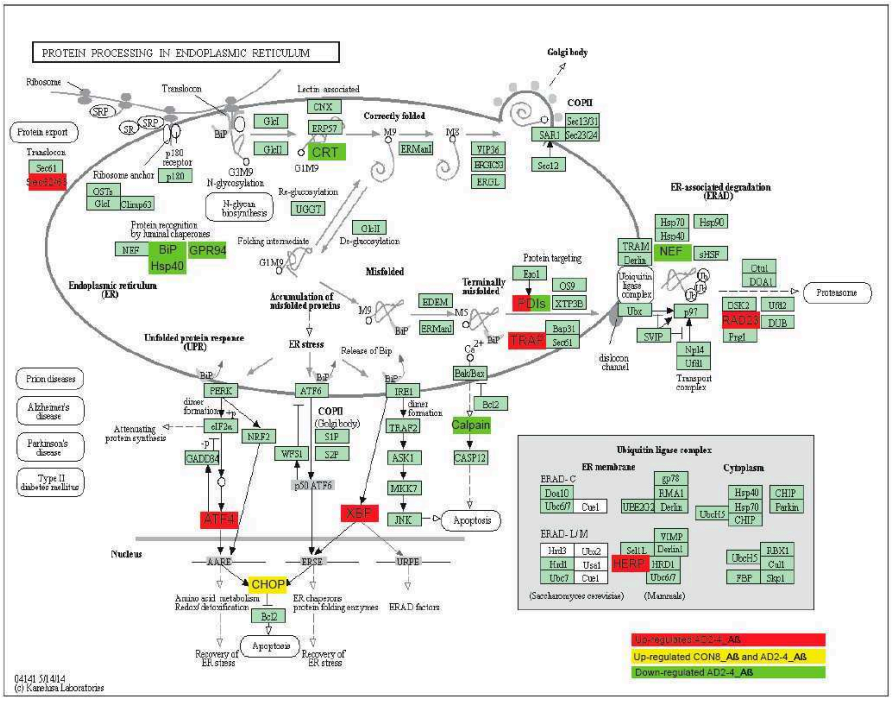


Figure 8. Representation of the KEGG protein process in endoplasmic reticulum pathway. Upregulated DEGs genes in response to Aβ-S8C stimulation in AD2-4 neuronal cultures are shown as red boxes, upregulated DEGs genes in response to Aβ-S8C stimulation in CON 8 and AD2-4 neuronal cultures are shown in yellow boxes and downregulated DEGs in AD2-4 neuronal cultures are shown as green boxes.

3. Discussion

While the mechanisms underlying the etiology of AD have been a focus of study over several decades, the current knowledge about the etiology and pathogenesis of AD are still incomplete. The use of primary neurons from animal models and immortalized cell lines based on modifications in *APP*, *PSEN1* and *PSEN2* has provided some insights into EOAD. While these models are helpful for studying a specific causal mutation (EOAD), there are several hurdles and limitations associated with studying LOAD, which requires the endogenous expression of genetic mutations and their genetic interactions. Understanding the biological implications of the recently identified genetic risk variants, namely the R47H substitution in *TREM2*, is essential to enable the establishment of genotype–phenotype correlations, which can lead to potential novel therapeutic approaches. The breakthrough development of iPSCs technology provides the most applicable tool to create an in vitro sporadic patient-derived model. Although modeling AD using patient-derived iPSCs has been prominent, a handful of studies to date have generated and characterized iPSC-derived neuronal cultures from LOAD patients [32,33,35,37,38]. This is the first study describing the generation and characterization of a model of LOAD based on Aβ dimer stimulated neuronal cultures originating from lymphoblast-derived iPSCs derived from LOAD patients carrying the missense mutation R47H in *TREM2*.

First, we differentiated the iPSCs to neurons using a modified protocol described by Liu et al., 2013 and analyzed the distinct progression steps during the differentiation process. Transcriptome analysis and immunocytochemistry confirmed the ability of our modified protocol to derive neurons and glia cells within our neuronal cultures. Based on gene expression comparison between the iPSC-neuronal

cultures with commercially bought fetal, adult and AD brain RNA we could show that our cells expressed the expected maturation markers. Thus, our results imply that lymphoblast-derived iPSCs from LOAD patients and healthy donors can be robustly differentiated into neuronal cultures. Moreover, we did not observe profound differences in the differentiation and maturation propensity between iPSCs derived from LOAD patients and healthy donors, in agreement with previous reports [33,35,37,38]. Cheng-Hathaway et al. and Sudom et al. reported that *Trem2* R47H knock-in mice showed reduced *Trem2* mRNA and protein expression in the brain as well as reduced soluble fragments of Trem2 (sTrem2) in plasma [66,67]. More recently, Xiang et al. reported that a mouse-specific splicing caused this reduction and *TREM2* mRNA levels were normal in both iPSC-derived microglia and in patient brains with the *TREM2* R47H variant [68]. We therefore evaluated if T*REM2* expression was different in LOAD patients carrying the *TREM2* R47H variant compared to the control. Our results indicate that *TREM2* mRNA was significantly upregulated in AD2-2 but not in the AD2-4 neuronal cultures compared to the control. This ambiguity is probably due to the limitations imposed by our small sample size. In addition, our neuronal culture is composed mainly of neurons so the TREM2 positive cells are in low abundance.

Although the neuronal cultures derived from LOAD patients and healthy donors did not exhibit differences in morphology or expression of differentiation markers, transcriptome analysis showed a distinct profile. Interestingly, GO analysis revealed that the proteins encoded by the 4990 genes exclusively expressed in AD neuronal cultures were predominantly mapped in the cell membrane and in the extracellular space. These genes were involved in (i) biological processes (BP) terms such as response to stimulus and secretion and (ii) molecular functions (MF) terms such as signal transducer activity, receptor activity, transporter activity, channel activity and ion transmembrane transporter activity. As part of these exclusively expressed genes we also identified genes of the matrix metalloproteinases (MMPs) family, for example *MMP2* and *MMP9*. Metalloproteinases play an important role in the pathogenesis of AD. While MMP2 might have a protective role, MMP9 expression, which is increased in AD patients, is induced by Aβ and it can influence TAU aggregation [69]. Furthermore, members of the ATP-binding cassette (ABC) and the solute carrier (SLC) families were over-represented in the GO_MF. ABC transporters have been implicated in AD pathophysiology, associated with processes leading to the accumulation of Aβ in the CNS. Importantly, we observed the exclusively expression of GLUT4 (*SLC2A4)*, a crucial insulin sensitive glucose transporter upregulated in AD patients, which is responsible for regulating glucose metabolism in neurons [70,71]. As anticipated, we also identified GO terms related to the regulation of the innate and adaptive immune response as significantly enriched. Implications of these results are that our AD neuronal cultures show a distinct signal transducer and transporter activity that may contribute to metabolic alterations, to an inadequate immune response and ultimately to neurotoxicity. According to the amyloid cascade theory, accumulation of Aβ plays a key role in triggering the cascade of events underlying the pathogenesis of EOAD. However some studies have shown that Aβ secretion is not altered in LOAD-derived neurons [37,72]. In accordance, our results show that AD cultures secreted Aβ with a similar $A\beta_{1\text{-}42}$ to $A\beta_{1\text{-}40}$ ratio as the control. Nonetheless, the levels of $A\beta_{1\text{-}40}$ and $A\beta_{1\text{-}42}$ were highly reproducible across multiple differentiations (six) and lines (four), thus establishing our cell culture model as robust for manipulating the production of Aβ. We subsequently aimed at evaluating the potential effect of Aβ in our neuronal cultures in order to close the gap in our understanding of the mechanisms that are underlying the early stages of AD. Aβ-S8C dimer can induce neurotoxicity and abnormal synaptic signaling, together with impaired cognitive functions in the absence of plaque pathology, thus mimicking the early stages of AD [55]. Aβ-induced TAU hyper-phosphorylation has been described to initiate the signaling cascade alterations that culminate in NFT formation and neuronal degeneration [73]. Phosphorylation of the AT8 epitope (Ser202/Thr205) has been found to be elevated in sAD-derived neurons [37]. We were not able to detect an increase in phosphorylation of TAU at Ser202/Thr205 upon Aβ-S8C dimer treatment. In addition, there were no differences in the levels of phosphorylation detected between control and AD neuronal cultures in the non-stimulated conditions. TAU can be phosphorylated on more than 80 residues,

and it is known that Ser422 is phosphorylated earlier than Ser202/Thr205 during NFTs formation [74]. Based on these facts and the results obtained, we can assume that the duration of incubation of the Aβ-S8C dimer was presumably not long enough to detect increased phosphorylation at Ser202/Thr205. Our results show that independent of the genetic background, incubation with the Aβ-S8C dimer increased the levels of total APP. A more in-depth analysis of APP processing will provide more insights into the pathogenic role of the *TREM2* R47H variant in EOAD.

In addition to interfering with total APP levels, Aβ-S8C dimer stimulation induced a remarkable and significant transcriptome change in the control as well as in the AD neuronal cultures. Annotation and enrichment analysis revealed that the upregulated DEGs induced by Aβ-S8C stimulation in the control neuronal cultures are related to immune system activation (interferon-gamma-mediated signaling pathway, cellular response to cytokine stimulus and adaptive immune response). Aβ soluble species have also been linked to an attenuation of the Wnt signaling pathway, in addition to putative effects on cell cycle, contributing to synaptic dysfunction and neurodegeneration. In accord, our data revealed that Wnt signaling and cell cycle were downregulated after Aβ-S8C stimulation in the control neuronal cultures. On the contrary, the AD neuronal cultures responded in a completely different manner to stimulation with the Aβ-S8C dimer.

The effect of the AD-associated TREM2 mutations on phagocytosis is an active area of study but so far variable results have been obtained. While R47H transduced HEK cells displayed a reduced up-take of latex beads and $A\beta_{1\text{-}42}$, no changes were observed in the fluorescent pH-sensitive rhodamine *Escherichia coli* (pHrodo-linked *E. coli)* uptake assay [61]. Additionally, *TREM2*$^{+/R47H}$ transdifferentiated microglia-like cells [58] and microglia-like cells derived from *TREM2* T66M $^{+/-}$, T66M $^{-/-}$ and W50C$^{-/-}$ hPSCs, also showed no defects in the *E. coli* uptake [59,63]. However, Piers et al. showed that iPSC-derived microglia harboring the *TREM2* R47H mutation exhibit a substantial deficil in the ability to phagocytose β-Amyloid [75]. We found that the control neuronal cultures upregulated the phagosome pathway after Aβ-S8C stimulation, namely the genes associated with MHCII. These observations are in line with previous reports where incubation with Aβ led to an accumulation of MHC-II and AD patients also showed upregulation of MHC-II [76]. On the contrary, these genes were not differentially regulated in our AD neuronal cultures, but interestingly other genes associated with the phagosome pathway were downregulated. Calreticulin is encoded by the *CALR* gene. It is an endoplasmic reticulum protein that interacts with Aβ, and is considered as a scavenger for $A\beta_{1\text{-}42}$ [77]. Low levels of calreticulin have been observed in AD brains, and it has been suggested that this down-regulation can lead to the pathological processes of AD [78]. Notably, the levels of tubulins *TUBB4A* and *TUBB4B* were downregulated, supported by Hondius et al., where the levels of these tubulins identified by mass spectrometry analysis in human post-mortem brain tissue were significantly decreased over the progressive stages of AD [79]. On the other hand, lysosome-associated membrane protein 2 (LAMP-2) together with other lysosome-related proteins was found to be increased in CSF from AD patients [80]. Interestingly, *LAMP2* was downregulated in our AD neuronal cultures leading us to speculate that R47H AD carriers have a unique response to phagocytosis, probably due to the partial loss of function of TREM2 activity.

The analysis of pro-inflammatory cytokines at the levels of mRNA and the secretome of the AD neuronal cultures in response to Aβ-S8C stimulation are of particular interest. Although the mRNA expression of the *IL-1β*, *IL-6*, *TNF-α* and *MIP-1α* proinflammatory cytokines was upregulated in some of the samples, the secretion of these cytokines was downregulated. A recent study using iPSC-derived microglia-like cells from patients carrying the T66M and W50C missense mutation within *TREM2* showed that these cells have a deficit in the cytokine release [63]. Indeed, *SPP1* and *GPNMB*, encoding osteopontin and osteoactivin, were also downregulated in AD neuronal cultures and not in the control after Aβ-S8C stimulation. SPP1 and GPNMB are microglia activation-related transcripts that are upregulated in AD models and associated with Aβ accumulation. In support of our data, it was recently reported that SPP1 and GPNMB reflect TREM2 signaling and the expression is highly sensitive to the R47H variant [26]. Interestingly, there was a cluster of genes associated with insulin

resistance, which was upregulated after Aβ-S8C stimulation. Increased levels of RBP4 were found in APP/PSEN1 mice and in insulin resistant humans [81]. Along the same track, it has been suggested that IGFBP2 plays a role in AD progression [82]. Both of these genes were indeed upregulated in response to Aβ-S8C in our AD neuronal cultures, thus further lending credence to the fact that metabolic dysfunction plays an important role in the pathogenesis of AD. It is noticeable that Aβ-S8C triggers a unique response in AD neuronal cultures, when compared to the control. The creation of a PPI network between the exclusively expressed genes in the AD after Aβ-S8C stimulation revealed *HSPA5* as the core of the Aβ-S8C signature. HSPA5, a chaperone protein that upon accumulation of unfolded proteins controls the activation of the UPR sensors [65], was found down-regulated after Aβ-S8C in our AD TREM2 neuronal cultures. Katayama et al. found that HSPA5 levels are reduced in the brains of AD patients [83]. Although Aβ-S8C stimulation upregulates *ATF3* and *DDIT3* (CHOP) in both CON and AD neuronal cultures, the prominent alteration in the UPR was observed in the AD TREM2 cultures with the upregulation of *XBP1, ATF4, BBC3, HERPUD1* and *CALR*. In support of our data, several studies have shown upregulation of the UPR in brain samples of AD patients [84,85]. According to Han et al. insufficient protein-folding homeostasis by URP increases expression of ATF4 and CHOP and initiates the ER-stress-mediated cell death, activating target genes involved in protein synthesis like aminoacyl-tRNA synthetases and RNA metabolic processes leading to oxidative stress and cell death [86]. Interestingly, biological processes related with increased protein synthesis such as amino acid activation and RNA metabolic process together with the KEGG pathway protein processing in the endoplasmic reticulum were upregulated in AD neuronal cultures. It seems that Aβ-S8C stimulation leads to the activation of the UPR that initially might be protective, however if the balance in proteostasis is not re-established, ER-stress-mediated cell death might mediate neurodegeneration in AD.

4. Materials and Methods

4.1. iPSC Lines

The iPSC lines derived from AD patients as well as control individuals without dementia used in this study have been characterized and published [40–43], as detailed in Table 1. All participants and/or their legal guardian provided written informed consent for participation in the study. Ethical approval was obtained by the Ethics Committee of the University Hospital Antwerp and the University of Antwerp (Approval number 13/15/161 obtained on 22 April 2013). AD patients were ascertained at the memory clinic of the ZNA Middelheim, Antwerpen, Belgium in the frame of a prospective study of neurodegenerative and vascular dementia in Flanders, Belgium. Ethnicity-matched healthy individuals were screened for neurological or psychiatric antecedents, neurological complaints and organic disease involving the central nervous system. Ascertainment and *TREM2* p.R47H genotyping are described in detail in [24]. iPSCs were maintained on Matrigel-coated (Corning, Bedford, MA, USA) plates in StemMACs culture medium (Miltenyi Biotec, Bergisch Gladbach, Germany). The medium was changed every day and the cells were passaged every 5–6 days using PBS without calcium and magnesium (Gibco, Life Technologies, Karlsruhe, Germany).

4.2. Neural Differentiation of the IPSC Lines

For the induction of GABAergic interneurons, iPSCs were differentiated using an embryoid body-based protocol [47] with modifications. On day 1, the iPS cells were harvested and recultivated in suspension in neural induction medium (NIM; DMEM/F-12 (Gibco, Life Technologies, Karlsruhe, Germany), 1% NEAA (Lonza, Basel, Switzerland), 1% N2 supplement (Gibco, Life Technologies, Karlsruhe, Germany), 2 µg/mL of Heparin (Sigma-Aldrich, Steinheim, Germany) and 1% P/S) supplemented with 1 µM purmorphamine (Tocris, Bristol, UK), a SHH agonist. At day 5 the formed aggregates, called embryoid bodies (EBs), were harvested and replated as adherent cells in the same medium and the same concentration of purmorphamine. From day 10 to 18, primitive

neuroepithelia structures were formed and neural rosettes were selected with STEMDiff Neural Rosette Selection reagent (Stem Cell Technologies, Vancouver, Canada) and recultured in suspension in NIM plus a B27 supplement (Gibco, Life Technologies, Karlsruhe, Germany; without retinoic acid) and 20 ng/mL of EGF and FGF2 (both PrepoTech, Hamburg, Germany). After 10 days the cells maintained as aggregates (neurospheres) were dissociated into single cells with accutase (Gibco, Life Technologies, Karlsruhe, Germany) and replated on Matrigel (Corning, Bedford, MA, USA) for the final differentiation in neural differentiating medium (NDM; Neurobasal 1% NEAA, 1% N2 supplement and 1% P/S) supplemented with 1µM of cAMP (Thermo Fisher Scientific, Rockford, IL, USA) and 10 ng/mL of BDNF, GDNF and IGF-1 (all Immuno Tools, Friesoythe, Germany). The iPSC-derived neurons were cultivated for approximately 80 days.

4.3. Cryosection of Neurospheres

Neurospheres were fixed in 4% paraformaldehyde (PFA) for 30 min at room temperature, washed with PBS and cryoprotected in 30% sucrose in PBS overnight at 4 °C. Subsequently, these neurospheres were transferred into embedding medium (Tissue-Tek OCT Compound 4583, Sakura Finetek), snap-frozen on dry ice and stored at −80 °C. Neurospheres were cut into 10 µm thin slides using a Leica CM3050 S cryostat (Leica Biosystems, Wetzlar, Germany).

4.4. Immunofluorescence Stainings

Cells were fixed with 4% paraformaldehyde for 15 min at room temperature (RT). Neurosphere slides were thawed, dried and rehydrated in PBS. Fixed cells and neurosphere slides were permeabilized with 0.2% Triton X-100 for 10 min and blocked with 3% BSA in PBS for 1 h. Samples were then incubated with the following primary antibodies overnight at 4 °C: mouse anti-PAX6 (1:1000, SySy, Goettingen, Germany # 153011), rabbit anti-Nestin (1:400, Sigma Aldrich, Steinheim, Germany #N5413), mouse anti-NKX2.1 (1:1000, Merck Millipore, Burlington, MA, USA #MAB5460), goat anti-SOX1 (1:200, R&D, Bristol, UK # MAB3369), mouse anti-FOXG1 (1:1000, Biozol, Eching, Germany # LS-C197226), mouse anti-βIII-tubulin (1:200, Cell Signaling, Danvers, MA, USA #TU-20), rabbit anti-MAP2 (1:1000, SySy, Goettingen, Germany #188002), guinea pig anti-GFAP (1:500, SySy, Goettingen, Germany #173004), guinea pig anti-Synapsin 1 (1:500, SySy, Goettingen, Germany #106004), mouse anti-SMI-3 (1:2000, Biolegend, San Diego, CA, USA #SMI-312R) and rabbit anti-GABA (1:1000, Sigma Aldrich, Steinheim, Germany #A2052). After washing with PBS, cells were then incubated with the appropriate secondary antibody conjugated with Alexa-488, Alexa-555 or Alexa-647 (1:500, Invitrogen, Thermo Fisher Scientific, Rockford, IL, USA) for 1 h at RT. The nuclear stain Hoechst 33258 (2 ug/mL, Sigma-Aldrich, Steinheim, Germany) was added at the time of the secondary antibody incubation. Slices were mounted in ImmuMount (Thermo Fisher Scientific, Rockford, IL, USA) and fluorescent images were obtained by a LSM 700 microscope (Carl Zeiss AG, Jena, Germany), and analyzed in Adobe Photoshop software CS6 (Adobe, USA).

4.5. Immunoblotting of Lysates from Aß-S8C Dimer Stimulated Cells

iPSC-derived neurons were differentiated for six weeks and then stimulated with 500 nM of oxidized S8C dimers [55] for 72 h. Cells were then washed three times with PBS and then lysed in PBS/1% NP40 + complete protease inhibitor cocktail (Sigma-Aldrich, Steinheim, Germany) and phosphatase inhibitor cocktail 2 (Sigma-Aldrich, Steinheim, Germany). Lysates were cleared by centrifugation at 20.000*g* for 10 min and quantified with the DC Protein assay Kit (Bio-Rad, Hercules, CA, USA). Of the lysates 25 µg were then separated on NuPAGE 4-12% Bis-Tris gels (Invitrogen, Thermo Fisher Scientific, Rockford, IL, USA) and blotted to a 0.2 µm nitrocellulose membrane for 2 h at 400 mA. The blots were blocked in PBS containing 5% skim milk and then probed with the following primary antibodies over night at 4 °C: mouse anti-total TAU (HT7, 1:1000, Thermo Fisher Scientific, Rockford, IL, USA#MN1000), mouse anti-phospho TAU Ser202/Thr205 (AT8, 1:1000, Thermo Fisher Scientific, Rockford, IL, USA#MN1020), rabbit anti-APP (CT15, 1:3500), rabbit anti-βactin (1:5000,

Sigma-Aldrich, Steinheim, Germany #A2066) and mouse anti βIII-tubulin (1:1000, Cell Signaling, Danvers, MA, USA #TU-20). After washing the blots three times with PBS/0.05%Tween20 they were incubated with the appropriate secondary antibody: goat anti-mouse IRDye 680RD and 800CW as well as goat anti-rabbit IRDye 680RD and 800CW (all from LI-COR Biosciences, Lincoln, NE, USA). Following three times washing with PBS/0.05% Tween20 the fluorescent signals were quantified by applying the Odyssey infrared imaging system (LI-COR Biosciences, Lincoln, NE, USA).

4.6. Measurement of Aß1-40 and Aß1-42 by ELISA

Aß1-40 and Aß1-42 concentrations from cleared supernatants of differentiated iPSCs were quantified by using the Amyloid beta 40/42 Human ELISA Kits (#KHB3441 and KHB3481; Thermo Fisher Scientific, Rockford, IL, USA) according to the manufacturer's recommendations. Results were normalized to the protein concentration of the cells. The cells were washed three times with PBS and lysed in PBS/ 1% NP40. The protein content was then measured with the DC Protein assay Kit (Bio-Rad, Hercules, CA, USA).

4.7. RNA Isolation and Quantitative RT-PCR

Total RNA was extracted from cell lysates using Direct-zol RNA Mini Prep kit (Zymo Research, Freiburg, Germany) in combination with peqGOLD TriFast (PeqLab Biotechnologie, Darmstadt, Germany) according to the manufacturer´s protocol. Of purified RNA 0.5 µg was used for first-strand cDNA synthesis using TaqMan reverse transcription reagent (Applied Biosystems, Rockford, IL, USA). cDNA was used for subsequent PCR. Real-time quantification of genes was conducted for three independent cultures from each iPSC-derived interneuron line using the SYBR®Green RT-PCR assay (Applied Biosystems, Waltham, Massachusetts, USA). Primer sequences are provided in Table S6 (Primers were purchased from Eurofins Genomics). Amplification, detection of specific gene products and quantitative analysis were performed using a 'ViiA7' sequence detection system (Applied Biosystem, Waltham, Massachusetts, USA). The expression levels were normalized relative to the expression of the housekeeping gene RPS16 using the comparative Ct-method $2^{-\Delta\Delta Ct}$.

4.8. Generation of Deep Sequencing Data

Deep sequencing data of cDNA from iPSC-derived neuronal cultures were generated at the Neuromics Support Facility at the VIB- University Antwerpen Center for Molecular Neurology. Sequence libraries were constructed using QuantSeq 3′ mRNA-Seq Library Prep Kit (Lexogen, Greenland, NH, USA). Sequencing was performed by Illumina NextSeq sequencing. Reads were single-end with a read length of 151. Samples from two independent experiments (n = 4 cell lines) were multiplexed onto the sequencing flow cell and the measured reads were demultiplexed for follow-up processing. Total RNA from human adult brain, human brain clinically diagnosed with AD and human fetal brain were purchased from BioChain®, Newark, CA USA

4.9. Analysis of Deep Sequencing Data

The demultiplexed fastq files were aligned against the GRCh38 genome with the HISAT2 (version 2.1.0) alignment software [87] using options for clipping the 50 bases at the 3′ end of each read. The exact HISAT2 command, which was mainly derived from the parameter optimizations of Barruzzo et al. [88], was:

hisat2 -p 7 –trim3 50 -N 1 -L 20 -i S,1,0.5 -D 25 -R 5 –mp 1,0 –sp 3,0 -x hisatindex/grch38 -U input.fastq.gz -S output.sam

The resulting BAM files were sorted by coordinates applying SAMtools software [89]. Reads were summarized per gene with the subread (1.6.1) featurecounts software [90] against the gencode.v22.annotation.gtf using parameter –t exon –g gene_id. Summarized reads were normalized in R using the voom normalization [91] algorithm from the limma package [92] filtering genes, which were expressed with CPM (counts per million) > 2 in at least two samples.

4.10. Analysis of Microarray Data

cDNA from iPSC-derived GABAergic interneurons from CON8 and AD-TREM2-4 untreated (CTR) and treated with Aß-S8C dimer was subjected to hybridization in duplicates on the GeneChip PrimeView Human Gene Expression Array (Affymetrix, Thermo Fisher Scientific, Rockford, IL, USA) at the BMFZ (Biomedizinisches Forschungszentrum) core facility of the Heinrich-Heine University, Düsseldorf. Data analysis of the Affymetrix raw data was performed in the R/Bioconductor [93] environment using the package affy [94]. The obtained data were background-corrected and normalized by employing the robust multi-array average (RMA) method from the package affy. Hierarchical clustering dendrograms and heatmaps were generated using the heatmap.2 function from the gplots package with Pearson correlation as similarity measure and color scaling per genes [95]. Expressed genes were compared in Venn diagrams employing package VennDiagram [96]. Gene expression was assessed with a threshold of 0.05 for the detection-*p*-value, which was calculated as described in the supplementary methods in Graffmann et al. [97]. The datasets generated and analyzed during the current study are available in the GEO repository (https://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE143951.

4.11. Protein Interaction Network

A protein interaction network was constructed from the set of 95 genes expressed exclusively in Aβ-S8C stimulated TREM2 neurons in the Venn diagram analysis. Interactions associated with *Homo sapiens* (taxonomy id 9606) were filtered from the Biogrid database version 3.4.161 [98]. From this dataset interactors and additionally interactors of these interactors starting at the proteins coded by the above-mentioned set of 95 genes were extracted. The resulting complex network was reduced by searching the shortest paths between the original set via the method get.shortest.paths () from the R package igraph [99]. The protein network consisting of these shortest paths was plotted employing the R package network [100] marking proteins from the original set in green and inferred proteins in red.

4.12. Gene Ontology and Pathway Analysis

Based on the set of 95 genes expressed exclusively in Aβ-S8C stimulated TREM2 neurons in the Venn diagram analysis over-represented gene ontology terms and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways [101] were determined. The hypergeometric test was used for over-representation analysis—in the version from the GOstats package [102] for GO terms and the version from the R base package for KEGG pathways, which had been downloaded from the KEGG database in March 2018. Dot plots of the most significant GO terms and KEGG pathways were done via the function ggplot() from the R package ggplot2 indicating *p*-values from the hypergeometric test on a red-blue color scale, number of significant genes in the dedicated pathway (G) by the size of the dots and ratios of the number of significant genes in the dedicated pathway/GO to the total number of genes in that pathway/GO on the x-axis.

4.13. Human Cytokine Array

The secretion of cytokines in AD neuronal cultures before and after stimulation with the Aß-S8C dimer was measured employing the Proteome Profiler Human Cytokine Array kit (R&D System, USA). The assay was performed following the manufacturer's instructions. Briefly, AD2-2 and AD2-4 cell culture supernatants from control and 72 h of Aß-S8C dimer stimulation were collected, pooled and mixed with a cocktail of biotinylated detection antibodies for further incubation in a nitrocellulose cytokine array membrane with the immobilized capture antibodies spotted in duplicates. Chemiluminescent detection of the streptavidin-HRP secondary antibody was performed and the average signal (pixel density) was determined for the pair of duplicate spots using Image J (U.S. National Institutes of Health, Bethesda, Maryland, MD, USA). The relative change in cytokine levels

was performed comparing the intensity of the spots in the Aß-S8C dimer stimulated membrane with the control membrane, which was set to 100%.

4.14. Statistical Analysis

Statistical analysis was performed with GraphPad Prism Software version 6.01 (GraphPad software, San Diego, CA, USA). For comparisons of the mean between two groups, one-tail Student´s *t*-test was performed. One-way ANOVA was used for statistical significance analysis for comparisons of the mean among 4 groups, followed by a post hoc test with the use of Tukey´s multiple comparison test. Statistical significance was assumed at $p < 0.05$. All data are expressed as mean ± standard error of the mean (SEM).

5. Conclusions

Our established neuronal cultures using lymphoblast-derived iPSCs from patients harboring the R47H mutation in TREM2 is a relevant model for investigating the effect of this variant in the etiology of LOAD. Comparative global transcriptome analysis identified a distinct gene expression profile in AD neuronal cultures, further suggesting that these lines exhibit alteration in key signaling pathways related to metabolism and immune system in comparison to control, thus implying the partial loss of function of TREM2 due to the R47H substitution. In addition, stimulation with the Aβ-S8C dimer revealed metabolic dysregulation, impaired phagocytosis-related pathway and altered inflammatory responses. Furthermore, our data strongly suggests that the Aβ-S8C dimer signature seems to be centered in the ER-stress response. In conclusion, our AD in vitro model is capable of efficiently responding to signaling cascades associated with the AD pathogenesis and thus is a promising cellular tool for investigating the molecular mechanisms underlying LOAD. Additionally, this cellular model can facilitate the discovery of new AD biomarkers, enable toxicology studies as well as the identification of potential drug targets for future therapy of this devastating disease.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/21/12/4516/s1.

Author Contributions: Conceptualization: S.M. and J.A.; Methodology: S.M., A.M.-S., L.E., M.B.; Formal analysis: S.M., A.M.-S., W.W., J.A.; Data curation: W.W.; Investigation: S.M., A.M.-S., M.B.; Resources: K.S., C.V.B. and C.K.; Writing—original draft preparation: S.M.; Writing—review and editing: J.A., W.W., A.M.-S., K.S., C.V.B., C.K.; Supervision: J.A. All authors have read and agreed to the published version of the manuscript.

Funding: JA acknowledges support from the Medical Faculty, Heinrich-Heine-University, Düsseldorf. JA, KS and CVB are members of the EU project—AgedBrainSYSBIO. The AgedBrainSYSBIO project received funding from the European Union's Seventh Framework Program for research, technological development and demonstration under grant agreement no 305299. Research at the Antwerp site is supported in part by the University of Antwerp Research Fund.

Acknowledgments: We thank Friederike Schröter for helping with the neuronal differentiations. We also thank the Zentrum für Informations- und Medientechnologie (ZIM) at the Heinrich-Heine University for the computational support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Patterson, C. World Alzheimer Report 2018 The state of the art of dementia research: New frontiers. *Alzheimer's Dis. Int.* **2018**. [CrossRef]
2. Prince, M.; Comas-Herrera, A. Dementia Fact Sheet. *WHO*. 2017. Available online: https://www.who.int/news-room/fact-sheets/detail/dementia (accessed on 27 May 2020).
3. Hardy, J.A.; Higgins, G.A. Alzheimer's disease: The amyloid cascade hypothesis. *Science* **1992**, *256*, 184–185. [CrossRef] [PubMed]
4. Polanco, J.C.; Li, C.; Bodea, L.G.; Martinez-Marmol, R.; Meunier, F.A.; Götz, J. Amyloid-β and tau complexity - Towards improved biomarkers and targeted therapies. *Nat. Rev. Neurol.* **2018**, *14*, 22–40. [CrossRef] [PubMed]

5. Hardy, J.; Selkoe, D.J. The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science* **2002**, *297*, 353–356. [CrossRef]
6. Cacace, R.; Sleegers, K.; van Broeckhoven, C. Molecular genetics of early-onset Alzheimer's disease revisited. *Alzheimer's Dement.* **2016**, *12*, 733–748. [CrossRef]
7. Carmona, S.; Hardy, J.; Guerreiro, R. The genetic landscape of Alzheimer disease. *Handb. Clin. Neurol.* **2018**, *148*, 395–408.
8. Bettens, K.; Sleegers, K.; Van Broeckhoven, C. Genetic insights in Alzheimer's disease. *Lancet Neurol.* **2013**, *12*, 92–104. [CrossRef]
9. Farrer, L.A.; Cupples, L.A.; Haines, J.L.; Hyman, B.; Kukull, W.A.; Mayeux, R.; Myers, R.H.; Pericak-Vance, M.A.; Risch, N.; Duijn, C.M. van Effects of Age, Sex, and Ethnicity on the Association Between Apolipoprotein E Genotype and Alzheimer Disease. *JAMA* **1997**, *278*, 1349. [CrossRef]
10. Bickeböller, H.; Campion, D.; Brice, A.; Amouyel, P.; Hannequin, D.; Didierjean, O.; Penet, C.; Martin, C.; Pérez-Tur, J.; Michon, A.; et al. Apolipoprotein E and Alzheimer disease: Genotype-specific risks by age and sex. *Am. J. Hum. Genet.* **1997**, *60*, 439–446.
11. Harold, D.; Abraham, R.; Hollingworth, P.; Sims, R. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.* **2009**, *41*, 1008–1093.
12. Hollingworth, P.; Harold, D.; Sims, R.; Gerrish, A.; Lambert, J.C.; Carrasquillo, M.M.; Abraham, R.; Hamshere, M.L.; Pahwa, J.S.; Moskvina, V.; et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat. Genet.* **2011**, *43*, 429–435. [CrossRef] [PubMed]
13. Lambert, J.C.; Heath, S.; Even, G.; Campion, D.; Sleegers, K.; Hiltunen, M.; Combarros, O.; Zelenika, D.; Bullido, M.J.; Tavernier, B.; et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat. Genet.* **2009**, *41*, 1094–1099. [CrossRef] [PubMed]
14. Naj, A.C.; Jun, G.; Beecham, G.W.; Wamg, L.-S.; Vardarajan, B.N.; Buros, J.; Gallins, P.J.; Buxbaum, J.D.; Jarvik, G.P.; Crane, P.k.; et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* **2011**, *43*, 436–441. [CrossRef]
15. Seshadri, S.; AL, F.; MA, I.; AL, D.; Gudnason, V.; Boada, M.; JC, B.; AV, S.; MM, C.; JC, L.; et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA J. Am. Med. Assoc.* **2010**, *303*, 1832–1840. [CrossRef] [PubMed]
16. Bellenguez, C.; Grenier-Boley, B.; Lambert, J.-C. Genetics of Alzheimer's disease: Where we are, and where we are going. *Curr. Opin. Neurobiol.* **2020**, *61*, 40–48. [CrossRef] [PubMed]
17. Verheijen, J.; Sleegers, K. Understanding Alzheimer Disease at the Interface between Genetics and Transcriptomics. *Trends Genet.* **2018**, *34*, 434–447. [CrossRef]
18. Van Cauwenberghe, C.; van Broeckhoven, C.; Sleegers, K. The genetic landscape of Alzheimer disease: Clinical implications and perspectives. *Genet. Med.* **2016**, *18*, 421–430. [CrossRef]
19. Guerreiro, R.; Wojtas, A.; Bras, J.; Carrasquillo, M.; Rogaeva, E.; Majounie, E.; Cruchaga, C.; Sassi, C.; Kauwe, J.S.K.; Younkin, S.; et al. TREM2 variants in Alzheimer's disease. *N. Engl. J. Med.* **2013**, *368*, 117–127. [CrossRef]
20. Jonsson, T.; Stefansson, H.; Steinberg, S.; Jonsdottir, I.; Jonsson, P.V.; Snaedal, J.; Bjornsson, S.; Huttenlocher, J.; Levey, A.I.; Lah, J.J.; et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* **2013**, *368*, 107–116. [CrossRef]
21. Song, W.; Hooli, B.; Mullin, K.; Jin, S.C.; Cella, M.; Ulland, T.K.; Wang, Y.; Tanzi, R.E.; Colonna, M. Alzheimer's disease-associated TREM2 variants exhibit either decreased or increased ligand-dependent activation. *Alzheimer's Dement.* **2017**, *13*, 381–387. [CrossRef]
22. Sims, R.; van der Lee, S.J.; Naj, A.C.; Bellenguez, C.; Badarinarayan, N.; Jakobsdottir, J.; Kunkle, B.W.; Boland, A.; Raybould, R.; Bis, J.C.; et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat. Genet.* **2017**, *49*, 1373–1384. [CrossRef] [PubMed]
23. Ruiz, A.; Dols-Icardo, O.; Bullido, M.J.; Pastor, P.; Rodríguez-Rodríguez, E.; López de Munain, A.; de Pancorbo, M.M.; Pérez-Tur, J.; Álvarez, V.; Antonell, A.; et al. Assessing the role of the TREM2 p.R47H variant as a risk factor for Alzheimer's disease and frontotemporal dementia. *Neurobiol. Aging* **2014**, *35*, 444.e1–444.e4. [CrossRef] [PubMed]

24. Cuyvers, E.; Bettens, K.; Philtjens, S.; Van Langenhove, T.; Gijselinck, I.; van der Zee, J.; Engelborghs, S.; Vandenbulcke, M.; Van Dongen, J.; Geerts, N.; et al. Investigating the role of rare heterozygous TREM2 variants in Alzheimer's disease and frontotemporal dementia. *Neurobiol. Aging* **2014**, *35*. [CrossRef] [PubMed]
25. Colonna, M.; Wang, Y. TREM2 variants: New keys to decipher Alzheimer disease pathogenesis. *Nat. Rev. Neurosci.* **2016**, *17*, 201–207. [CrossRef] [PubMed]
26. Song, W.M.; Joshita, S.; Zhou, Y.; Ulland, T.K.; Gilfillan, S.; Colonna, M. Humanized TREM2 mice reveal microglia-intrinsic and -extrinsic effects of R47H polymorphism. *J. Exp. Med.* **2018**, *215*, 745–760. [CrossRef]
27. Wang, Y.; Cella, M.; Mallinson, K.; Ulrich, J.D.; Young, K.L.; Robinette, M.L.; Gilfillan, S.; Krishnan, G.M.; Sudhakar, S.; Zinselmeyer, B.H.; et al. TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. *Cell* **2015**, *160*, 1061–1072. [CrossRef]
28. Jay, T.R.; Miller, C.M.; Cheng, P.J.; Graham, L.C.; Bemiller, S.; Broihier, M.L.; Xu, G.; Margevicius, D.; Karlo, J.C.; Sousa, G.L.; et al. TREM2 deficiency eliminates TREM2+ inflammatory macrophages and ameliorates pathology in Alzheimer's disease mouse models. *J. Exp. Med.* **2015**, *212*, 287–295. [CrossRef]
29. Yuan, P.; Condello, C.; Keene, C.D.; Wang, Y.; Bird, T.D.; Paul, S.M.; Luo, W.; Colonna, M.; Baddeley, D.; Grutzendler, J. TREM2 Haplodeficiency in Mice and Humans Impairs the Microglia Barrier Function Leading to Decreased Amyloid Compaction and Severe Axonal Dystrophy. *Neuron* **2016**, *90*, 724–739. [CrossRef]
30. Lee, C.Y.D.; Daggett, A.; Gu, X.; Jiang, L.L.; Langfelder, P.; Li, X.; Wang, N.; Zhao, Y.; Park, C.S.; Cooper, Y.; et al. Elevated TREM2 Gene Dosage Reprograms Microglia Responsivity and Ameliorates Pathological Phenotypes in Alzheimer's Disease Models. *Neuron* **2018**, *97*, 1032–1048.e5. [CrossRef]
31. Zhao, Y.; Wu, X.; Li, X.; Jiang, L.-L.; Gui, X.; Liu, Y.; Sun, Y.; Zhu, B.; Piña-Crespo, J.C.; Zhang, M.; et al. TREM2 Is a Receptor for β-Amyloid that Mediates Microglial Function. *Neuron* **2018**, *97*, 1023–1031.e7. [CrossRef]
32. Muratore, C.R.; Rice, H.C.; Srikanth, P.; Callahan, D.G.; Shin, T.; Benjamin, L.N.P.; Walsh, D.M.; Selkoe, D.J.; Young-Pearse, T.L. The familial Alzheimer's disease APPV717I mutation alters APP processing and Tau expression in iPSC-derived neurons. *Hum. Mol. Genet.* **2014**, *23*, 3523–3536. [CrossRef] [PubMed]
33. Hossini, A.M.; Megges, M.; Prigione, A.; Lichtner, B.; Toliat, M.R.; Wruck, W.; Schröter, F.; Nuernberg, P.; Kroll, H.; Makrantonaki, E.; et al. Induced pluripotent stem cell-derived neuronal cells from a sporadic Alzheimer's disease donor as a model for investigating AD-associated gene regulatory networks. *BMC Genomics* **2015**, *16*, 84. [CrossRef]
34. Yagi, T.; Ito, D.; Okada, Y.; Akamatsu, W.; Nihei, Y.; Yoshizaki, T.; Yamanaka, S.; Okano, H.; Suzuki, N. Modeling familial Alzheimer's disease with induced pluripotent stem cells. *Hum. Mol. Genet.* **2011**, *20*, 4530–4539. [CrossRef] [PubMed]
35. Israel, M.A.; Yuan, S.H.; Bardy, C.; Reyna, S.M.; Mu, Y.; Herrera, C.; Hefferan, M.P.; Van Gorp, S.; Nazor, K.L.; Boscolo, F.S.; et al. Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. *Nature* **2012**, *482*, 216–220. [CrossRef]
36. Duan, L.; Bhattacharyya, B.J.; Belmadani, A.; Pan, L.; Miller, R.J.; Kessler, J.A. Stem cell derived basal forebrain cholinergic neurons from Alzheimer's disease patients are more susceptible to cell death. *Mol. Neurodegener.* **2014**, *9*, 3. [CrossRef]
37. Ochalek, A.; Mihalik, B.; Avci, H.X.; Chandrasekaran, A.; Téglási, A.; Bock, I.; Lo Giudice, M.; Táncos, Z.; Molnár, K.; László, L.; et al. Neurons derived from sporadic Alzheimer's disease iPSCs reveal elevated TAU hyperphosphorylation, increased amyloid levels, and GSK3B activation. *Alzheimers. Res. Ther.* **2017**, *9*, 90. [CrossRef]
38. Flamier, A.; El Hajjar, J.; Adjaye, J.; Fernandes, K.J.; Abdouh, M.; Bernier, G. Modeling Late-Onset Sporadic Alzheimer's Disease through BMI1 Deficiency. *Cell Rep.* **2018**, *23*, 2653–2666. [CrossRef]
39. Zhang, B.; Gaiteri, C.; Bodea, L.-G.; Wang, Z.; McElwee, J.; Podtelezhnikov, A.A.; Zhang, C.; Xie, T.; Tran, L.; Dobrin, R.; et al. Integrated Systems Approach Identifies Genetic Nodes and Networks in Late-Onset Alzheimer's Disease. *Cell* **2013**, *153*, 707–720. [CrossRef]
40. Martins, S.; Bohndorf, M.; Schröter, F.; Assar, F.; Wruck, W.; Sleegers, K.; Van Broeckhoven, C.; Adjaye, J. Lymphoblast-derived integration-free ISRM-CON9 iPS cell line from a 75 year old female. *Stem Cell Res.* **2018**, *26*. [CrossRef]

41. Schröter, F.; Sleegers, K.; Cuyvers, E.; Bohndorf, M.; Wruck, W.; Van Broeckhoven, C.; Adjaye, J. Lymphoblast-derived integration-free iPS cell line from a female 67-year-old Alzheimer's disease patient with TREM2 (R47H) missense mutation. *Stem Cell Res.* **2016**, *17*, 553–555. [CrossRef]
42. Schröter, F.; Sleegers, K.; Cuyvers, E.; Bohndorf, M.; Wruck, W.; Van Broeckhoven, C.; Adjaye, J. Lymphoblast-derived integration-free iPS cell line from a 65-year-old Alzheimer's disease patient expressing the TREM2 p.R47H variant. *Stem Cell Res.* **2016**, *16*, 113–115. [CrossRef] [PubMed]
43. Schröter, F.; Sleegers, K.; Bohndorf, M.; Wruck, W.; Van Broeckhoven, C.; Adjaye, J. Lymphoblast-derived integration-free iPS cell line from a 69-year-old male. *Stem Cell Res.* **2016**, *16*, 29–31. [CrossRef] [PubMed]
44. Levenga, J.; Krishnamurthy, P.; Rajamohamedsait, H.; Wong, H.; Franke, T.F.; Cain, P.; Sigurdsson, E.M.; Hoeffer, C.A. Tau pathology induces loss of GABAergic interneurons leading to altered synaptic plasticity and behavioral impairments. *Acta Neuropathol. Commun.* **2013**, *1*, 34. [CrossRef] [PubMed]
45. Burgos-Ramos, E.; Hervás-Aguilar, A.; Aguado-Llera, D.; Puebla-Jiménez, L.; Hernández-Pinto, A.M.; Barrios, V.; Arilla-Ferreiro, E. Somatostatin and Alzheimer's disease. *Mol. Cell. Endocrinol.* **2008**, *286*, 104–111. [CrossRef] [PubMed]
46. Jo, S.; Yarishkin, O.; Hwang, Y.J.; Chun, Y.E.; Park, M.; Woo, D.H.; Bae, J.Y.; Kim, T.; Lee, J.; Chun, H.; et al. GABA from reactive astrocytes impairs memory in mouse models of Alzheimer's disease. *Nat. Med.* **2014**, *20*, 886–896. [CrossRef] [PubMed]
47. Liu, Y.; Liu, H.; Sauvey, C.; Yao, L.; Zarnowska, E.D.; Zhang, S.-C. Directed differentiation of forebrain GABA interneurons from human pluripotent stem cells. *Nat. Protoc.* **2013**, *8*, 1670–1679. [CrossRef]
48. Li, J.T.; Zhang, Y. TREM2 regulates innate immunity in Alzheimer's disease. *J. Neuroinflammation* **2018**, *15*, 107. [CrossRef]
49. Sengupta, U.; Nilson, A.N.; Kayed, R. The Role of Amyloid-β Oligomers in Toxicity, Propagation, and Immunotherapy. *EBioMedicine* **2016**, *6*, 42. [CrossRef]
50. Viola, K.L.; Klein, W.L. Amyloid β oligomers in Alzheimer's disease pathogenesis, treatment, and diagnosis. *Acta Neuropathol.* **2015**, *129*, 183–206. [CrossRef]
51. Klyubin, I.; Betts, V.; Welzel, A.T.; Blennow, K.; Zetterberg, H.; Wallin, A.; Lemere, C.A.; Cullen, W.K.; Peng, Y.; Wisniewski, T.; et al. Amyloid Protein Dimer-Containing Human CSF Disrupts Synaptic Plasticity: Prevention by Systemic Passive Immunization. *J. Neurosci.* **2008**, *28*, 4231–4237. [CrossRef]
52. Shankar, G.M.; Li, S.; Mehta, T.H.; Garcia-Munoz, A.; Shepardson, N.E.; Smith, I.; Brett, F.M.; Farrell, M.A.; Rowan, M.J.; Lemere, C.A.; et al. Amyloid-β protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nat. Med.* **2008**, *14*, 837–842. [CrossRef] [PubMed]
53. Mc Donald, J.M.; Savva, G.M.; Brayne, C.; Welzel, A.T.; Forster, G.; Shankar, G.M.; Selkoe, D.J.; Ince, P.G.; Walsh, D.M.; Medical Research Council. Cognitive Function and Ageing Study The presence of sodium dodecyl sulphate-stable Aβ dimers is strongly associated with Alzheimer-type dementia. *Brain* **2010**, *133*, 1328–1341. [CrossRef] [PubMed]
54. Brinkmalm, G.; Hong, W.; Wang, Z.; Liu, W.; O'Malley, T.T.; Sun, X.; Frosch, M.P.; Selkoe, D.J.; Portelius, E.; Zetterberg, H.; et al. Identification of neurotoxic cross-linked amyloid-β dimers in the Alzheimer's brain. *Brain* **2019**, *142*, 1441–1457. [CrossRef] [PubMed]
55. Müller-Schiffmann, A.; Herring, A.; Abdel-Hafiz, L.; Chepkova, A.N.; Schäble, S.; Wedel, D.; Horn, A.H.C.; Sticht, H.; de Souza Silva, M.A.; Gottmann, K.; et al. Amyloid-β dimers in the absence of plaque pathology impair learning and synaptic plasticity. *Brain* **2016**, *139*, 509–525. [CrossRef]
56. Abdel-Hafiz, L.; Müller-Schiffmann, A.; Korth, C.; Fazari, B.; Chao, O.Y.; Nikolaus, S.; Schäble, S.; Herring, A.; Keyvani, K.; Lamounier-Zepter, V.; et al. Aβ dimers induce behavioral and neurochemical deficits of relevance to early Alzheimer's disease. *Neurobiol. Aging* **2018**, *69*, 1–9. [CrossRef]
57. Müller-Schiffmann, A.; Andreyeva, A.; Horn, A.H.C.; Gottmann, K.; Korth, C.; Sticht, H. Molecular Engineering of a Secreted, Highly Homogeneous, and Neurotoxic Aβ Dimer. *ACS Chem. Neurosci.* **2011**, *2*, 242–248. [CrossRef]
58. Claes, C.; Van Den Daele, J.; Boon, R.; Schouteden, S.; Colombo, A.; Monasor, L.S.; Fiers, M.; Ordovás, L.; Nami, F.; Bohrmann, B.; et al. Human stem cell–derived monocytes and microglia-like cells reveal impaired amyloid plaque clearance upon heterozygous or homozygous loss of TREM2. *Alzheimer's Dement.* **2018**, *15*, 453–464. [CrossRef]

59. Brownjohn, P.W.; Smith, J.; Solanki, R.; Lohmann, E.; Houlden, H.; Hardy, J.; Dietmann, S.; Livesey, F.J. Functional Studies of Missense TREM2 Mutations in Human Stem Cell-Derived Microglia. *Stem Cell Rep.* **2018**, *10*, 1294–1307. [CrossRef]
60. Hsieh, C.L.; Koike, M.; Spusta, S.C.; Niemi, E.C.; Yenari, M.; Nakamura, M.C.; Seaman, W.E. A role for TREM2 ligands in the phagocytosis of apoptotic neuronal cells by microglia. *J. Neurochem.* **2009**, *109*, 1144–1156. [CrossRef]
61. Kleinberger, G.; Yamanishi, Y.; Suarez-Calvet, M.; Czirr, E.; Lohmann, E.; Cuyvers, E.; Struyfs, H.; Pettkus, N.; Wenninger-Weinzierl, A.; Mazaheri, F.; et al. TREM2 mutations implicated in neurodegeneration impair cell surface transport and phagocytosis. *Sci. Transl. Med.* **2014**, *6*, 243ra86. [CrossRef]
62. Takahashi, K.; Rochford, C.D.P.; Neumann, H. Clearance of apoptotic neurons without inflammation by microglial triggering receptor expressed on myeloid cells-2. *J. Exp. Med.* **2005**, *201*, 647–657. [CrossRef] [PubMed]
63. Garcia-Reitboeck, P.; Phillips, A.; Piers, T.M.; Villegas-Llerena, C.; Butler, M.; Mallach, A.; Rodrigues, C.; Arber, C.E.; Heslegrave, A.; Zetterberg, H.; et al. Human Induced Pluripotent Stem Cell-Derived Microglia-Like Cells Harboring TREM2 Missense Mutations Show Specific Deficits in Phagocytosis. *Cell Rep.* **2018**, *24*, 2300–2311. [CrossRef] [PubMed]
64. Domingues, C.; da Cruz e Silva, O.A.B.; Henriques, A.G. Impact of Cytokines and Chemokines on Alzheimer's Disease Neuropathological Hallmarks. *Curr. Alzheimer Res.* **2017**, *14*, 870–882. [CrossRef] [PubMed]
65. Doyle, K.M.; Kennedy, D.; Gorman, A.M.; Gupta, S.; Healy, S.J.M.; Samali, A. Unfolded proteins and endoplasmic reticulum stress in neurodegenerative disorders. *J. Cell. Mol. Med.* **2011**, *15*, 2025–2039. [CrossRef]
66. Cheng-Hathaway, P.J.; Reed-Geaghan, E.G.; Jay, T.R.; Casali, B.T.; Bemiller, S.M.; Puntambekar, S.S.; von Saucken, V.E.; Williams, R.Y.; Karlo, J.C.; Moutinho, M.; et al. The Trem2 R47H variant confers loss-of-function-like phenotypes in Alzheimer's disease. *Mol. Neurodegener.* **2018**, *13*, 29. [CrossRef]
67. Sudom, A.; Talreja, S.; Danao, J.; Bragg, E.; Kegel, R.; Min, X.; Richardson, J.; Zhang, Z.; Sharkov, N.; Marcora, E.; et al. Molecular basis for the loss-of-function effects of the Alzheimer's disease-associated R47H variant of the immune receptor TREM2. *J. Biol. Chem.* **2018**, *293*, 12634–12646. [CrossRef]
68. Xiang, X.; Piers, T.M.; Wefers, B.; Zhu, K.; Mallach, A.; Brunner, B.; Kleinberger, G.; Song, W.; Colonna, M.; Herms, J.; et al. The Trem2 R47H Alzheimer's risk variant impairs splicing and reduces Trem2 mRNA and protein in mice but not in humans. *Mol. Neurodegener.* **2018**, *13*, 49. [CrossRef]
69. Stomrud, E.; Björkqvist, M.; Janciauskiene, S.; Minthon, L.; Hansson, O. Alterations of matrix metalloproteinases in the healthy elderly with increased risk of prodromal Alzheimer's disease. *Alzheimers. Res. Ther.* **2010**, *2*, 20. [CrossRef]
70. Shah, K.; Desilva, S.; Abbruscato, T. The role of glucose transporters in brain disease: Diabetes and Alzheimer's Disease. *Int. J. Mol. Sci.* **2012**, *13*, 12629–12655. [CrossRef]
71. Katsel, P.; Roussos, P.; Beeri, M.S.; Gama-Sosa, M.A.; Gandy, S.; Khan, S.; Haroutunian, V. Parahippocampal gyrus expression of endothelial and insulin receptor signaling pathway genes is modulated by Alzheimer's disease and normalized by treatment with anti-diabetic agents. *PLoS ONE* **2018**, *13*, e0206547. [CrossRef]
72. Kondo, T.; Asai, M.; Tsukita, K.; Kutoku, Y.; Ohsawa, Y.; Sunada, Y.; Imamura, K.; Egawa, N.; Yahata, N.; Okita, K.; et al. Cell Stem Cell Modeling Alzheimer's Disease with iPSCs Reveals Stress Phenotypes Associated with Intracellular Ab and Differential Drug Responsiveness. *Cell Stem Cell* **2013**, *12*, 487–496. [CrossRef] [PubMed]
73. Wada, Y.; Ishiguro, K.; Itoh, T.J.; Uchida, T.; Hotani, H.; Saito, T.; Kishimoto, T.; Hisanaga, S. Microtubule-stimulated phosphorylation of tau at Ser202 and Thr205 by cdk5 decreases its microtubule nucleation activity. *J. Biochem.* **1998**, *124*, 738–746. [CrossRef] [PubMed]
74. Cruchaga, C.; Kauwe, J.S.K.; Harari, O.; Jin, S.C.; Cai, Y.; Karch, C.M.; Benitez, B.A.; Jeng, A.T.; Skorupa, T.; Carrell, D.; et al. GWAS of Cerebrospinal Fluid Tau Levels Identifies Risk Variants for Alzheimer's Disease. *Neuron* **2013**, *78*, 256–268. [CrossRef] [PubMed]
75. Piers, T.M.; Cosker, K.; Mallach, A.; Johnson, G.T.; Guerreiro, R.; Hardy, J.; Pocock, J.M. A locked immunometabolic switch underlies TREM2 R47H loss of function in human iPSC-derived microglia. *FASEB J.* **2020**, *34*, 2436–2450. [CrossRef] [PubMed]
76. Schetters, S.T.T.; Gomez-Nicola, D.; Garcia-Vallejo, J.J.; Van Kooyk, Y. Neuroinflammation: Microglia and T Cells Get Ready to Tango. *Front. Immunol.* **2017**, *8*, 1905. [CrossRef] [PubMed]

77. Duus, K.; Hansen, P.R.; Houen, G. Interaction of calreticulin with amyloid beta peptide 1-42. *Protein Pept. Lett.* **2008**, *15*, 103–107.
78. Lin, Q.; Cao, Y.; Gao, J. Serum Calreticulin Is a Negative Biomarker in Patients with Alzheimer's Disease. *Int. J. Mol. Sci.* **2014**, *15*, 21740–21753. [CrossRef]
79. Hondius, D.C.; van Nierop, P.; Li, K.W.; Hoozemans, J.J.M.; van der Schors, R.C.; van Haastert, E.S.; van der Vies, S.M.; Rozemuller, A.J.M.; Smit, A.B. Profiling the human hippocampal proteome at all pathologic stages of Alzheimer's disease. *Alzheimer's Dement.* **2016**, *12*, 654–668. [CrossRef]
80. Armstrong, A.; Mattsson, N.; Appelqvist, H.; Janefjord, C.; Sandin, L.; Agholme, L.; Olsson, B.; Svensson, S.; Blennow, K.; Zetterberg, H.; et al. Lysosomal network proteins as potential novel CSF biomarkers for Alzheimer's disease. *Neuromolecular Med.* **2014**, *16*, 150–160. [CrossRef]
81. Mody, N.; Agouni, A.; Mcilroy, G.D.; Platt, B.; Delibegovic, M. Susceptibility to diet-induced obesity and glucose intolerance in the APP SWE/PSEN1 A246E mouse model of Alzheimer's disease is associated with increased brain levels of protein tyrosine phosphatase 1B (PTP1B) and retinol-binding protein 4 (RBP4), and basal phosphorylation of S6 ribosomal protein. *Diabetologia* **2011**, *54*, 2143–2151. [CrossRef]
82. Bonham, L.W.; Geier, E.G.; Steele, N.Z.R.; Holland, D.; Miller, B.L.; Dale, A.M.; Desikan, R.S.; Yokoyama, J.S.; Initiative, A.D.N. Insulin-Like Growth Factor Binding Protein 2 Is Associated With Biomarkers of Alzheimer's Disease Pathology and Shows Differential Expression in Transgenic Mice. *Front. Neurosci.* **2018**, *12*, 476. [CrossRef] [PubMed]
83. Katayama, T.; Imaizumi, K.; Sato, N.; Miyoshi, K.; Kudo, T.; Hitomi, J.; Morihara, T.; Yoneda, T.; Gomi, F.; Mori, Y.; et al. Presenilin-1 mutations downregulate the signalling pathway of the unfolded-protein response. *Nat. Cell Biol.* **1999**, *1*, 479–485. [CrossRef] [PubMed]
84. Hamos, J.E.; Oblas, B.; Pulaski-Salo, D.; Welch, W.J.; Bole, D.G.; Drachman, D.A. Expression of heat shock proteins in Alzheimer's disease. *Neurology* **1991**, *41*, 345–350. [CrossRef] [PubMed]
85. Hoozemans, J.J.M.; Veerhuis, R.; Van Haastert, E.S.; Rozemuller, J.M.; Baas, F.; Eikelenboom, P.; Scheper, W. The unfolded protein response is activated in Alzheimer's disease. *Acta Neuropathol.* **2005**, *110*, 165–172. [CrossRef]
86. Han, J.; Back, S.H.; Hur, J.; Lin, Y.-H.; Gildersleeve, R.; Shan, J.; Yuan, C.L.; Krokowski, D.; Wang, S.; Hatzoglou, M.; et al. ER-stress-induced transcriptional regulation increases protein synthesis leading to cell death. *Nat. Cell Biol.* **2013**, *15*, 481–490. [CrossRef]
87. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [CrossRef]
88. Baruzzo, G.; Hayer, K.E.; Kim, E.J.; Di Camillo, B.; FitzGerald, G.A.; Grant, G.R. Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nat. Methods* **2017**, *14*, 135–139. [CrossRef]
89. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
90. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef]
91. Law, C.W.; Chen, Y.; Shi, W.; Smyth, G.K. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **2014**, *15*, R29. [CrossRef]
92. Smyth, G.K. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat. Appl. Genet. Mol. Biol.* **2004**, *3*. [CrossRef] [PubMed]
93. Gentleman, R.C.; Carey, V.J.; Bates, D.M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; et al. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **2004**, *5*, R80. [CrossRef] [PubMed]
94. Gautier, L.; Cope, L.; Bolstad, B.M.; Irizarry, R.A. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **2004**, *20*, 307–315. [CrossRef] [PubMed]
95. Warnes, G.R.; Bolker, B.; Bonebakker, L.; Gentleman, R.; Liaw, W.H.A.; Lumley, T.; Maechler, M.; Magnusson, A.; Moeller, S.; Schwartz, M.; et al. gplots: Various R Programming Tools for Plotting Data. *R Package Version* **2015**.
96. Chen, H.; Boutros, P.C. VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinf.* **2011**, *12*, 35. [CrossRef] [PubMed]

97. Graffmann, N.; Ring, S.; Kawala, M.-A.; Wruck, W.; Ncube, A.; Trompeter, H.-I.; Adjaye, J. Modeling Nonalcoholic Fatty Liver Disease with Human Pluripotent Stem Cell-Derived Immature Hepatocyte-Like Cells Reveals Activation of PLIN2 and Confirms Regulatory Functions of Peroxisome Proliferator-Activated Receptor Alpha. *Stem Cells Dev.* **2016**, *25*, 1119–1133. [CrossRef]
98. Chatr-aryamontri, A.; Oughtred, R.; Boucher, L.; Rust, J.; Chang, C.; Kolas, N.K.; O'Donnell, L.; Oster, S.; Theesfeld, C.; Sellam, A.; et al. The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* **2017**, *45*, D369–D379. [CrossRef]
99. Csardi, G.; Nepusz, T. The igraph software package for complex network research. *InterJournal* **2006**, *1695*, 1–9.
100. Butts, C. network: A Package for Managing Relational Data in R. *J. Stat. Softw.* **2008**, *24*, 2. [CrossRef]
101. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [CrossRef]
102. Falcon, S.; Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **2007**, *23*, 257–258. [CrossRef] [PubMed]

Chapter 3

Discussion

3.1 Analysis of stem cell based disease models in liver

Reprogramming adult cells into the pluripotent state has become a routine procedure and the same holds for the derivation of HLCs. As the name says these cells are only "hepatocyte-like" but not fully mature hepatocytes. Their phenotype resembles more the fetal hepatocyte phenotype and we have shown that the gene expression patterns of HLCs have more overlap with fetal liver (Matz et al. 2017). This behavior is in line with the expectations considering the differentiation of iPSCs into HLCs as a process recapitulating liver development in the fetus. Nevertheless, we could show that most of the functionality which can be measured in functional assays is already existent such as glycogen storage, urea and bile acid production, indocyanine green (ICG) uptake and release, and cytochrome activity (Matz et al. 2017). However, cytochromes only in parts resemble the mature phenotyp in a way that while already some, e.g. CYP3A4, are expressed, many other cytochromes have lower expression than in the mature cells. In addition, Albumin (*ALB*) - the most important gene expression marker in liver - in HLCs does not reach the level of mature hepatocytes. Gerets *et al.* have shown that gene expression and cytochrome activity in the usually employed hepatocarcinoma-derived or transformed hepatocyte cell lines differ from primary human hepatocytes (Gerets et al. 2012). Thus, HLCs are a promising alternative that however needs further improvements of cytochrome and ALB expression. Gieseck et al. proposed 3D-culturing of hepatocytes, Berger et al. a micropatterned co-culture platform with murine embryonic fibroblasts to improve maturity (Gieseck III et al. 2014), (Berger et al. 2015). Other approaches tackle maturity and the expression of ALB via transcription factors. Our transcription factor analysis revealed several transcription factors including HNF4A, HNF1A, FOXA2, FOXA1 and CEBPA which are regulating ALB and hence are potential targets for inducing maturity via a higher expression level of ALB (Matz et al. 2017) thus expanding and partly confirming earlier results (J. Li, Ning, and Duncan 2000).

3.2 Analysis of stem cell based disease models in neurons

Differentiation of pluripotent stem cells into neurons has become a routine procedure. Challenges exist in reaching an aged phenotype that is needed for modeling of neurodegenerative diseases such as Alzheimer's and Parkinson's disease and in the differentiation into specific neuron types. However, several iPSC-based neuron models exist already for neurons specific for dedicated neurotransmitters. During differentiation of pluripotent stem cells into neurons multiple neuron types can arise as e.g. Brennand et al. report a mixture of glutamatergic, GABAergic and dopaminergic in their differentiation approach for modeling Schizophrenia (Brennand et al. 2011). Cao et al. generated serotonergic neurons from iPSCs (Cao et al. 2017). Moreno et al. differentiated iPSCs into cholinergic neurons in an Alzheimer's disease study (Moreno et al. 2018).

For neurodegenerative diseases several iPSC-based disease models exist including the Alzheimer's disease models from our group (Hossini et al. 2015) and from Israel et al. (Israel et al. 2012), the Parkinson's disease model from Cooper et al. (Cooper et al. 2012) and the Huntinton's disease models by Mehta et al. (Mehta et al. 2018) and An et al. including genetic correction (An et al. 2012). Wu et al. provide an overview of existing iPSC-based models of neurodegenerative disease (Wu et al. 2019).

The disease models can be distinguished into models for sporadic disease without known mutation, such as our publication about an iPSC-model of sporadic AD (Hossini et al. 2015) and models for familial disease such as the iPSC- AD-model by Kondo et al. (Kondo et al. 2013) with a APP mutation and the iPSC-AD-model by Yagi et al. with a PSEN2 mutation (Yagi et al. 2011) . In the familial disease iPSC-models the mutation usually persists after reprogramming but has to be verified - as we showed for instance in our publication about an iPSC-model of the Nijmegen Breakage Syndrome (Mlody et al. 2017).

The aim to detect disease-driving mechanisms can be pursued by comparing iPSC-derived disease neurons to iPSC-derived healthy neurons as control. Applying CRISPR-Cas9 techniques (Doudna and Charpentier 2014) also an isogenic control can be generated by editing the mutated gene (Park et al. 2015). Differential gene expression can then be assessed between the disease and the control model to identify involved genes and the resulting gene sets can be further analysed for functional annotations such as pathways, transcription factors, GOs, etc.. However, as many disease arise from a complex interplay of genetic and other external factors, e.g. aging as the most important risk factor for AD, it has to be taken into account that only the genetic component of a more complex system is observed. Nonetheless, in our iPSC-model of sporadic AD we could recapitulate essential disease mechanisms such as phosphorylated tau and down-regulation of AD-related genes, in particular genes from the proteasome or genes reacting on oxidative stress (Hossini et al. 2015). The down-regulation of proteasome-related genes led us to

the hypothesis that early occurring impairment of the proteasome could result in later accumulation of amyloid plaques.

3.3 Meta-analysis of biopsy-based data in liver and comparison to stem cell based disease models

The focus of the biopsy-based data analysis was the study of NAFLD and also the comparison to ALD. In a first multi-omics analysis we could compare the early steatosis stages of NAFLD which we stratified into a simplified scheme combining stages 2 and 3 (> 33% steatosis area) from the scheme by Kleiner et al. as high-grade and stage 1 (5% - 33% steatosis area) as low-grade steatosis (Kleiner et al. 2005). However, even in this simplified classification we observed a high heterogeneity in patients manifesting in the transcriptome, metabolome and proteome. We considered this heterogeneity as a feature of the disease requiring personalized approaches for a full elucidation of the causes. We could nevertheless find some marker genes such as PLIN2 in the transcriptome and proteins such as Adiponectin in ELISA assays from the patients' serum.

In a meta-analysis along with a review we investigated the progression from simple steatosis to NASH for which non-invasive markers are needed (Wruck et al. 2017). Existing non-invasive markers for NASH such as keratin 18 (KRT18) (Feldstein et al. 2009) or test panels such as FibroTest (Ratziu et al. 2006) are still missing sensitivity or specifity. We could condense the results of our meta-analysis to signature of 22 genes most significantly correlating with the progression of NAFLD. This signature could be associated with cholesterol and lipid metabolic pathways thus confirming a body of literature (Min et al. 2012), (Musso, Gambino, and Cassader 2013) indicating cholesterol as important disease accelerating factor and cholesterol clearance as disease alleviation (Neuschwander-Tetri et al. 2015).

In a follow-up meta-analysis we compared NAFLD and alcoholic liver disease (ALD) and found predominating commonalities at the transcriptome level (Wruck and Adjaye 2017). Numerous down-regulated metabolic pathways and cytochrome-related pathways were down-regulated and a few pathways including ECM-receptor interaction, phagosome and lysosome were up-regulated in both dieases. Among the pathways going into opposite directions was glycolysis which was down-regulated in ALD and up-regulated in NAFLD. Also rate-limiting genes of cholesterol pathways such as HMGCR , SQLE and CYP7A1 were down-regulated in ALD and up-regulated in NAFLD. We hypothesize that the nevertheless similar phenotypes in both diseases may be due to a lower level of CYP7A1 in comparison to cholesterol synthesizing genes HMGCR and SQLE. CYP7A1 is responsible for cholesterol secretion into bile (Ikonen 2008).

Comparing the results of biopsy-based analysis to stem cell based NAFLD models shows that the latter are a promising tool for the exploration of NAFLD. While they are just in the beginning of their development they could already recapitulate the mostimportant marker for lipid droplets PLIN2 which we found in both approaches (Graffmann et al. 2016). The models will have to be further elaborated in order to capture the progression of the disease to NASH, cirrhosis and hepatocellular carcinoma. Methods pursued in NAFLD animal models such as methionine-choline deficient diet (Rinella et al. 2008) or carbon tetrachloride (CCl_4) and high-fat, high-fructose, high-cholesterol western diet (Tsuchida et al. 2018) may be exploited to achieve stem cell based models of later disease stages. With that in hand they will be an invaluable tool to explore NAFLD in a personalized medicine approach which is indispensable to get insights into the individually varying metabolic mechanisms.

3.4 Meta-analysis of biopsy-based data in neurons and comparison to stem cell based disease models

Central topics of the meta-analyses of transcriptome data from human post-mortem brain biopsies described here were the study of Alzheimer's disease (AD) and brain aging. The AD study was focused on hippocampus biopsies (Wruck, Schröter, and Adjaye 2016) while the brain aging study (Wruck and Adjaye 2020) was focused on pre-frontal cortex biopsies.

The AD meta-analysis compared hippocampus biopsy studies among each other and additionally to stem cell based AD models (Wruck, Schröter, and Adjaye 2016) . Hippocampus is the brain region associated with learning and memory and therefore plays an important role in AD where decreased short-term memory is one of the earlier symptoms (M. P. Laakso et al. 1995), (Mikko P Laakso et al. 2000), (Mu and Gage 2011). However, AD is not restricted to hippocampus and during the course of the disease affects the whole brain. Other brain regions were excluded from this meta-analysis because the high variability of gene expression between different brain regions would have confounded the detection of genes relevant for the etiology of AD. The meta-analysis resulted in a gene signature which could distinguish AD and healthy control biopsies into two clusters. The prominent AD-related genes APP and APOE only had low correlation with the AD phenotype in this analysis and therefore were not in the signature. This is likely to be originated in the data basis containing not only familial early-onset cases but many sporadic late-onset cases without genetic predisposition. Gene ontology and pathway analysis of the signature indicated that reactive oxygen species, response to stress and metabolic processes play a major role in AD. Transcription factor analysis additionally suggested that FOXA1 and FOXA2 gene regulatory networks are involved in the etiology of the disease.

The aging brain meta-analysis characterized 591 transcriptomes of prefrontal cortex biopsies for the full range of ages in males and females (Wruck and Adjaye 2020). In this study we identified a decline of synaptic transmission and up-regulation of glial fibrillary acidic protein (GFAP) along with progressing age. We found an antagonistic connection between the down-regulated CAMK4 (calcium/calmodulin dependent protein kinase IV) and the up-regulated GFAP. Decreased CAMK4 may be involved in up-regulation of GFAP via a mechanism involving cAMP responsive element binding protein (CREB) and mitogen-activated protein kinase (MAPK, alias ERK). Besides the decline in synaptic transmission also decreased neurogenesis and an increased base-level of inflammatory and immune-related processes could be detected in both sexes. Differences between both sexes were found in dendritic spine morphogenesis, catecholamine signaling and cellular responses to external stimuli. Also on the global transcriptome level principal component analysis (PCA) and hierarchical clustering dendrograms showed a clustering of samples by sex which was much more apparent than clustering by age. Detailed analysis showed that this can be attributed in large parts to genes on the sex chromosomes.

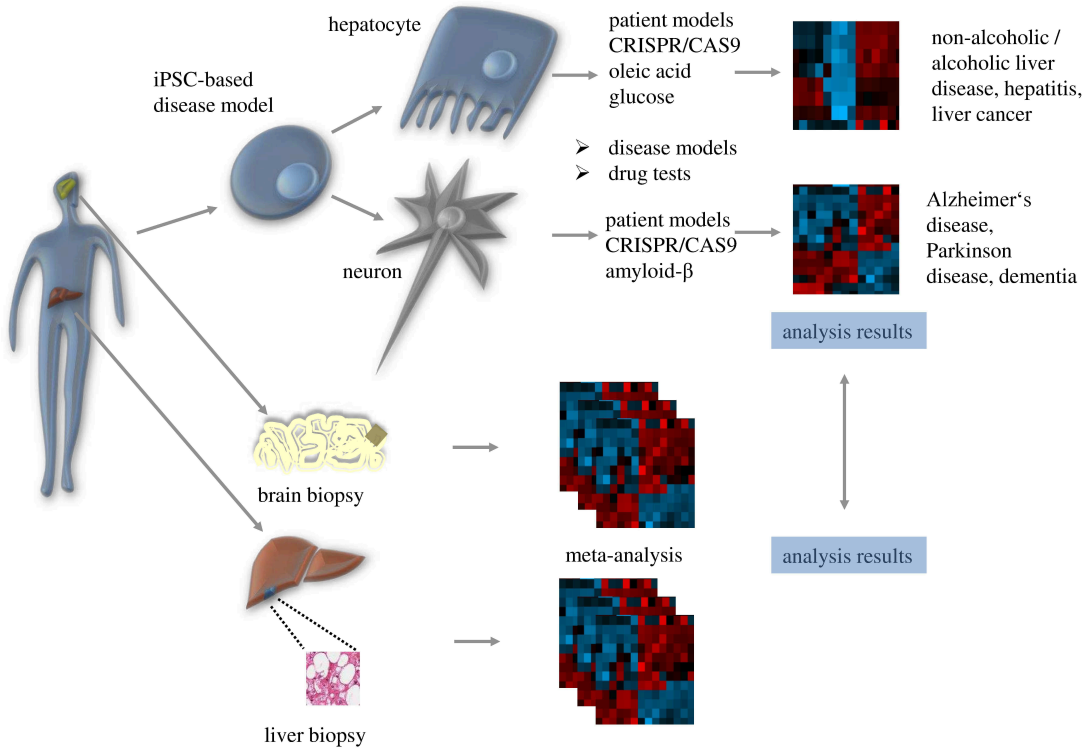


Figure 3: Comparison of meta-analyses and iPSC-based disease models.

In Figure 3 a scheme is laid out about how the meta-analysis approach and the iPSC-disease model approach are compared in this thesis. Meta-analyses are based on microarray and NGS transcriptome data retrieved from public repositories. Here, the focus was on liver and brain for which datasets reflecting the investigated diseases were accessed accounting for comparability and reliability of the datasets, e.g. comparing only specific regions of the brain such as hippocampus or prefrontal cortex. For the iPSC-based approach iPSCs were reprogrammed based on fibroblasts or urine-derived cells from patients and healthy controls. These iPSCs in turn were differentiated into hepatocyte or neuronal lineage and subjected to further challenges, e.g. oleic acid for induction of steatosis, if needed. Gene expression data of the iPSC-disease models was then evaluated and compared to the results of the meta-analyses.

3.5 Time series analysis of gene expression data

In the above mentioned meta-analysis of 591 prefrontal cortex transcriptomes I developed methods for the analysis of time series of gene expression data (Wruck and Adjaye 2020). Gene expression data of the post-mortem prefrontal cortex biopsies covered a range of ages from 16 to 106 years for both sexes. I reconstructed time series from these gene expression data and employed the Granger test - which was originally designed for econometric models (Granger 1969) - to test causality between them. This leads to a null hypothesis stating that the time series g of one gene x does not cause the time series h of gene y.

The test is realized by an auto-regression model of h . If adding of lagged values of g would improve the model this would mean that h would depend on previous values of g. In that case at least one coefficient b_i has to be unequal to zero. The test can then be realized by checking the coefficients b_i for equality to zero:

$$h_t = \sum_{i=1}^{L} a_i h_{t-i} + \sum_{i=1}^{L} b_i g_{t-i} + \epsilon_t \tag{3.1}$$

$$H_0 : b_1 = ... = b_L = 0 \, (gene\, h\, does\, not\, Granger\, cause\, gene\, g) \tag{3.2}$$

Here, h_t and g_t are the time series and a_i are coefficients of the auto-regression model of h and b_i coefficients for the added lagged values of g, $å_t$ is the error.

As we were also interested in relationships between GOs and between genes and GOs we developed the above described time series analysis further. For each GO we condensed the means of the expression values of genes significantly correlated or anti-correlated with age to a consensus time series.

Let G_{gu} and G_{gd} be the sets of genes significantly correlated and anti-correlated with age from the GO g and X_{aui} and X_{adi} be the corresponding gene expression values over all existing ages. Then the consensus time series $\overline{X}_{gu}$ and $\overline{X}_{gd}$ for GO g are:

$$\overline{X}_{gu} = \frac{1}{|G_{gu}|} \sum_{i \in G_{gu}} X_{aui} \tag{3.3}$$

$$\overline{X}_{gd} = \frac{1}{|G_{gd}|} \sum_{i \in G_{gd}} X_{adi} \tag{3.4}$$

Figure 4 shows an example of these GO consensus time series for the GO *Weibel-Palade body* and the gene *GFAP* which was found to "Granger-cause" the GO *Weibel-Palade body* with p=0.04.

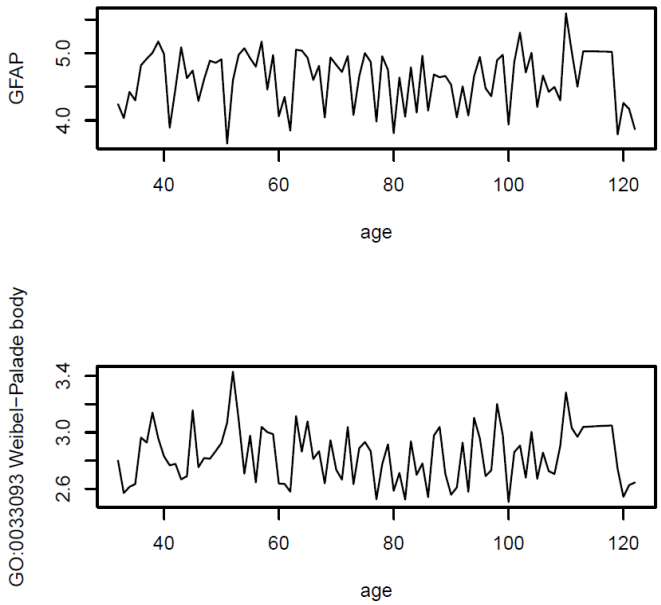


Figure 4: GO consensus time series for the GO *Weibel-Palade body* and the gene *GFAP*.

3.6 Meta-analysis of epithelial lung cells infected with the coronavirus SARS-CoV-2

The main focus of this meta-analyses of transcriptome data from diverse human epithelial lung cells was to find genes connected with the already established SARS-CoV-2 receptor gene *ACE2*. I retrieved several transcriptome datasets from public repositories some of which had been infected with SARS-CoV-2 in cell culture. In this particular case the interest was not to find gene expression clusters or networks grouping genes with similar expression unsupervisedly as would have been done by approaches such as the ones by Langfelder et al. (Langfelder and Horvath 2008), Eisen et al. (Eisen et al. 1998) or Herwig et al. (Herwig 1999) but instead to find exactly the genes with the highest similarity with *ACE2*. Therefore, the Pearson correlation coefficient of the expression of each gene to the expression of *ACE2* was calculated across all experiments.

$$r_i = cor(e(i), e(ACE2)) \tag{3.5}$$

Here, r_i is the Pearson correlation coefficient of the expression of gene i (*e(i)*) with the expression of *ACE2* (e(*ACE2*)), *cor* is the function to calculate the Pearson correlation. For the Pearson correlation the corresponding test for association between paired samples using a correlation of zero as null hypothesis is calculated and all genes are ranked by the resulting p-value. This way the genes with highest positive and lowest negative correlation with *ACE2* are determined which are suggested to interact with *ACE2* in the processing of the virus. The most prominent gene that was found in this study was the gene *TMPRSS4* among others from the transmembrane serine protease TMPRSS family. These would raise the question if besides *TMPRSS2* as already known interactor with *ACE2* also other members of the TMPRSS family play important roles in the entry of SARS-CoV-2.

Figure 5 illustrates how the results from the SARS-CoV-2 meta-analysis can be further investigated in 2D- and 3D stem-cell based cellular models. The meta-analysis was focussed on transcriptome data from SARS-CoV-2 infected lung epithelial cells but to date the knowledge is already settled that SARS-CoV-2 can attack multiple if not all organs of the human body. One major finding of the meta-analysis was that several members of the Transmembrane serine proteases family may be involved in SARS-CoV-2 endocytosis for instance was already reported in similar way for a distinct organ type: TMPRSS4 was identified as major actor in SARS-CoV-2 infection of intestine by Zang et al. (Zang

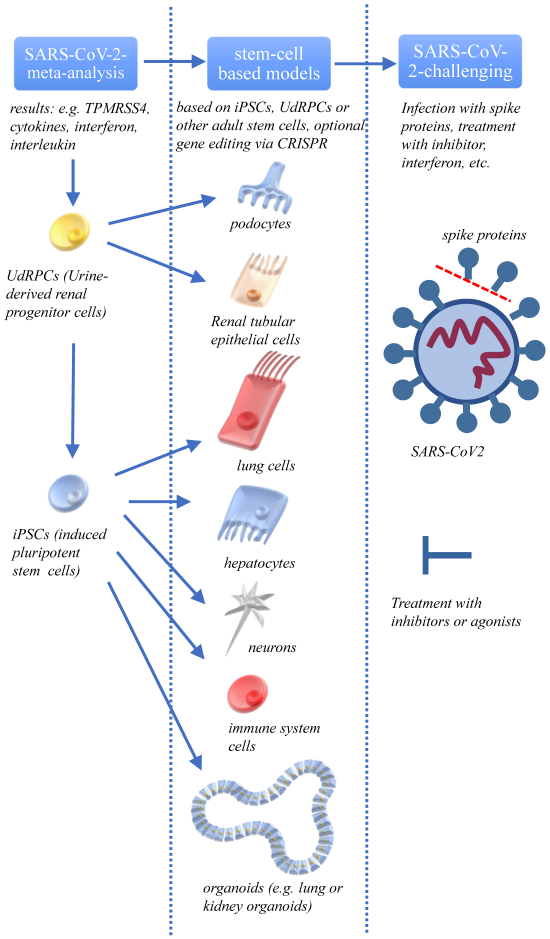


Figure 5: Stem-cell-based 2D- and 3D-models (organoids) can be employed to further explore findings from the meta-analysis and possible CRISPR or drug interventions.

et al. 2020). Thus, it can be expected that there are common ways of infection of distinct organs while on the other hand individual differences in infected organs show that there are also organ-specific mechanisms. Therefore, it is of great importance to study the infection in multiple organs as illustrated in figure 5. Besides the preference for the respiratory tract renal tropism has been reported for SARS-CoV-2 (Puelles et al. 2020) what can be seen in the context of the major role of the SARS-CoV-2 receptor ACE2 in kidneys participating in the regulation of blood pressure via the renin-angiotensin system . Figure 5 shows that UdRPCs (Urine-derived renal progenitor cells) can be derived from urine and in turn be differentiated into podocytes and renal tubular epithelial cells (Rahman et al. 2020). This represents a personalized approach as this way from each individual cells can be harvested non-invasively and rapidly. UdRPCs as well as podocytes and renal tubular epithelial cells can then directly be used for research on

SARS-CoV-2: target genes can be edited via CRISPR or can be inhibited or stimulated by drug candidates. Possible locations for pharmaceutical intervention have been listed in the SARS-CoV-2 publication including the endocytosis via TMPRSS family members, virus replication and packaging or interferons and cytokines (Wruck and Adjaye 2020b, 2).

However, reprogramming of the UdRPCs into iPSCs is the route leading to a full spectrum of organs. iPSCs can then be differentiated into cell types such as lung cells, hepatocytes, neurons and cells of the immune system. In more sophisticated 3D-approaches organoids could be generated which represent small pieces of an organ. For exploration of SARS-CoV-2 lung and kidney organoids will provide in vitro models capturing multiple organ-specific cell types and thus improve the realism of the models. In all these approaches target genes found as results of the meta-analysis can be edited via CRISPR or perturbated via other methods such as silencing RNA.

SARS-CoV-2 infection can be simulated by infection with the spike protein only what would remove the infectivity from the virus and make it hazard-free. This method could also account for mutations inside the spike protein which in future may pose severe challenges to the worlwide efforts to tackle the virus. Phylogenetic analysis as described in the following chapter is another essential pillar in tracking the virus mutations.

3.7 Phylogenetic analysis of transmission routes of the coronavirus SARS-CoV-2 to West-Africa

In this phylogenetic analysis a profound dataset of sequences of the coronavirus SARS-CoV-2 from the West-African countries Ghana, Nigeria, Senegal and Gambia were investigated together with a representative set of reference sequences from Asia, Europe and America. The analysis employed the pipeline from the NextStrain project (Hadfield et al. 2018) and visualized geographic transmission routes within this framework. With the development of the novel powerful sequencing methods described above in the chapter “Next-generation sequencing (NGS)” phylogenetic analysis has been advanced on the one hand by the availability of a much larger base of sequences and on the other hand by new methods enabling efficient analysis of them. Alignment algorithms have been improved to tackle the needs in terms of performance, storage and accuracy for NGS (Dobin et al. 2013), (Langmead et al. 2009) and for phylogenetic analysis (Katoh et al. 2002). Phylogenetic analysis aims at finding developmental relationships between genetic sequences such as DNA or proteins. The analysis results open up a multitude of applications such as to find out how a gene evolved over various species, e.g. the pluripotency gene POU5F1 which is related to the gene POU2 in earlier evolved species

(Frankenberg and Renfree 2013). With respect to viruses, phylogenetic analysis enables to track transmission routes focussing on single mutations or to identify new variants with characteristic patterns of mutations which when changing proteins may also have impact on the severity and infectivity. For the research of the evolution of influenza viruses the GISAID database has been set up harboring a plethora of virus sequences. On the outbreak of the COVID-19 pandemic this database has been instrumental to store a multitude of SARS-CoV-2 sequences the COVID-19 causing virus. Novel phylogenetic algorithms use maximum-likelihood approaches to estimate dates of relevant events such as the date of zoonosis or the date when a specific virus variant emerged (Sagulenko, Puller, and Neher 2018). For influenza viruses and SARS-CoV-2 these techniques have been employed to provide the NextStrain framework to epidemiologists visualizing the geographic evolution of a virus (Hadfield et al. 2018).

Here, in the SARS-CoV-2 sequences retrieved from the GISAID database and analyzed via the NextStrain pipeline country-specific patterns of viral clades were found. Clades are lineages of a virus related to major clusters in the phylogenetic tree and characterized by specific mutations such as the D614G mutation in the G-clades (Korber et al. 2020). Employing the GISAID clade nomenclature the study revealed a higher abundance of the Europe-associated G-clades in Gambia and Senegal, and a mixture of the G-clades and Asian-associated clades (clades L, S, V) in Nigeria and Ghana. In the analyzed West-African sequences the Asian-associated clades which globally appeared earlier emerged after the Europe-associated later G-clades. This unexpected finding can be caused by founder effects meaning that just by accident G-clade viruses were introduced first. Another explanation suggesting latent circulation of the Asian-associated clades appeared more likely as it was confirmed by detailed analysis of distinct samples which additionally pointed at migration routes via Mali and Tunisia.

3.8 Future bioinformatics challenges in stem cell-based disease modeling and regenerative medicine

One main advantage of stem cell based disease models is to provide access to otherwise hard or even impossible accessible tissues such as brain, liver or kidney. Another advantage is the capability to pursue a personalized approach where a model for each patient or individual can be established. On the other hand, these models have problems in capturing the maturity or even aged state of cells which play a crucial role in many diseases and particularly in neurodegenerative diseases. To that end bioinformatics tools may support these models by providing information about the quality of the differentiations derived

from the stem cells. They can assess if the target tissue type was reached by determining gene expression of crucial marker genes, comparisons to huge databases providing associations between gene expression and tissues such as GTEX (GTEx Consortium 2013) or in silico classification approaches (Roost et al. 2015). Furthermore, they can check the pluripotent state of cells with algorithms such as the Pluritest (Müller et al. 2011) or simply by comparison to other pluripotent stem cells (Schröter et al. 2016). Future challenges here may be located in the detection of hitherto unknown or only poorly specified cell types which may be assessed by novel single-cell sequencing methods such as for instance Lindström et al. characterize in the developing kidney (Lindström et al. 2018).

Personalized approaches will be challenging as they introduce additional complexity into the already complex regulatory networks orchestrating many diseases such as AD, NAFLD, diabetes, cancer, etc.. However - as we already showed the strong heterogeneity in NAFLD (Wruck et al. 2015) - they will be indispensable to understand many diseases and to find appropriate personalized therapies. An feasible avenue to cope with the additional complexity due to the individual differences will be systems biology modeling (Jozefczuk et al. 2012) providing the means to use one model for several individuals by populating it with the personalized parameters.

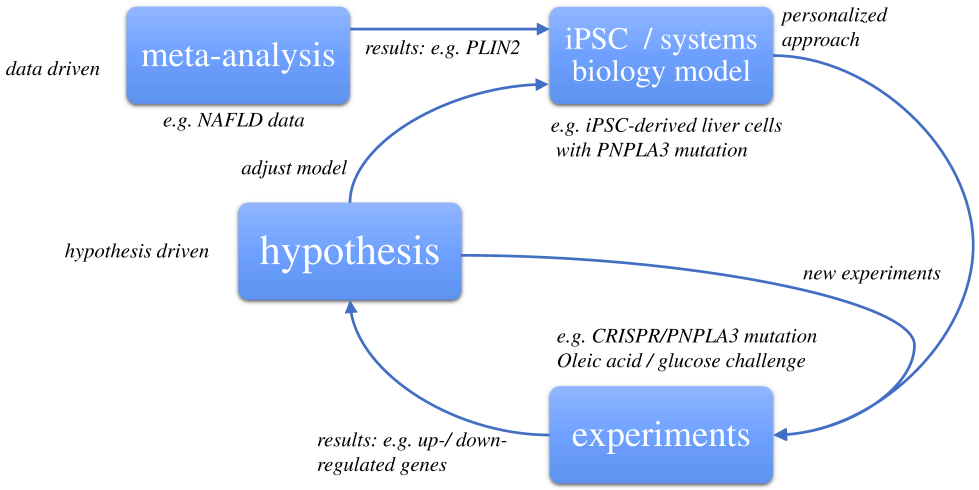


Figure 6: Meta-analysis and modeling cycle.

I propose the scheme in Figure 6 for the combination of meta-analysis and iPSC- or also Systems Biology-based models. The meta-analysis of transcriptome data has the advan-

tage to integrate huge numbers of datasets which directly reflect disease characteristics. On the other hand, because of the complex production process which also may introduce additional variation iPSC-based models are usually only generated in small sample size. Although mutations are kept alongside reprogramming and differentiation, as we had seen e.g. in the NBS study, other factors such as age or epigenetics are also influenced by the reprogramming. Thus, the iPSC-models cannot compete with meta-analyses of disease tissues in screening of the disease-causing genes but are superior in follow-up analyses such as silencing of genes or introducing and removing mutations via CRISPR. In Figure 6 I show the example that we found *PLIN2* correlated with the degree of NAFLD in the meta-analysis and in subsequent studies we investigated the effect of silencing *PLIN2* in iPSC-models. Analogously, *PNPLA3* or *TM6SF2* which had been found in genome-wide association studies associated with NAFLD could be explored in more detail in personalized iPSC-models of patients carrying these mutations. Moreover, the mutations could be introduced or removed in iPSC-models via CRISPR. In that sense the results of the meta-analysis can be explored in more detail by entering a cycle of modeling – iPSC based or also Systems biology based - experiments and hypothesis generation. New hypotheses can be tested by adjustment of the model or by new experiments with the model.

3.9 Conclusion

In this thesis, I have compared the transcriptome analysis of stem-cell based disease models with meta-analyses of biopsy-derived transcriptome data. Main focuses were the liver and the brain considering NAFLD in liver and Alzheimer's disease and aging in the brain. However, also urine-derived renal progenitor cells were assessed laying the foundation for kidney disease models (Rahman et al. 2020). Basis of the stem-cell disease models is the quality of the reprogramming and differentiation process. Here, I could show with bioinformatics methods that transcriptomes of the derived cell types resembled the requested target cell types. However, in comparison to a fully mature phenotype often some factors are missing what could be shown by differences in marker gene expression between stem-cell- and biopsy-derived cells, e.g. not fully mature albumin and cytochrome expression in hepatocyte-like-cells. Nevertheless, the comparison showed that the stem-cell-based models - although not fully mature in some details of gene expression - can very well capture most phenotypes of diseases: the essential marker of lipid droplets (PLIN2) found in NAFLD patients' biopsies (Wruck et al. 2015) could be recapitulated with the iPSC-based NAFLD model (Graffmann et al. 2016). The same hold for Alzheimer's disease where in the iPSC-models of sporadic AD patients features of the disease could be detected (Hossini et al. 2015) such as the expression of p-tau and differential expression of genes involved in e.g. the ubiquitin-proteasome system and response to oxidative stress. In a study of brain aging in prefrontal cortex samples I developed algorithms for the analysis of gene expression time series (Wruck and Adjaye 2020). These provide the means to elucidate causality between genes via the Granger test, i.e. a test if a time series of a gene can be better explained by a linear model using additional lagged time series data from another potentially causal gene than by the data

of the gene alone. The notion of Granger causality originates in the econometric field and was first presented by the name-giving author Granger giving a somewhat ironic example by answering the question if chicken or eggs came first (Granger 1969). In the prefrontal cortex meta-analysis I could detect a granger causality between the gene CAMK4 and GFAP which was the most outstanding gene increasing with age indicating the growing number of astrocytes as reaction to neural destruction (Wruck and Adjaye 2020).

In summary, I want to conclude that analyses of biopsy data have the advantage that they are closest to the target tissue - possibly only biased by post-mortem interval, individual variability, biopsy location while stem-cell based models have to cope with reaching the mature phenotype of the target cell. However, here I could show that they are already able to reproduce many disease features and that there is considerable overlap in expression of essential marker genes between biopsy- and stem-cell-based models. In future, with optimized stem cell based models it will be possible to generate even more realistic disease models and for particularly hard or impossible to access tissues they can provide personalized disease models during life-time of the patient. These models then would also allow testing of manifold research hypotheses and therapy options.

References

Adjaye, James. 2005. "Whole-Genome Approaches for Large-Scale Gene Identification and Expression Analysis in Mammalian Preimplantation Embryos." *Reproduction, Fertility, and Development* 17 (1-2): 37-45. https://doi.org/10.1071/rd04075.

Adjaye, James, John Huntriss, Ralf Herwig, Alia BenKahla, Thore C. Brink, Christoph Wierling, Claus Hultschig, et al. 2005. "Primary Differentiation in the Human Blastocyst: Comparative Molecular Portraits of Inner Cell Mass and Trophectoderm Cells." *STEM CELLS* 23 (10): 1514-25. https://doi.org/10.1634/stemcells.2005-0113.

An, Mahru C., Ningzhe Zhang, Gary Scott, Daniel Montoro, Tobias Wittkop, Sean Mooney, Simon Melov, and Lisa M. Ellerby. 2012. "Genetic Correction of Huntington's Disease Phenotypes in Induced Pluripotent Stem Cells." *Cell Stem Cell* 11 (2): 253-63. https://doi.org/10.1016/j.stem.2012.04.026.

Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, et al. 2000. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25 (1): 25-29. https://doi.org/10.1038/75556.

Babaie, Yasmin, Ralf Herwig, Boris Greber, Thore C Brink, Wasco Wruck, Detlef Groth, Hans Lehrach, Tom Burdon, and James Adjaye. 2007. "Analysis of Oct4-Dependent Transcriptional Networks Regulating Self-Renewal and Pluripotency in Human Embryonic Stem Cells." *Stem Cells (Dayton, Ohio)* 25 (2): 500-510. https://doi.org/10.1634/stemcells.2006-0426.

Bar-Joseph, Z. 2004. "Analyzing Time Series Gene Expression Data." *Bioinformatics* 20 (16): 2493-2503. https://doi.org/10.1093/bioinformatics/bth283.

Berger, Dustin R., Brenton R. Ware, Matthew D. Davidson, Samuel R. Allsup, and Salman R. Khetani. 2015. "Enhancing the Functional Maturity of Induced Pluripotent Stem Cell-Derived Human Hepatocytes by Controlled Presentation of Cell-Cell Interactions in Vitro." *Hepatology (Baltimore, Md.)* 61 (4): 1370-81. https://doi.org/10.1002/hep.27621.

Bohndorf, Martina, Audrey Ncube, Lucas-Sebastian Spitzhorn, Jürgen Enczmann, Wasco Wruck, and James Adjaye. 2017. "Derivation and Characterization of Integration-Free IPSC Line ISRM-UM51 Derived from SIX2-Positive Renal Cells Isolated from Urine of an African Male Expressing the CYP2D6 *4/*17 Variant Which Confers Intermediate Drug Metabolizing Activity." *Stem Cell Research* 25 (December): 18-21. https://doi.org/10.1016/j.scr.2017.10.004.

Brennand, Kristen J., Anthony Simone, Jessica Jou, Chelsea Gelboin-Burkhart, Ngoc Tran, Sarah Sangar, Yan Li, et al. 2011. "Modelling Schizophrenia Using Human Induced Pluripotent Stem Cells." *Nature* 473 (7346): 221-25. https://doi.org/10.1038/nature09915.

Cahan, Patrick, Hu Li, Samantha A. Morris, Edroaldo Lummertz da Rocha, George Q. Daley, and James J. Collins. 2014. "CellNet: Network Biology Applied to Stem Cell Engineering." *Cell* 158 (4): 903-15. https://doi.org/10.1016/j.cell.2014.07.020.

Canzar, Stefan, Sandro Andreotti, David Weese, Knut Reinert, and Gunnar W. Klau. 2016. "CIDANE: Comprehensive Isoform Discovery and Abundance Estimation." *Genome Biology* 17 (January): 16. https://doi.org/10.1186/s13059-015-0865-0.

Cao, Lining, Rui Hu, Ting Xu, Zhen-Ning Zhang, Weida Li, and Jianfeng Lu. 2017. "Characterization of Induced Pluripotent Stem Cell-Derived Human Serotonergic Neurons." *Frontiers in Cellular Neuroscience* 11 (May). https://doi.org/10.3389/fncel.2017.00131.

Chang, T. W. 1983. "Binding of Cells to Matrixes of Distinct Antibodies Coated on Solid Surface." *Journal of Immunological Methods* 65 (1-2): 217-23. https://doi.org/10.1016/0022-1759(83)90318-6.

Clarke, Laura, Xiangqun Zheng-Bradley, Richard Smith, Eugene Kulesha, Chunlin Xiao, Iliana Toneva, Brendan Vaughan, et al. 2012. "The 1000 Genomes Project: Data Management and Community Access." *Nature Methods* 9 (5): 459-62. https://doi.org/10.1038/nmeth.1974.

Cleveland, William S. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74 (368): 829-36. https://doi.org/10.1080/01621459.1979.10481038.

Cooper, Oliver, Hyemyung Seo, Shaida Andrabi, Cristina Guardia-Laguarta, John Graziotto, Maria Sundberg, Jesse R. McLean, et al. 2012. "Pharmacological Rescue of Mitochondrial Deficits in IPSC-Derived Neural Cells from Patients with Familial Parkinson's Disease." *Science Translational Medicine* 4 (141): 141ra90. https://doi.org/10.1126/scitranslmed.3003985.

De Los Angeles, Alejandro, Francesco Ferrari, Ruibin Xi, Yuko Fujiwara, Nissim Benvenisty, Hongkui Deng, Konrad Hochedlinger, et al. 2015. "Hallmarks of Pluripotency." *Nature* 525 (7570): 469-78. https://doi.org/10.1038/nature15515.

Delcher, Arthur L., Adam Phillippy, Jane Carlton, and Steven L. Salzberg. 2002. "Fast Algorithms for Large-Scale Genome Alignment and Comparison." *Nucleic Acids Research* 30 (11): 2478-83. https://doi.org/10.1093/nar/30.11.2478.

Devika, A. S., Wasco Wruck, James Adjaye, and Smita Sudheer. 2019. "The Quest for Pluripotency: A Comparative Analysis across Mammalian Species." *Reproduction (Cambridge, England)*, April. https://doi.org/10.1530/REP-18-0083.

Dittrich, M. T., G. W. Klau, A. Rosenwald, T. Dandekar, and T. Muller. 2008. "Identifying Functional Modules in Protein-Protein Interaction Networks: An Integrated Exact Approach." *Bioinformatics* 24 (13): i223-31. https://doi.org/10.1093/bioinformatics/btn161.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics* 29 (1): 15-21. https://doi.org/10.1093/bioinformatics/bts635.

Döring, Andreas, David Weese, Tobias Rausch, and Knut Reinert. 2008. "SeqAn an Efficient, Generic C++ Library for Sequence Analysis." *BMC Bioinformatics* 9 (January): 11. https://doi.org/10.1186/1471-2105-9-11.

Doudna, Jennifer A., and Emmanuelle Charpentier. 2014. "The New Frontier of Genome Engineering with CRISPR-Cas9." *Science* 346 (6213): 1258096. https://doi.org/10.1126/science.1258096.

Du, Pan, Warren A. Kibbe, and Simon M. Lin. 2008. "Lumi: A Pipeline for Processing Illumina Microarray." *Bioinformatics* 24 (13): 1547-48. https://doi.org/10.1093/bioinformatics/btn224.

Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein. 1998. "Cluster Analysis and Display of Genome-Wide Expression Patterns." *Proceedings of the National Academy of Sciences of the United States of America* 95 (25): 14863-68. https://doi.org/10.1073/pnas.95.25.14863.

Evans, M. J., and M. H. Kaufman. 1981. "Establishment in Culture of Pluripotential Cells from Mouse Embryos." *Nature* 292 (5819): 154-56. https://doi.org/10.1038/292154a0.

Falcon, S., and R. Gentleman. 2007. "Using GOstats to Test Gene Lists for GO Term Association." *Bioinformatics (Oxford, England)* 23 (2): 257-58. https://doi.org/10.1093/bioinformatics/btl567.

Feldstein, Ariel E., Anna Wieckowska, A. Rocio Lopez, Yao-Chang Liu, Nizar N. Zein, and Arthur J. McCullough. 2009. "Cytokeratin-18 Fragment Levels as Noninvasive Biomarkers for Nonalcoholic Steatohepatitis: A Multicenter Validation Study." *Hepatology* 50 (4): 1072-1078. https://doi.org/10.1002/hep.23050.

Frankenberg, Stephen, and Marilyn B. Renfree. 2013. "On the Origin of POU5F1." *BMC Biology* 11 (1): 56. https://doi.org/10.1186/1741-7007-11-56.

Gautier, Laurent, Leslie Cope, Benjamin M. Bolstad, and Rafael A. Irizarry. 2004. "Affy–Analysis of Affymetrix GeneChip Data at the Probe Level." *Bioinformatics (Oxford, England)* 20 (3): 307-15. https://doi.org/10.1093/bioinformatics/btg405.

Gentleman, Robert C, Vincent J Carey, Douglas M Bates, Ben Bolstad, Marcel Dettling, Sandrine Dudoit, Byron Ellis, et al. 2004. "Bioconductor: Open Software Development for Computational Biology and Bioinformatics." *Genome Biology* 5 (10): R80. https://doi.org/10.1186/gb-2004-5-10-r80.

Gerets, H. H. J., K. Tilmant, B. Gerin, H. Chanteux, B. O. Depelchin, S. Dhalluin, and F. A. Atienzar. 2012. "Characterization of Primary Human Hepatocytes, HepG2 Cells, and HepaRG Cells at the MRNA Level and CYP Activity in Response to Inducers and Their Predictivity for the Detection of Human Hepatotoxins." *Cell Biology and Toxicology* 28 (2): 69-87. https://doi.org/10.1007/s10565-011-9208-4.

Gieseck III, Richard L., Nicholas R. F. Hannan, Roque Bort, Neil A. Hanley, Rosemary A. L. Drake, Grant W. W. Cameron, Thomas A. Wynn, and Ludovic Vallier. 2014. "Maturation of Induced Pluripotent Stem Cell Derived Hepatocytes by 3D-Culture." Edited by Majlinda Lako. *PLoS ONE* 9 (1): e86372. https://doi.org/10.1371/journal.pone.0086372.

Goeman, Jelle J., Sara A. van de Geer, Floor de Kort, and Hans C. van Houwelingen. 2004. "A Global Test for Groups of Genes: Testing Association with a Clinical Outcome." *Bioinformatics (Oxford, England)* 20 (1): 93-99. https://doi.org/10.1093/bioinformatics/btg382.

Goodwin, Sara, John D. McPherson, and W. Richard McCombie. 2016. "Coming of Age: Ten Years of next-Generation Sequencing Technologies." *Nature Reviews. Genetics* 17 (6): 333-51. https://doi.org/10.1038/nrg.2016.49.

Graffmann, Nina, Sarah Ring, Marie-Ann Kawala, Wasco Wruck, Audrey Ncube, Hans-Ingo Trompeter, and James Adjaye. 2016. "Modeling Nonalcoholic Fatty Liver Disease with Human Pluripotent Stem Cell-Derived Immature Hepatocyte-Like Cells Reveals Activation of PLIN2 and Confirms Regulatory Functions of Peroxisome Proliferator-Activated Receptor Alpha." *Stem Cells and Development* 25 (15): 1119-33. https://doi.org/10.1089/scd.2015.0383.

Granger, C. W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37 (3): 424-38. https://doi.org/10.2307/1912791.

GTEx Consortium. 2013. "The Genotype-Tissue Expression (GTEx) Project." *Nature Genetics* 45 (6): 580-85. https://doi.org/10.1038/ng.2653.

Gunderson, Kevin L., Semyon Kruglyak, Michael S. Graige, Francisco Garcia, Bahram G. Kermani, Chanfeng Zhao, Diping Che, et al. 2004. "Decoding Randomly Ordered DNA Arrays." *Genome Research* 14 (5): 870-77. https://doi.org/10.1101/gr.2255804.

Hadfield, James, Colin Megill, Sidney M. Bell, John Huddleston, Barney Potter, Charlton Callender, Pavel Sagulenko, Trevor Bedford, and Richard A. Neher. 2018. "Nextstrain: Real-Time Tracking of Pathogen Evolution." *Bioinformatics* 34 (23): 4121-23. https://doi.org/10.1093/bioinformatics/bty407.

Hänzelmann, Sonja, Robert Castelo, and Justin Guinney. 2013. "GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data." *BMC Bioinformatics* 14 (January): 7. https://doi.org/10.1186/1471-2105-14-7.

Hay, David C., Debiao Zhao, Judy Fletcher, Zoë A. Hewitt, Doris McLean, Alai Urruticoechea-Uriguen, James R. Black, et al. 2008. "Efficient Differentiation of Hepatocytes from Human Embryonic Stem Cells Exhibiting Markers Recapitulating Liver Development in Vivo." *Stem Cells (Dayton, Ohio)* 26 (4): 894-902. https://doi.org/10.1634/stemcells.2007-0718.

Heinemeyer, T., X. Chen, H. Karas, A. E. Kel, O. V. Kel, I. Liebich, T. Meinhardt, I. Reuter, F. Schacherer, and E. Wingender. 1999. "Expanding the TRANSFAC Database towards an Expert System of Regulatory Molecular Mechanisms." *Nucleic Acids Research* 27 (1): 318-22. https://doi.org/10.1093/nar/27.1.318.

Herwig, R. 1999. "Large-Scale Clustering of CDNA-Fingerprinting Data." *Genome Research* 9 (11): 1093-1105. https://doi.org/10.1101/gr.9.11.1093.

Herwig, R., P. Aanstad, M. Clark, and H. Lehrach. 2001. "Statistical Evaluation of Differential Expression on CDNA Nylon Arrays with Replicated Experiments." *Nucleic Acids Research* 29 (23): E117. https://doi.org/10.1093/nar/29.23.e117.

Hossini, Amir M., Matthias Megges, Alessandro Prigione, Bjoern Lichtner, Mohammad R. Toliat, Wasco Wruck, Friederike Schröter, et al. 2015. "Induced Pluripotent Stem Cell-Derived Neuronal Cells from a Sporadic Alzheimer's Disease Donor as a Model for Investigating AD-Associated Gene Regulatory Networks." *BMC Genomics* 16: 84. https://doi.org/10.1186/s12864-015-1262-5.

Huang, Da Wei, Brad T Sherman, and Richard A Lempicki. 2009. "Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources." *Nature Protocols* 4 (1): 44-57. https://doi.org/10.1038/nprot.2008.211.

Hubbard, T., D. Barker, E. Birney, G. Cameron, Y. Chen, L. Clark, T. Cox, et al. 2002. "The Ensembl Genome Database Project." *Nucleic Acids Research* 30 (1): 38-41. https://doi.org/10.1093/nar/30.1.38.

Huber, Wolfgang, Anja von Heydebreck, Holger Sültmann, Annemarie Poustka, and Martin Vingron. 2002. "Variance Stabilization Applied to Microarray Data Calibration and to the Quantification of Differential Expression." *Bioinformatics (Oxford, England)* 18 Suppl 1: S96-104. https://doi.org/10.1093/bioinformatics/18.suppl_1.s96.

Ideker, Trey, Owen Ozier, Benno Schwikowski, and Andrew F. Siegel. 2002. "Discovering Regulatory and Signalling Circuits in Molecular Interaction Networks." *Bioinformatics (Oxford, England)* 18 Suppl 1: S233-240. https://doi.org/10.1093/bioinformatics/18.suppl_1.s233.

Ikonen, Elina. 2008. "Cellular Cholesterol Trafficking and Compartmentalization." *Nature Reviews Molecular Cell Biology* 9 (2): 125-38. https://doi.org/10.1038/nrm2336.

Israel, Mason A., Shauna H. Yuan, Cedric Bardy, Sol M. Reyna, Yangling Mu, Cheryl Herrera, Michael P. Hefferan, et al. 2012. "Probing Sporadic and Familial Alzheimer's Disease Using Induced Pluripotent Stem Cells." *Nature* 482 (7384): 216-20. https://doi.org/10.1038/nature10821.

Jones, Allan R., Caroline C. Overly, and Susan M. Sunkin. 2009. "The Allen Brain Atlas: 5 Years and Beyond." *Nature Reviews. Neuroscience* 10 (11): 821-28. https://doi.org/10.1038/nrn2722.

Joshi-Tope, G., M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, et al. 2005. "Reactome: A Knowledgebase of Biological Pathways." *Nucleic Acids Research* 33 (Database issue): D428-432. https://doi.org/10.1093/nar/gki072.

Józefczuk, Justyna. 2009. *Differentiation of Human Embryonic Stem Cells into Hepatocytes as a Tool to Analyse Dynamic Regulatory Events during Hepatogenesis in Vitro.* Berlin: Dissertation at Free University Berlin.

Jozefczuk, Justyna, Karl Kashofer, Ramesh Ummanni, Frauke Henjes, Samrina Rehman, Suzanne Geenen, Wasco Wruck, et al. 2012. "A Systems Biology Approach to Deciphering the Etiology of Steatosis Employing Patient-Derived Dermal Fibroblasts and IPS Cells." *Frontiers in Physiology* 3: 339. https://doi.org/10.3389/fphys.2012.00339.

Jozefczuk, Justyna, Alessandro Prigione, Lukas Chavez, and James Adjaye. 2010. "Comparative Analysis of Human Embryonic Stem Cell and Induced Pluripotent Stem Cell-Derived Hepatocyte-Like Cells Reveals Current Drawbacks and Possible Strategies for Improved Differentiation." *Stem Cells and Development* 20 (7): 1259-75. https://doi.org/10.1089/scd.2010.0361.

Kamburov, Atanas, Konstantin Pentchev, Hanna Galicka, Christoph Wierling, Hans Lehrach, and Ralf Herwig. 2011. "ConsensusPathDB: Toward a More Complete Picture of Cell Biology." *Nucleic Acids Research* 39 (suppl 1): D712-17. https://doi.org/10.1093/nar/gkq1156.

Kanehisa, Minoru, Susumu Goto, Miho Furumichi, Mao Tanabe, and Mika Hirakawa. 2010. "KEGG for Representation and Analysis of Molecular Networks Involving Diseases and Drugs." *Nucleic Acids Research* 38 (suppl 1): D355-60. https://doi.org/10.1093/nar/gkp896.

Katoh, Kazutaka, Kazuharu Misawa, Kei-ichi Kuma, and Takashi Miyata. 2002. "MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform." *Nucleic Acids Research* 30 (14): 3059-66. https://doi.org/10.1093/nar/gkf436.

Kim, Daehwan, Ben Langmead, and Steven L Salzberg. 2015. "HISAT: A Fast Spliced Aligner with Low Memory Requirements." *Nature Methods* 12 (4): 357-60. https://doi.org/10.1038/nmeth.3317.

Kleiner, David E., Elizabeth M. Brunt, Mark Van Natta, Cynthia Behling, Melissa J. Contos, Oscar W. Cummings, Linda D. Ferrell, et al. 2005. "Design and Validation of a Histological Scoring System for Nonalcoholic Fatty Liver Disease." *Hepatology* 41 (6): 1313-1321. https://doi.org/10.1002/hep.20701.

Kondo, Takayuki, Masashi Asai, Kayoko Tsukita, Yumiko Kutoku, Yutaka Ohsawa, Yoshihide Sunada, Keiko Imamura, et al. 2013. "Modeling Alzheimer's Disease with IPSCs Reveals Stress Phenotypes Associated with Intracellular Aâ and Differential Drug Responsiveness." *Cell Stem Cell* 12 (4): 487-96. https://doi.org/10.1016/j.stem.2013.01.009.

Korber, Bette, Will M. Fischer, Sandrasegaram Gnanakaran, Hyejin Yoon, James Theiler, Werner Abfalterer, Nick Hengartner, et al. 2020. "Tracking Changes in SARS-CoV-2 Spike: Evidence That D614G Increases Infectivity of the COVID-19 Virus." *Cell* 182 (4): 812-827.e19. https://doi.org/10.1016/j.cell.2020.06.043.

Kwon, Andrew T., David J. Arenillas, Rebecca Worsley Hunt, and Wyeth W. Wasserman. 2012. "OPOSSUM-3: Advanced Analysis of Regulatory Motif over-Representation across Genes or ChIP-Seq Datasets." *G3 (Bethesda, Md.)* 2 (9): 987-1002. https://doi.org/10.1534/g3.112.003202.

Laakso, M. P., H. Soininen, K. Partanen, E. -L. Helkala, P. Hartikainen, P. Vainio, M. Hallikainen, T. Hänninen, and P. J. Riekkinen Sr. 1995. "Volumes of Hippocampus, Amygdala and Frontal Lobes in the MRI-Based Diagnosis of Early Alzheimer's Disease: Correlation with Memory Functions." *Journal of Neural Transmission - Parkinson's Disease and Dementia Section* 9 (1): 73-86. https://doi.org/10.1007/BF02252964.

Laakso, Mikko P, Merja Hallikainen, Tuomo Hänninen, Kaarina Partanen, and Hilkka Soininen. 2000. "Diagnosis of Alzheimer's Disease: MRI of the Hippocampus vs Delayed Recall." *Neuropsychologia* 38 (5): 579-84. https://doi.org/10.1016/S0028-3932(99)00111-6.

Langfelder, Peter, and Steve Horvath. 2008. "WGCNA: An R Package for Weighted Correlation Network Analysis." *BMC Bioinformatics* 9 (1). https://doi.org/10.1186/1471-2105-9-559.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. "Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome." *Genome Biology* 10 (3): R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Li, Heng, and Richard Durbin. 2010. "Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform." *Bioinformatics (Oxford, England)* 26 (5): 589-95. https://doi.org/10.1093/bioinformatics/btp698.

Li, J., G. Ning, and S. A. Duncan. 2000. "Mammalian Hepatocyte Differentiation Requires the Transcription Factor HNF-4alpha." *Genes & Development* 14 (4): 464-74.

Lindström, Nils O., Guilherme De Sena Brandine, Tracy Tran, Andrew Ransick, Gio Suh, Jinjin Guo, Albert D. Kim, et al. 2018. "Progressive Recruitment of Mesenchymal Progenitors Reveals a Time-Dependent Process of Cell Fate Acquisition in Mouse and Human Nephrogenesis." *Developmental Cell* 45 (5): 651-660.e4. https://doi.org/10.1016/j.devcel.2018.05.010.

Lippert, Ross A. 2005. "Space-Efficient Whole Genome Comparisons with Burrows-Wheeler Transforms." *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 12 (4): 407-15. https://doi.org/10.1089/cmb.2005.12.407.

Liu, Yan, Huisheng Liu, Conall Sauvey, Lin Yao, Ewa D. Zarnowska, and Su-Chun Zhang. 2013. "Directed Differentiation of Forebrain GABA Interneurons from Human Pluripotent Stem Cells." *Nature Protocols* 8 (9): 1670-79. https://doi.org/10.1038/nprot.2013.106.

Luo, W., and C. Brouwer. 2013. "Pathview: An R/Bioconductor Package for Pathway-Based Data Integration and Visualization." *Bioinformatics* 29 (14): 1830-31. https://doi.org/10.1093/bioinformatics/btt285.

Maere, Steven, Karel Heymans, and Martin Kuiper. 2005. "BiNGO: A Cytoscape Plugin to Assess Overrepresentation of Gene Ontology Categories in Biological Networks." *Bioinformatics (Oxford, England)* 21 (16): 3448-49. https://doi.org/10.1093/bioinformatics/bti551.

Mane, Shrinivasrao P., Clive Evans, Kristal L. Cooper, Oswald R. Crasta, Otto Folkerts, Stephen K. Hutchison, Timothy T. Harkins, Danielle Thierry-Mieg, Jean Thierry-Mieg, and Roderick V. Jensen. 2009. "Transcriptome Sequencing of the Microarray Quality Control (MAQC) RNA Reference Samples Using next Generation Sequencing." *BMC Genomics* 10: 264. https://doi.org/10.1186/1471-2164-10-264.

MAQC Consortium, Leming Shi, Laura H. Reid, Wendell D. Jones, Richard Shippy, Janet A. Warrington, Shawn C. Baker, et al. 2006. "The MicroArray Quality Control (MAQC) Project Shows Inter- and Intraplatform Reproducibility of Gene Expression Measurements." *Nature Biotechnology* 24 (9): 1151-61. https://doi.org/10.1038/nbt1239.

Martins, Soraia, Martina Bohndorf, Nina Graffmann, Wasco Wruck, Krystyna H. Chrzanowska, and James Adjaye. 2019. "Fibroblast-Derived Integration-Free IPSC Line ISRM-NBS1 from an 18-Year-Old Nijmegen Breakage Syndrome Patient Carrying the Homozygous NBN c.657_661del5 Mutation." *Stem Cell Research* 34 (January): 101372. https://doi.org/10.1016/j.scr.2018.101372.

Martins, Soraia, Andreas Müller-Schiffmann, Lars Erichsen, Martina Bohndorf, Wasco Wruck, Kristel Sleegers, Christine Van Broeckhoven, Carsten Korth, and James Adjaye. 2020. "IPSC-Derived Neuronal Cultures Carrying the Alzheimer's Disease Associated TREM2 R47H Variant Enables the Construction of an Aβ-Induced Gene Regulatory Network." *International Journal of Molecular Sciences* 21 (12): 4516. https://doi.org/10.3390/ijms21124516.

Matz, Peggy, Wasco Wruck, Beatrix Fauler, Diran Herebian, Thorsten Mielke, and James Adjaye. 2017. "Footprint-Free Human Fetal Foreskin Derived IPSCs: A Tool for Modeling Hepatogenesis Associated Gene Regulatory Networks." *Scientific Reports* 7 (1): 6294. https://doi.org/10.1038/s41598-017-06546-9.

McCall, Matthew N., Benjamin M. Bolstad, and Rafael A. Irizarry. 2010. "Frozen Robust Multiarray Analysis (FRMA)." *Biostatistics (Oxford, England)* 11 (2): 242-53. https://doi.org/10.1093/biostatistics/kxp059.

Mehta, Shagun R., Colton M. Tom, Yizhou Wang, Catherine Bresee, David Rushton, Pranav P. Mathkar, Jie Tang, and Virginia B. Mattis. 2018. "Human Huntington's Disease IPSC-Derived Cortical Neurons Display Altered Transcriptomics, Morphology, and Maturation." *Cell Reports* 25 (4): 1081-1096.e6. https://doi.org/10.1016/j.celrep.2018.09.076.

Min, Hae-Ki, Ashwani Kapoor, Michael Fuchs, Faridoddin Mirshahi, Huiping Zhou, James Maher, John Kellum, Russell Warnick, Melissa J. Contos, and Arun J. Sanyal. 2012. "Increased Hepatic Synthesis and Dysregulation of Cholesterol Metabolism Is Associated with the Severity of Nonalcoholic Fatty Liver Disease." *Cell Metabolism* 15 (5): 665-74. https://doi.org/10.1016/j.cmet.2012.04.004.

Mlody, Barbara, Wasco Wruck, Soraia Martins, Karl Sperling, and James Adjaye. 2017. "Nijmegen Breakage Syndrome Fibroblasts and IPSCs: Cellular Models for Uncovering Disease-Associated Signaling Pathways and Establishing a Screening Platform for Anti-Oxidants." *Scientific Reports* 7 (1). https://doi.org/10.1038/s41598-017-07905-2.

Moreno, Cesar L., Lucio Della Guardia, Valeria Shnyder, Maitane Ortiz-Virumbrales, Ilya Kruglikov, Bin Zhang, Eric E. Schadt, et al. 2018. "IPSC-Derived Familial Alzheimer's PSEN2 N141I Cholinergic Neurons Exhibit Mutation-Dependent Molecular Pathology Corrected by Insulin Signaling." *Molecular Neurodegeneration* 13 (1): 33. https://doi.org/10.1186/s13024-018-0265-5.

Mu, Yangling, and Fred H Gage. 2011. "Adult Hippocampal Neurogenesis and Its Role in Alzheimer's Disease." *Molecular Neurodegeneration* 6 (1): 85. https://doi.org/10.1186/1750-1326-6-85.

Müller, Franz-Josef, Bernhard M. Schuldt, Roy Williams, Dylan Mason, Gulsah Altun, Eirini P. Papapetrou, Sandra Danner, et al. 2011. "A Bioinformatic Assay for Pluripotency in Human Cells." *Nature Methods* 8 (4): 315-17. https://doi.org/10.1038/nmeth.1580.

Musso, Giovanni, Roberto Gambino, and Maurizio Cassader. 2013. "Cholesterol Metabolism and the Pathogenesis of Non-Alcoholic Steatohepatitis." *Progress in Lipid Research* 52 (1): 175-91. https://doi.org/10.1016/j.plipres.2012.11.002.

Neuschwander-Tetri, Brent A, Rohit Loomba, Arun J Sanyal, Joel E Lavine, Mark L Van Natta, Manal F Abdelmalek, Naga Chalasani, et al. 2015. "Farnesoid X Nuclear Receptor Ligand Obeticholic Acid for Non-Cirrhotic, Non-Alcoholic Steatohepatitis (FLINT): A Multicentre, Randomised, Placebo-Controlled Trial." *The Lancet* 385 (9972): 956-65. https://doi.org/10.1016/S0140-6736(14)61933-4.

Nichols, J., B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Schöler, and A. Smith. 1998. "Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4." *Cell* 95 (3): 379-91. https://doi.org/10.1016/s0092-8674(00)81769-9.

Oughtred, Rose, Chris Stark, Bobby-Joe Breitkreutz, Jennifer Rust, Lorrie Boucher, Christie Chang, Nadine Kolas, et al. 2019. "The BioGRID Interaction Database: 2019 Update." *Nucleic Acids Research* 47 (D1): D529-41. https://doi.org/10.1093/nar/gky1079.

Pan, Jian-Bo, Shi-Chang Hu, Dan Shi, Mei-Chun Cai, Yin-Bo Li, Quan Zou, and Zhi-Liang Ji. 2013. "PaGenBase: A Pattern Gene Database for the Global and Dynamic Understanding of Gene Function." *PloS One* 8 (12): e80747. https://doi.org/10.1371/journal.pone.0080747.

Park, Chul-Yong, Tomer Halevy, Dongjin R. Lee, Jin Jea Sung, Jae Souk Lee, Ofra Yanuka, Nissim Benvenisty, and Dong-Wook Kim. 2015. "Reversion of FMR1 Methylation and Silencing by Editing the Triplet Repeats in Fragile X IPSC-Derived Neurons." *Cell Reports* 13 (2): 234-41. https://doi.org/10.1016/j.celrep.2015.08.084.

Pertea, Mihaela, Geo M. Pertea, Corina M. Antonescu, Tsung-Cheng Chang, Joshua T. Mendell, and Steven L. Salzberg. 2015. "StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads." *Nature Biotechnology* 33 (3): 290-95. https://doi.org/10.1038/nbt.3122.

Pesce, M., M. K. Gross, and H. R. Schöler. 1998. "In Line with Our Ancestors: Oct-4 and the Mammalian Germ." *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 20 (9): 722-32. https://doi.org/10.1002/(SICI)1521-1878(199809)20:9<722::AID-BIES5>3.0.CO;2-I.

Puelles, Victor G., Marc Lütgehetmann, Maja T. Lindenmeyer, Jan P. Sperhake, Milagros N. Wong, Lena Allweiss, Silvia Chilla, et al. 2020. "Multiorgan and Renal Tropism of SARS-CoV-2." *New England Journal of Medicine* 383 (6): 590–92. https://doi.org/10.1056/NEJMc2011400.

Rahman, Md Shaifur, Wasco Wruck, Lucas-Sebastian Spitzhorn, Lisa Nguyen, Martina Bohndorf, Soraia Martins, Fatima Asar, et al. 2020. "The FGF, TGFâ and WNT Axis Modulate Self-Renewal of Human SIX2+ Urine Derived Renal Progenitor Cells." *Scientific Reports* 10 (1): 739. https://doi.org/10.1038/s41598-020-57723-2.

Ratziu, Vlad, Julien Massard, Frederic Charlotte, Djamila Messous, Françoise Imbert-Bismut, Luninita Bonyhay, Mohamed Tahiri, et al. 2006. "Diagnostic Value of Biochemical Markers (FibroTest-FibroSURE) for the Prediction of Liver Fibrosis in Patients with Non-Alcoholic Fatty Liver Disease." *BMC Gastroenterology* 6: 6. https://doi.org/10.1186/1471-230X-6-6.

Rinella, Mary E., Marc S. Elias, Robin R. Smolak, Tao Fu, Jayme Borensztajn, and Richard M. Green. 2008. "Mechanisms of Hepatic Steatosis in Mice Fed a Lipogenic Methionine Choline-Deficient Diet." *Journal of Lipid Research* 49 (5): 1068-76. https://doi.org/10.1194/jlr.M800042-JLR200.

Roost, Matthias S., Liesbeth van Iperen, Yavuz Ariyurek, Henk P. Buermans, Wibowo Arindrarto, Harsha D. Devalla, Robert Passier, et al. 2015. "KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas." *Stem Cell Reports* 4 (6): 1112-24. https://doi.org/10.1016/j.stemcr.2015.05.002.

Sagulenko, Pavel, Vadim Puller, and Richard A. Neher. 2018. "TreeTime: Maximum-Likelihood Phylodynamic Analysis." *Virus Evolution* 4 (1). https://doi.org/10.1093/ve/vex042.

Sandelin, Albin, Wynand Alkema, Pär Engström, Wyeth W. Wasserman, and Boris Lenhard. 2004. "JASPAR: An Open-Access Database for Eukaryotic Transcription Factor Binding Profiles." *Nucleic Acids Research* 32 (Database issue): D91-94. https://doi.org/10.1093/nar/gkh012.

Schena, M., D. Shalon, R. W. Davis, and P. O. Brown. 1995. "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray." *Science* 270 (5235): 467-70. https://doi.org/10.1126/science.270.5235.467.

Schröter, Friederike, Kristel Sleegers, Elise Cuyvers, Martina Bohndorf, Wasco Wruck, Christine Van Broeckhoven, and James Adjaye. 2016. "Lymphoblast-Derived Integration-Free IPS Cell Line from a 65-Year-Old Alzheimer's Disease Patient Expressing the TREM2 p.R47H Variant." *Stem Cell Research* 16 (1): 113-15. https://doi.org/10.1016/j.scr.2015.12.017.

Slenter, Denise N., Martina Kutmon, Kristina Hanspers, Anders Riutta, Jacob Windsor, Nuno Nunes, Jonathan Mélius, et al. 2018. "WikiPathways: A Multifaceted Pathway Database Bridging Metabolomics to Other Omics Research." *Nucleic Acids Research* 46 (D1): D661-67. https://doi.org/10.1093/nar/gkx1064.

Smyth, Gordon K. 2004. "Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments." *Statistical Applications in Genetics and Molecular Biology* 3 (1). https://doi.org/10.2202/1544-6115.1027.

Spitzhorn, Lucas-Sebastian, Matthias Megges, Wasco Wruck, Md Shaifur Rahman, Jörg Otte, Özer Degistirici, Roland Meisel, Rüdiger Volker Sorg, Richard O. C. Oreffo, and James Adjaye. 2019. "Human IPSC-Derived MSCs (IMSCs) from Aged Individuals Acquire a Rejuvenation Signature." *Stem Cell Research & Therapy* 10 (1): 100. https://doi.org/10.1186/s13287-019-1209-x.

Steinfath, Matthias, Wasco Wruck, Henrik Seidel, Hans Lehrach, Uwe Radelof, and John O'Brien. 2001. "Automated Image Analysis for Array Hybridization Experiments." *Bioinformatics* 17 (7): 634-641.

Storey, John D. 2002. "A Direct Approach to False Discovery Rates." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (3): 479-98. https://doi.org/10.1111/1467-9868.00346.

Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, et al. 2005. "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles." *Proceedings of the National Academy of Sciences of the United States of America* 102 (43): 15545-50. https://doi.org/10.1073/pnas.0506580102.

Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. "REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms." Edited by Cynthia Gibas. *PLoS ONE* 6 (7): e21800. https://doi.org/10.1371/journal.pone.0021800.

Szklarczyk, Damian, Andrea Franceschini, Stefan Wyder, Kristoffer Forslund, Davide Heller, Jaime Huerta-Cepas, Milan Simonovic, et al. 2015. "STRING V10: Protein-Protein Interaction Networks, Integrated over the Tree of Life." *Nucleic Acids Research* 43 (Database issue): D447-452. https://doi.org/10.1093/nar/gku1003.

Takahashi, Kazutoshi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda, and Shinya Yamanaka. 2007. "Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors." *Cell* 131 (5): 861-72. https://doi.org/10.1016/j.cell.2007.11.019.

Takahashi, Kazutoshi, and Shinya Yamanaka. 2006. "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors." *Cell* 126 (4): 663-76. https://doi.org/10.1016/j.cell.2006.07.024.

Tarca, Adi L., Gaurav Bhatti, and Roberto Romero. 2013. "A Comparison of Gene Set Analysis Methods in Terms of Sensitivity, Prioritization and Specificity." *PloS One* 8 (11): e79217. https://doi.org/10.1371/journal.pone.0079217.

Thomas, Paul D., David P. Hill, Huaiyu Mi, David Osumi-Sutherland, Kimberly Van Auken, Seth Carbon, James P. Balhoff, et al. 2019. "Gene Ontology Causal Activity Modeling (GO-CAM) Moves beyond GO Annotations to Structured Descriptions of Biological Functions and Systems." *Nature Genetics* 51 (10): 1429-33. https://doi.org/10.1038/s41588-019-0500-1.

Thomson, J. A., J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, and J. M. Jones. 1998. "Embryonic Stem Cell Lines Derived from Human Blastocysts." *Science (New York, N.Y.)* 282 (5391): 1145-47.

Thurman, Walter N., and Mark E. Fisher. 1988. "Chickens, Eggs, and Causality, or Which Came First?" *American Journal of Agricultural Economics* 70 (2): 237-38. https://doi.org/10.2307/1242062.

Tomfohr, John, Jun Lu, and Thomas B. Kepler. 2005. "Pathway Level Analysis of Gene Expression Using Singular Value Decomposition." *BMC Bioinformatics* 6 (September): 225. https://doi.org/10.1186/1471-2105-6-225.

Trapnell, Cole, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R. Kelley, Harold Pimentel, Steven L. Salzberg, John L. Rinn, and Lior Pachter. 2012. "Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks." *Nature Protocols* 7 (3): 562-78. https://doi.org/10.1038/nprot.2012.016.

Tsuchida, Takuma, Youngmin A. Lee, Naoto Fujiwara, Maria Ybanez, Brittany Allen, Sebastiao Martins, M. Isabel Fiel, et al. 2018. "A Simple Diet- and Chemical-Induced Murine NASH Model with Rapid Progression of Steatohepatitis, Fibrosis and Liver Cancer." *Journal of Hepatology* 69 (2): 385-95. https://doi.org/10.1016/j.jhep.2018.03.011.

Wruck, Wasco, and James Adjaye. 2017. "Meta-Analysis Reveals up-Regulation of Cholesterol Processes in Non-Alcoholic and down-Regulation in Alcoholic Fatty Liver

Disease." *World Journal of Hepatology* 9 (8): 443. https://doi.org/10.4254/wjh.v9.i8.443.———

. 2018. "Human Pluripotent Stem Cell Derived HLC Transcriptome Data Enables Molecular Dissection of Hepatogenesis." *Scientific Data* 5: 180035. https://doi.org/10.1038/sdata.2018.35.———

. 2020. "Meta-Analysis of Human Prefrontal Cortex Reveals Activation of GFAP and Decline of Synaptic Transmission in the Aging Brain." *Acta Neuropathologica Communications* 8 (1): 26. https://doi.org/10.1186/s40478-020-00907-8.

Wruck, Wasco, Nina Graffmann, Marie-Ann Kawala, and James Adjaye. 2017. "Concise Review: Current Status and Future Directions on Research Related to Nonalcoholic Fatty Liver Disease." *Stem Cells (Dayton, Ohio)* 35 (1): 89-96. https://doi.org/10.1002/stem.2454.

Wruck, Wasco, Huw Griffiths, Matthias Steinfath, Hans Lehrach, Uwe Radelof, and John O'Brien. 2002. "Xdigitise: Visualization of Hybridization Experiments." *Bioinformatics* 18 (5): 757-60. https://doi.org/10.1093/bioinformatics/18.5.757.

Wruck, Wasco, Karl Kashofer, Samrina Rehman, Andriani Daskalaki, Daniela Berg, Ewa Gralka, Justyna Jozefczuk, et al. 2015. "Multi-Omic Profiles of Human Non-Alcoholic Fatty Liver Disease Tissue Highlight Heterogenic Phenotypes." *Scientific Data* 2 (December): 150068.

Wruck, Wasco, Friederike Schröter, and James Adjaye. 2016. "Meta-Analysis of Transcriptome Data Related to Hippocampus Biopsies and IPSC-Derived Neuronal Cells from Alzheimer's Disease Patients Reveals an Association with FOXA1 and FOXA2 Gene Regulatory Networks." *Journal of Alzheimer's Disease: JAD* 50 (4): 1065-82. https://doi.org/10.3233/JAD-150733.

Wu, Yi-Ying, Feng-Lan Chiu, Chan-Shien Yeh, and Hung-Chih Kuo. 2019. "Opportunities and Challenges for the Use of Induced Pluripotent Stem Cells in Modelling Neurodegenerative Disease." *Open Biology* 9 (1): 180177. https://doi.org/10.1098/rsob.180177.

Yagi, Takuya, Daisuke Ito, Yohei Okada, Wado Akamatsu, Yoshihiro Nihei, Takahito Yoshizaki, Shinya Yamanaka, Hideyuki Okano, and Norihiro Suzuki. 2011. "Modeling Familial Alzheimer's Disease with Induced Pluripotent Stem Cells." *Human Molecular Genetics* 20 (23): 4530-39. https://doi.org/10.1093/hmg/ddr394.

Zang, Ruochen, Maria Florencia Gomez Castro, Broc T. McCune, Qiru Zeng, Paul W. Rothlauf, Naomi M. Sonnek, Zhuoming Liu, et al. 2020. "TMPRSS2 and TMPRSS4 Promote SARS-CoV-2 Infection of Human Small Intestinal Enterocytes." *Science Immunology* 5 (47): eabc3582. https://doi.org/10.1126/sciimmunol.abc3582.

Zaret, Kenneth S. 2002. "Regulatory Phases of Early Liver Development: Paradigms of Organogenesis." *Nature Reviews. Genetics* 3 (7): 499-512. https://doi.org/10.1038/nrg837.

Zeeberg, Barry R., Weimin Feng, Geoffrey Wang, May D. Wang, Anthony T. Fojo, Margot Sunshine, Sudarshan Narasimhan, et al. 2003. "GoMiner: A Resource for Biological Interpretation of Genomic and Proteomic Data." *Genome Biology* 4 (4): R28. https://doi.org/10.1186/gb-2003-4-4-r28.

Zhang, Miao, Cristina D'Aniello, Arie O. Verkerk, Eva Wrobel, Stefan Frank, Dorien Ward-van Oostwaard, Ilaria Piccini, et al. 2014. "Recessive Cardiac Phenotypes in Induced Pluripotent Stem Cell Models of Jervell and Lange-Nielsen Syndrome: Disease Mechanisms and Pharmacological Rescue." *Proceedings of the National Academy of Sciences of the United States of America* 111 (50): E5383-5392. https://doi.org/10.1073/pnas.1419553111.

Zhang, Quanwei, and Zhengdong D. Zhang. 2013. "SubNet: A Java Application for Subnetwork Extraction." *Bioinformatics* 29 (19): 2509-11. https://doi.org/10.1093/bioinformatics/btt430.

Zhang, S. C., M. Wernig, I. D. Duncan, O. Brüstle, and J. A. Thomson. 2001. "In Vitro Differentiation of Transplantable Neural Precursors from Human Embryonic Stem Cells." *Nature Biotechnology* 19 (12): 1129-33. https://doi.org/10.1038/nbt1201-1129.

Zhao, Roong, and Stephen A. Duncan. 2005. "Embryonic Development of the Liver." *Hepatology (Baltimore, Md.)* 41 (5): 956-67. https://doi.org/10.1002/hep.20691.

Acknowledgements

I would like to thank Prof. Dr. James Adjaye for giving me the opportunity to work on so many interesting projects and to write my dissertation. I would like to express my great appreciation to Prof. Dr. Gunnar Klau for kindly acting as reviewer of my thesis as representative of the Faculty of Mathematics and Natural Sciences of the Heinrich-Heine-University Düsseldorf. Furthermore, I would like to thank all my colleagues at the Institute for Stem Cell Research and Regenerative Medicine for their support and the nice working atmosphere.

Appendix – Other publications

A.1. J Cell Mol Med. 2021 Jan 14. doi: 10.1111/jcmm.16222. [Epub ahead of print]

The pioneer and differentiation factor FOXA2 is a key driver of yolk-sac tumour formation and a new biomarker for paediatric and adult yolk-sac tumours.

Wruck W(1), Bremmer F(2), Kotthoff M(3), Fichtner A(2), Skowron MA(3), Schönberger S(4), Calaminus G(5), Vokuhl C(6), Pfister D(7), Heidenreich A(7), Albers P(8), Adjaye J(1), Nettersheim D(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, University Hospital Düsseldorf, Düsseldorf, Germany. (2)Institute of Pathology, University Medical Center Goettingen, Goettingen, Germany. (3)Department of Urology, Urological Research Lab, Translational UroOncology, University Hospital Düsseldorf, Düsseldorf, Germany. (4)Department of Pediatric Hematology and Oncology, University Children's Hospital, Essen, Germany. (5)Department of Pediatric Hematology and Oncology, University Hospital Bonn, Bonn, Germany. (6)Institute of Pathology, University Hospital Bonn, Bonn, Germany. (7)Department of Urology, University Hospital Cologne, Cologne, Germany. (8)Department of Urology, University Hospital Düsseldorf, Düsseldorf, Germany.

Yolk-sac tumours (YSTs), a germ cell tumour subtype, occur in newborns and infants as well as in young adults of age 14-44 years. In clinics, adult patients with YSTs face a poor prognosis, as these tumours are often therapy-resistant and count for many germ cell tumour related deaths. So far, the molecular and (epi)genetic mechanisms that control development of YST are far from being understood. We deciphered the molecular and (epi)genetic mechanisms regulating YST formation by meta-analysing high-throughput data of gene and microRNA expression, DNA methylation and mutational burden. We validated our findings by qRT-PCR and immunohistochemical analyses of paediatric and adult YSTs. On a molecular level, paediatric and adult YSTs were nearly indistinguishable, but were considerably different from embryonal carcinomas, the stem cell

precursor of YSTs. We identified FOXA2 as a putative key driver of YST formation, subsequently inducing AFP, GPC3, APOA1/APOB, ALB and GATA3/4/6 expression. In YSTs, WNT-, BMP- and MAPK signalling-related genes were up-regulated, while pluripotency- and (primordial) germ cell-associated genes were down-regulated. Expression of FOXA2 and related key factors seems to be regulated by DNA methylation, histone methylation / acetylation and microRNAs. Additionally, our results highlight FOXA2 as a promising new biomarker for paediatric and adult YSTs.

DOI: 10.1111/jcmm.16222 PMID: 33448076

A.2. BMC Genomics. 2020 Mar 30;21(1):265. doi: 10.1186/s12864-020-6684-z.

Transcriptomic analysis of marine endophytic fungi extract identifies highly enriched anti-fungal fractions targeting cancer pathways in HepG2 cell lines.

Blessie EJ(1), Wruck W(2), Abbey BA(1), Ncube A(2), Graffmann N(2), Amarh V(1), Arthur PK(3), Adjaye J(4)(5).

Author information: (1)West African Center for Cell Biology of Infectious Pathogens, Department of Biochemistry, Cell and Molecular Biology, University of Ghana, Accra, Ghana. (2)Institute for Stem Cell Research and Regenerative Medicine, Medical faculty, Heinrich-Heine University, Düsseldorf, Germany. (3)West African Center for Cell Biology of Infectious Pathogens, Department of Biochemistry, Cell and Molecular Biology, University of Ghana, Accra, Ghana. PArthur@ug.edu.gh. (4)West African Center for Cell Biology of Infectious Pathogens, Department of Biochemistry, Cell and Molecular Biology, University of Ghana, Accra, Ghana. James.Adjaye@med.uni-duesseldorf.de. (5)Institute for Stem Cell Research and Regenerative Medicine, Medical faculty, Heinrich-Heine University, Düsseldorf, Germany. James.Adjaye@med.uni-duesseldorf.de.

BACKGROUND: Marine endophytic fungi (MEF) are good sources of structurally unique and biologically active secondary metabolites. Due to the increase in antimicrobial resistance, the secondary metabolites from MEF ought to be fully explored to identify candidates which could serve as lead compounds for novel drug development. These secondary metabolites might also be useful for development of new cancer drugs. In this study, ethyl acetate extracts from marine endophytic fungal cultures were tested for their antifungal activity and anticancer properties against C. albicans and the human liver cancer cell line HepG2, respectively. The highly enriched fractions were also analyzed by high performance liquid chromatography coupled with high resolution mass spectrometry (HPLC-HRMS) and their effect on the HepG2 cells was assessed via transcriptomics and with a proliferation assay. RESULTS: We demonstrated that the fractions could reduce proliferation in HepG2 cells. The detailed transcriptome analysis revealed regulation of several cancer- and metabolism-related pathways and gene ontologies. The down-regulated pathways included, cell cycle, p53 signaling, DNA replication, sphingolipid

metabolism and drug metabolism by cytochrome P450. The upregulated pathways included HIF-1 signaling, focal adhesion, necroptosis and transcriptional mis-regulation of cancer. Furthermore, a protein interaction network was constructed based on the 26 proteins distinguishing the three treatment conditions from the untreated cells. This network was composed of central functional components associated with metabolism and cancer such as TNF, MAPK, TRIM21 and one component contained APP. CONCLUSIONS: The purified fractions from MEF investigated in this study showed antifungal activity against C. albicans and S. cerevisiae alone or both and reduced proliferation of the human liver cancer cell line HepG2 implicating regulation of several cancer- and metabolism-related pathways. The data from this study could be instrumental in identifying new pathways associated with liver cancer anti-proliferative processes which can be used for the development of novel antifungal and anti-cancer drugs.

DOI: 10.1186/s12864-020-6684-z PMCID: PMC7106652 PMID: 32228434

A.3. Reproduction. 2019 Sep;158(3):R97-R111. doi: 10.1530/REP-18-0083.

The quest for pluripotency: a comparative analysis across mammalian species.

Devika AS(1), Wruck W(2), Adjaye J(2), Sudheer S(1).

Author information: (1)Department of Genomic Science, Central University of Kerala, Kerala, India. (2)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany.

Pluripotency is the developmental potential of a cell to give rise to all the cells in the three embryonic germ layers, including germline cells. Pluripotent stem cells (PSCs) can be embryonic, germ cell or somatic cell in origin and can adopt alternative states of pluripotency: naïve or primed. Although several reports have described the differentiation of PSCs to extra-embryonic lineages, such as primitive endoderm and trophectoderm, this is still debated among scientists in the field. In this review, we integrate the recent findings on pluripotency among mammals, alternative states of pluripotency, signalling pathways associated with maintaining pluripotency and the nature of PSCs derived from various mammals. PSCs from humans and mouse have been the most extensively studied. In other mammalian species, more research is required for understanding the optimum in vitro conditions required for either achieving pluripotency or preservation of distinct pluripotent states. A comparative high-throughput analysis of PSCs of genes expressed in naïve or primed states of humans, nonhuman primates (NHP) and rodents, based on publicly available datasets revealed the probable prominence of seven signalling pathways common among these species, irrespective of the states of pluripotency. We conclude by highlighting some of the unresolved questions and future directions of research on pluripotency in mammals.

DOI: 10.1530/REP-18-0083 PMID: 31035255

A.4. Brief Bioinform. 2014 Jan;15(1):65-78. doi: 10.1093/bib/bbs064. Epub 2012 Oct 9.

Data management strategies for multinational large-scale systems biology projects.

Wruck W(1), Peuker M, Regenbrecht CR.

Author information: (1)Institute of Pathology, Charite - Universitaetsmedizin Berlin, Chariteplatz 1, 10117 Berlin. Tel.: +49 30 2093 8951; Fax: +49 30 450 536 909; wasco.wruck@charite.de.

Good accessibility of publicly funded research data is essential to secure an open scientific system and eventually becomes mandatory [Wellcome Trust will Penalise Scientists Who Don't Embrace Open Access. The Guardian 2012]. By the use of high-throughput methods in many research areas from physics to systems biology, large data collections are increasingly important as raw material for research. Here, we present strategies worked out by international and national institutions targeting open access to publicly funded research data via incentives or obligations to share data. Funding organizations such as the British Wellcome Trust therefore have developed data sharing policies and request commitment to data management and sharing in grant applications. Increased citation rates are a profound argument for sharing publication data. Pre-publication sharing might be rewarded by a data citation credit system via digital object identifiers (DOIs) which have initially been in use for data objects. Besides policies and incentives, good practice in data management is indispensable. However, appropriate systems for data management of large-scale projects for example in systems biology are hard to find. Here, we give an overview of a selection of open-source data management systems proved to be employed successfully in large-scale projects.

DOI: 10.1093/bib/bbs064 PMCID: PMC3896927 PMID: 23047157 [Indexed for MEDLINE]

A.5. Bioinformatics. 2002 May;18(5):757-60.

Xdigitise: visualization of hybridization experiments.

Wruck W(1), Griffiths H, Steinfath M, Lehrach H, Radelof U, O'Brien J.

Author information: (1)Max-Planck-Institute for Molecular Genetics Berlin-Dahlem, Ihnestrasse 73, 14195 Berlin, Germany. wruck_w@molgen.mpg.de

Xdigitise is a software system for visualization of hybridization experiments giving the user facilities to analyze the corresponding images manually or automatically. Images of the high-density DNA arrays are displayed as well as the results of an external image analysis bundled with Xdigitise, e.g. the spot locations are marked and the duplicate correlations are shown by a color scale.AVAILABILITY: Xdigitise can be downloaded from http://www.molgen.mpg.de/~xdigitise.

DOI: 10.1093/bioinformatics/18.5.757 PMID: 12050072 [Indexed for MEDLINE]

A.6 Biol Open. 2020 Dec 28. pii: bio.054189. doi: 10.1242/bio.054189. [Epub ahead of print]

A stem cell based in vitro model of NAFLD enables the analysis of patient specific individual metabolic adaptations in response to a high fat diet and AdipoRon interference.

Graffmann N(1), Ncube A(1), Martins S(1), Fiszl AR(1), Reuther P(2), Bohndorf M(1), Wruck W(1), Beller M(3)(4), Czekelius C(2), Adjaye J(5). Author information: (1)Insti-

tute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University Düsseldorf, Medical faculty, Moorenstrasse 5, 40225 Düsseldorf, Germany. (2)Institute of Organic Chemistry and Macromolecular Chemistry, Heinrich-Heine University, Düsseldorf, Germany. (3)Institute for Mathematical Modeling of Biological Systems, Heinrich-Heine University, Düsseldorf, Germany. (4)Systems Biology of Lipid Metabolism, Heinrich-Heine University, Düsseldorf, Germany. (5)Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University Düsseldorf, Medical faculty, Moorenstrasse 5, 40225 Düsseldorf, Germany nina.graffmann@med.uni-duesseldorf.de.

Non-alcoholic fatty liver disease (NAFLD) is a multifactorial disease. Its development and progression depend on genetically predisposed susceptibility of the patient towards several hits which induce fat storage first and later inflammation and fibrosis. Here, we differentiated induced pluripotent stem cells (iPSCs) derived from four distinct donors with varying disease stages into hepatocyte like cells (HLCs) and determined fat storage as well as metabolic adaptations after stimulations with oleic acid. We could recapitulate the complex networks that control lipid and glucose metabolism and we identified distinct gene expression profiles related to the steatosis phenotype of the donor. In an attempt to reverse the steatotic phenotype, cells were treated with the small molecule AdipoRon, a synthetic analogue of adiponectin. Although the responses varied between cells lines, they suggest a general influence of AdipoRon on metabolism, transport, immune system, cell stress and signalling.

DOI: 10.1242/bio.054189 PMID: 33372064

A.7. Sci Rep. 2020 Feb 24;10(1):3284. doi: 10.1038/s41598-020-60065-8.

Functional omics analyses reveal only minor effects of microRNAs on human somatic stem cell differentiation.

Schira-Heinen J(1)(2), Czapla A(3), Hendricks M(2), Kloetgen A(4)(5), Wruck W(6), Adjaye J(6), Kögler G(3), Werner Müller H(2)(7), Stühler K(8)(9), Trompeter HI(3).

Author information: (1)Institute for Molecular Medicine, University Hospital Düsseldorf, Moorenstr. 5, Düsseldorf, Germany. (2)Molecular Neurobiology Laboratory, Department of Neurology, University Hospital Düsseldorf, Moorenstr. 5, 40225, Düsseldorf, Germany. (3)Institute for Transplantation Diagnostics and Cell Therapeutics, University Hospital Düsseldorf, Moorenstr. 5, Düsseldorf, Germany. (4)Department of Pediatric Oncology, Hematology and Clinical Immunology, University Hospital Düsseldorf, Moorenstr. 5, Düsseldorf, Germany. (5)Computational Biology of Infection Research, Helmholtz Center for Infection Research, Inhoffenstr. 7, 38124, Braunschweig, Germany. (6)Institute for Stem Cell Research and Regenerative Medicine, University Hospital Düsseldorf, Moorenstr. 5, Düsseldorf, Germany. (7)Biologisch-Medizinisches Forschungszentrum (BMFZ), Heinrich Heine University, Universitätsstr. 1, 40225, Düsseldorf, Germany. (8)Institute for Molecular Medicine, University Hospital Düsseldorf, Moorenstr. 5, Düsseldorf, Germany. Kai.Stuehler@uni-duesseldorf.de. (9)Biologisch-Medizinisches Forschungszentrum (BMFZ), Heinrich Heine University, Universitätsstr. 1, 40225, Düsseldorf, Germany. Kai.Stuehler@uni-duesseldorf.de.

The contribution of microRNA-mediated posttranscriptional regulation on the final proteome in differentiating cells remains elusive. Here, we evaluated the impact of microRNAs (miRNAs) on the proteome of human umbilical cord blood-derived unrestricted somatic stem cells (USSC) during retinoic acid (RA) differentiation by a systemic approach using next generation sequencing analysing mRNA and miRNA expression and quantitative mass spectrometry-based proteome analyses. Interestingly, regulation of mRNAs and their dedicated proteins highly correlated during RA-incubation. Additionally, RA-induced USSC demonstrated a clear separation from native USSC thereby shifting from a proliferating to a metabolic phenotype. Bioinformatic integration of up- and downregulated miRNAs and proteins initially implied a strong impact of the miRNome on the XXL-USSC proteome. However, quantitative proteome analysis of the miRNA contribution on the final proteome after ectopic overexpression of downregulated miR-27a-5p and miR-221-5p or inhibition of upregulated miR-34a-5p, respectively, followed by RA-induction revealed only minor proportions of differentially abundant proteins. In addition, only small overlaps of these regulated proteins with inversely abundant proteins in non-transfected RA-treated USSC were observed. Hence, mRNA transcription rather than miRNA-mediated regulation is the driving force for protein regulation upon

RA-incubation, strongly suggesting that miRNAs are fine-tuning regulators rather than active primary switches during RA-induction of USSC.

DOI: 10.1038/s41598-020-60065-8 PMCID: PMC7040006 PMID: 32094412

A.8. Sci Rep. 2019 Nov 22;9(1):17365. doi: 10.1038/s41598-019-53907-7.

FGF Signalling in the Self-Renewal of Colon Cancer Organoids.

Otte J(1), Dizdar L(2), Behrens B(2), Goering W(3), Knoefel WT(2), Wruck W(1), Stoecklein NH(2), Adjaye J(4).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, University Hospital and Medical Faculty of the Heinrich-Heine University Düsseldorf, Düsseldorf, Germany. (2)General, Visceral and Paediatric Surgery, University Hospital and Medical Faculty of the Heinrich-Heine University Düsseldorf, Düsseldorf, Germany. (3)Institute for Pathology, University Hospital and Medical Faculty of the Heinrich-Heine University Düsseldorf, Düsseldorf, Germany. (4)Institute for Stem Cell Research and Regenerative Medicine, University Hospital and Medical Faculty of the Heinrich-Heine University Düsseldorf, Düsseldorf, Germany. James.Adjaye@med.uni-duesseldorf.de.

The progression of colorectal cancer (CRC) is supposedly driven by cancer stem cells (CSC) which are able to self-renew and simultaneously fuel bulk tumour mass with highly proliferative and differentiated tumour cells. However, the CSC-phenotype in CRC is unstable and dependent on environmental cues. Fibroblast growth factor 2 (FGF2) is essential and necessary for the maintenance of self-renewal in adult and embryonic stem cells. Investigating its role in self-renewal in advanced CRC patient-derived organoids, we unveiled that FGF-receptor (FGFR) inhibition prevents organoid formation in very early expanding cells but induces cyst formation when applied to pre-established organoids. Comprehensive transcriptome analyses revealed that the induction of the transcription factor activator-protein-1 (AP-1) together with MAPK activation was most prominent after FGFR-inhibition. These effects resemble mechanisms of an acquired resistance against other described tyrosine kinase inhibitors such as EGF-receptor targeted therapies. Furthermore, we detected elevated expression levels of several self-renewal and stemness-associated genes in organoid cultures with active FGF2 signalling. The combined data assume that CSCs are a heterogeneous population while self-renewal is a common feature regulated by distinct but converging pathways. Finally, we highlight FGF2 signalling as one of numerous components of the complex regulation of stemness in cancer.

DOI: 10.1038/s41598-019-53907-7 PMCID: PMC6874569 PMID: 31758153

A.9. Bone Res. 2019 Oct 24;7:32. doi: 10.1038/s41413-019-0069-4. eCollection 2019.

Human iPSC-derived iMSCs improve bone regeneration in mini-pigs.

Jungbluth P(#)(1), Spitzhorn LS(#)(2), Grassmann J(1), Tanner S(1), Latz D(1), Rahman MS(2), Bohndorf M(2), Wruck W(2), Sager M(3), Grotheer V(1), Kröpil P(4), Hakimi M(5), Windolf J(1), Schneppendahl J(1), Adjaye J(2).

Author information: (1)1Department of Trauma and Hand Surgery, Heinrich Heine University Hospital Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. (2)2Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. (3)3Animal Research Institute, Heinrich Heine University, Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. (4)4Department of Diagnostic and Interventional Radiology, Medical Faculty, Heinrich Heine University, Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. (5)5Vivantes Klinikum Am Urban, Dieffenbachstraße 1, 10967 Berlin, Germany. (#)Contributed equally

Autologous bone marrow concentrate (BMC) and mesenchymal stem cells (MSCs) have beneficial effects on the healing of bone defects. To address the shortcomings associated with the use of primary MSCs, induced pluripotent stem cell (iPSC)-derived MSCs (iMSCs) have been proposed as an alternative. The aim of this study was to investigate the bone regeneration potential of human iMSCs combined with calcium phosphate granules (CPG) in critical-size defects in the proximal tibias of mini-pigs in the early phase of bone healing compared to that of a previously reported autograft treatment and treatment with a composite made of either a combination of autologous BMC and CPG or CPG alone. iMSCs were derived from iPSCs originating from human fetal foreskin fibroblasts (HFFs). They were able to differentiate into osteoblasts in vitro, express a plethora of bone morphogenic proteins (BMPs) and secrete paracrine signaling-associated cytokines such as PDGF-AA and osteopontin. Radiologically and histomorphometrically, HFF-iMSC + CPG transplantation resulted in significantly better osseous consolidation than the transplantation of CPG alone and produced no significantly different outcomes compared to the transplantation of autologous BMC + CPG after 6 weeks. The results of this translational study imply that iMSCs represent a valuable future treatment option for load-bearing bone defects in humans.

DOI: 10.1038/s41413-019-0069-4 PMCID: PMC6813363 PMID: 31667001

Conflict of interest statement: Competing interestsThe authors declare no competing interests.

A.10. Stem Cell Res Ther. 2019 Mar 18;10(1):100. doi: 10.1186/s13287-019-1209-x.

Human iPSC-derived MSCs (iMSCs) from aged individuals acquire a rejuvenation signature.

Spitzhorn LS(1), Megges M(1), Wruck W(1), Rahman MS(1), Otte J(1), Degistirici Ö(2), Meisel R(2), Sorg RV(3), Oreffo ROC(4), Adjaye J(5).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Düsseldorf, Moorenstr. 5, 40225, Düsseldorf, Germany. (2)Division of Paediatric Stem Cell Therapy, Clinic for Pediatric Oncology, Hematology and Clinical Immunology, Medical Faculty, Heinrich Heine University, Düsseldorf, Moorenstr. 5, 40225, Düsseldorf, Germany. (3)Institute for Transplantation Diagnostics and Cell Therapeutics, Heinrich Heine University Hospital, Moorenstr, 5, 40225, Düsseldorf, Germany. (4)Bone and Joint Research Group, Centre for Human Development, Stem Cells and Regeneration, Institute of Developmental Sciences, University of Southampton, Southampton, SO16 6YD, UK. (5)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Düsseldorf, Moorenstr. 5, 40225, Düsseldorf, Germany. james.adjaye@med.uni-duesseldorf.de.

BACKGROUND: Primary mesenchymal stem cells (MSCs) are fraught with aging-related shortfalls. Human-induced pluripotent stem cell (iPSC)-derived MSCs (iMSCs) have been shown to be a useful clinically relevant source of MSCs that circumvent these aging-associated drawbacks. To date, the extent of the retention of aging-hallmarks in iMSCs differentiated from iPSCs derived from elderly donors remains unclear. METHODS: Fetal femur-derived MSCs (fMSCs) and adult bone marrow MSCs (aMSCs) were isolated, corresponding iPSCs were generated, and iMSCs were differentiated from fMSC-iPSCs, from aMSC-iPSCs, and from human embryonic stem cells (ESCs) H1. In addition, typical MSC characterization such as cell surface marker expression, differentiation capacity, secretome profile, and trancriptome analysis were conducted for the three distinct iMSC preparations-fMSC-iMSCs, aMSC-iMSCs, and ESC-iMSCs. To verify these results, previously published data sets were used, and also, additional aMSCs and iMSCs were analyzed. RESULTS: fMSCs and aMSCs both express the typical MSC cell surface markers and can be differentiated into osteogenic, adipogenic, and chondrogenic lineages in vitro. However, the transcriptome analysis revealed overlapping and distinct gene expression patterns and showed that fMSCs express more genes in common with ESCs than with aMSCs. fMSC-iMSCs, aMSC-iMSCs, and ESC-iMSCs met the criteria set out for MSCs. Dendrogram analyses confirmed that the transcriptomes of all iMSCs clustered together with the parental MSCs and separated from the MSC-iPSCs and ESCs. iMSCs irrespective of donor age and cell type acquired a rejuvenation-associated gene signature, specifically, the expression of INHBE, DNMT3B, POU5F1P1, CDKN1C, and GCNT2 which are also expressed in pluripotent stem cells (iPSCs and ESC) but not in the parental aMSCs. iMSCs expressed more genes in common with fMSCs than with aMSCs. Independent real-time PCR comparing aMSCs, fMSCs, and iMSCs confirmed the differential expression of the rejuvenation (COX7A, EZA2, and TMEM119) and aging (CXADR and IGSF3) signatures. Importantly, in terms of regenerative medicine, iMSCs acquired a secretome (e.g., angiogenin, DKK-1, IL-8, PDGF-AA, osteopontin, SERPINE1, and VEGF) similar to that of fMSCs and aMSCs, thus highlighting their

ability to act via paracrine signaling. CONCLUSIONS: iMSCs irrespective of donor age and cell source acquire a rejuvenation gene signature. The iMSC concept could allow circumventing the drawbacks associated with the use of adult MSCs und thus provide a promising tool for use in various clinical settings in the future.

DOI: 10.1186/s13287-019-1209-x PMCID: PMC6423778 PMID: 30885246

A.11. Stem Cell Res. 2019 Jan;34:101372. doi: 10.1016/j.scr.2018.101372. Epub 2018 Dec 27.

Fibroblast-derived integration-free iPSC line ISRM-NBS1 from an 18-year-old Nijmegen Breakage Syndrome patient carrying the homozygous NBN c.657_661del5 mutation.

Martins S(1), Bohndorf M(1), Graffmann N(1), Wruck W(1), Chrzanowska KH(2), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. (2)Department of Medical Genetics, Children's Memorial Health Institute, Warsaw, Poland. (3)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. Electronic address: james.adjaye@med.uni-duesseldorf.de.

Human fibroblasts cells from a female diagnosed with Nijmegen Breakage Syndrome (NBS) carrying the homozygous NBN c.657_661del5 mutation were used to generate integration-free induced pluripotent stem cells (iPSCs) by over-expressing episomal-based plasmids harbouring OCT4, SOX2, NANOG, KLF4, c-MYC and LIN28. The derived iPSC line - ISRM-NBS1 was defined as pluripotent based on (i) expression of pluripotency-associated markers (ii) embryoid body-based differentiation into cell types representative of the three germ layers and (iii) the similarity between the transcriptome of the iPSC line and the human embryonic stem cell line H1 with a Pearson correlation of 0.955.

DOI: 10.1016/j.scr.2018.101372 PMID: 30616142 [Indexed for MEDLINE]

A.12. Stem Cells Dev. 2018 Nov 20. doi: 10.1089/scd.2018.0010. [Epub ahead of print]

Transplanted Human Pluripotent Stem Cell-Derived Mesenchymal Stem Cells Support Liver Regeneration in Gunn Rats.

Spitzhorn LS(1), Kordes C(2), Megges M(1), Sawitza I(2), Götze S(2), Reichert D(2), Schulze-Matz P(1), Graffmann N(1), Bohndorf M(1), Wruck W(1), Köhler JP(2), Herebian D(3), Mayatepek E(3), Oreffo ROC(4), Häussinger D(2), Adjaye J(1).

Author information: (1)1 Institute for Stem Cell Research and Regenerative Medicine, Neonatolgy and Pediatric Cardiology, Heinrich Heine University , Düsseldorf, Germany . (2)2 Clinic of Gastroenterology, Hepatology and Infectious Diseases, and Neonatolgy and Pediatric Cardiology, Heinrich Heine University , Düsseldorf, Germany . (3)3 Department of General Pediatrics, Neonatolgy and Pediatric Cardiology, Heinrich Heine University , Düsseldorf, Germany . (4)4 Centre for Human Development, Stem Cells and Regeneration, Faculty of Medicine, University of Southampton , Southampton, United Kingdom .

Gunn rats bear a mutation within the uridine diphosphate glucuronosyltransferase-1a1 (Ugt1a1) gene resulting in high serum bilirubin levels as seen in Crigler-Najjar syndrome. In this study, the Gunn rat was used as an animal model for heritable liver dysfunction. Induced mesenchymal stem cells (iMSCs) derived from embryonic stem cells (H1) and induced pluripotent stem cells were transplanted into Gunn rats after partial hepatectomy. The iMSCs engrafted and survived in the liver for up to 2 months. The transplanted iMSCs differentiated into functional hepatocytes as evidenced by partially suppressed hyperbilirubinemia and expression of multiple human-specific hepatocyte markers such as albumin, hepatocyte nuclear factor 4α, UGT1A1, cytokeratin 18, bile salt export pump, multidrug resistance protein 2, Na/taurocholate-cotransporting polypeptide, and α-fetoprotein. These findings imply that transplanted human iMSCs can contribute to liver regeneration in vivo and thus represent a promising tool for the treatment of inherited liver diseases.

DOI: 10.1089/scd.2018.0010 PMID: 30280963

A.13. Stem Cell Res. 2018 Aug;31:131-134. doi: 10.1016/j.scr.2018.07.011. Epub 2018 Jul 27.

Establishment and characterization of an iPSC line from a 58 years old high grade patient with nonalcoholic fatty liver disease (70% steatosis) with homozygous wildtype PNPLA3 genotype.

Graffmann N(1), Bohndorf M(1), Ncube A(1), Wruck W(1), Kashofer K(2), Zatloukal K(2), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany. (2)Institute of

Pathology, Medical University of Graz, 8036 Graz, Austria. (3)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany. Electronic address: james.adjaye@med.uni-duesseldorf.de.

Nonalcoholic fatty liver disease (NAFLD) is the hepatic manifestation of the metabolic syndrome and its prevalence increases continuously. Here, we reprogrammed fibroblasts of a high grade NAFLD patient with homozygous wildtype PNPLA3 genotype. We characterized the induced pluripotent stem cells (iPSCs) by immunocytochemistry, flow cytometry, embryoid body formation, pluritest DNA-fingerprinting, and karyotype analysis.

DOI: 10.1016/j.scr.2018.07.011 PMID: 30081348 [Indexed for MEDLINE]

A.14. Stem Cell Res. 2018 Aug;31:113-116. doi: 10.1016/j.scr.2018.07.015. Epub 2018 Jul 25.

Establishment and characterization of an iPSC line from a 35 years old high grade patient with nonalcoholic fatty liver disease (30-40% steatosis) with homozygous wildtype PNPLA3 genotype.

Graffmann N(1), Bohndorf M(1), Ncube A(1), Kawala MA(1), Wruck W(1), Kashofer K(2), Zatloukal K(2), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, 40225 Düsseldorf, Germany. (2)Institute of Pathology, Medical University of Graz, 8036 Graz, Austria. (3)Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, 40225 Düsseldorf, Germany. Electronic address: james.adjaye@med.uni-duesseldorf.de.

Nonalcoholic fatty liver disease (NAFLD) is the hepatic manifestation of the metabolic syndrome and its prevalence increases continuously. Here, we reprogrammed fibroblasts of a high grade NAFLD patient with homozygous wildtype PNPLA3 genotype. The induced pluripotent stem cells (iPSCs) were characterized by immunocytochemistry, flow cytometry, embryoid body formation, pluritest, DNA-fingerprinting and karyotype analysis.

DOI: 10.1016/j.scr.2018.07.015 PMID: 30071394 [Indexed for MEDLINE]

A.15. PLoS One. 2018 Jul 10;13(7):e0200416. doi: 10.1371/journal.pone.0200416. eCollection 2018.

Cell fate decisions of human iPSC-derived bipotential hepatoblasts depend on cell density.

Graffmann N(1), Ncube A(1), Wruck W(1), Adjaye J(1).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical faculty, Heinrich-Heine University, Düsseldorf, Germany.

During embryonic development bipotential hepatoblasts differentiate into hepatocytes and cholangiocytes- the two main cell types within the liver. Cell fate decision depends on elaborate interactions between distinct signalling pathways, namely Notch, WNT, TGFβ, and Hedgehog. Several in vitro protocols have been established to differentiate human pluripotent stem cells into either hepatocyte or cholangiocyte like cells (HLC/CLC) to enable disease modelling or drug screening. During HLC differentiation we observed the occurrence of epithelial cells with a phenotype divergent from the typical hepatic polygonal shape- we refer to these as endoderm derived epithelial cells (EDECs). These cells do not express the mature hepatocyte marker ALB or the progenitor marker AFP. However they express the cholangiocyte markers SOX9, OPN, CFTR as well as HNF4α, CK18 and CK19. Interestingly, they express both E Cadherin and Vimentin, two markers that are mutually exclusive, except for cancer cells. EDECs grow spontaneously under low density cell culture conditions and their occurrence was unaffected by interfering with the above mentioned signalling pathways.

DOI: 10.1371/journal.pone.0200416 PMCID: PMC6039024 PMID: 29990377 [Indexed for MEDLINE]

Conflict of interest statement: The authors have declared that no competing interests exist.

A.16. Stem Cell Res. 2018 Jul;30:141-144. doi: 10.1016/j.scr.2018.05.018. Epub 2018 Jun 1.

Lymphoblast-derived integration-free iPSC line AD-TREM2-3 from a 74 year-old Alzheimer's disease patient expressing the TREM2 p.R47H variant.

Martins S(1), Yigit H(1), Bohndorf M(1), Graffmann N(1), Fiszl AR(1), Wruck W(1), Sleegers K(2), Van Broeckhoven C(2), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf,

Germany. (2)Neurodegenerative Brain Disease Groups, Department of Molecular Genetics, VIB, Antwerp, Belgium; Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Belgium. (3)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. Electronic address: james.adjaye@med.uni-duesseldorf.de.

Human lymphoblast cells from a male diagnosed with Alzheimer's disease (AD) expressing the TREM2 p.R47H variant were used to generate integration-free induced pluripotent stem cells (iPSCs) by over-expressing episomal-based plasmids harbouring OCT4, SOX2, KLF4, LIN28, L-MYC and p53 shRNA. The derived iPSC line - AD-TREM2-3 was defined as pluripotent based on (i) expression of pluripotency-associated markers (ii) embryoid body-based differentiation into cell types representative of the three germ layers and (iii) the similarity between the transcriptome of the iPSC line and the human embryonic stem cell line H1 with a Pearson correlation of 0.940.

DOI: 10.1016/j.scr.2018.05.018 PMID: 29902745 [Indexed for MEDLINE]

A.17. Stem Cell Res Ther. 2018 Apr 25;9(1):113. doi: 10.1186/s13287-018-0864-7.

The presence of human mesenchymal stem cells of renal origin in amniotic fluid increases with gestational time.

Rahman MS(1), Spitzhorn LS(1), Wruck W(1), Hagenbeck C(2), Balan P(2), Graffmann N(1), Bohndorf M(1), Ncube A(1), Guillot PV(3), Fehm T(2), Adjaye J(4).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Moorenstraße 5, 40225, Düsseldorf, Germany. (2)Department of Obstetrics and Gynaecology, Medical Faculty, Heinrich Heine University Düsseldorf, Moorenstraße 5, 40225, Düsseldorf, Germany. (3)Institute for Women's Health, Maternal and Fetal Medicine Department, University College London, London, WC1E 6HX, UK. (4)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Moorenstraße 5, 40225, Düsseldorf, Germany. james.adjaye@med.uni-duesseldorf.de.

BACKGROUND: Established therapies for managing kidney dysfunction such as kidney dialysis and transplantation are limited due to the shortage of compatible donated organs and high costs. Stem cell-based therapies are currently under investigation as an alternative treatment option. As amniotic fluid is composed of fetal urine harboring mesenchymal stem cells (AF-MSCs), we hypothesized that third-trimester amniotic fluid could be a novel source of renal progenitor and differentiated cells. METHODS: Human third-trimester amniotic fluid cells (AFCs) were isolated and cultured in distinct media. These cells were characterized as renal progenitor cells with respect to

cell morphology, cell surface marker expression, transcriptome and differentiation into chondrocytes, osteoblasts and adipocytes. To test for renal function, a comparative albumin endocytosis assay was performed using AF-MSCs and commercially available renal cells derived from kidney biopsies. Comparative transcriptome analyses of first, second and third trimester-derived AF-MSCs were conducted to monitor expression of renal-related genes. RESULTS: Regardless of the media used, AFCs showed expression of pluripotency-associated markers such as SSEA4, TRA-1-60, TRA-1-81 and C-Kit. They also express the mesenchymal marker Vimentin. Immunophenotyping confirmed that third-trimester AFCs are bona fide MSCs. AF-MSCs expressed the master renal progenitor markers SIX2 and CITED1, in addition to typical renal proteins such as PODXL, LHX1, BRN1 and PAX8. Albumin endocytosis assays demonstrated the functionality of AF-MSCs as renal cells. Additionally, upregulated expression of BMP7 and downregulation of WT1, CD133, SIX2 and C-Kit were observed upon activation of WNT signaling by treatment with the GSK-3 inhibitor CHIR99201. Transcriptome analysis and semiquantitative PCR revealed increasing expression levels of renal-specific genes (e.g., SALL1, HNF4B, SIX2) with gestational time. Moreover, AF-MSCs shared more genes with human kidney cells than with native MSCs and gene ontology terms revealed involvement of biological processes associated with kidney morphogenesis. CONCLUSIONS: Third-trimester amniotic fluid contains AF-MSCs of renal origin and this novel source of kidney progenitors may have enormous future potentials for disease modeling, renal repair and drug screening.

DOI: 10.1186/s13287-018-0864-7 PMCID: PMC5918774 PMID: 29695308 [Indexed for MEDLINE]

A.18. Stem Cell Res. 2018 May;29:60-63. doi: 10.1016/j.scr.2018.03.011. Epub 2018 Mar 20.

Lymphoblast-derived integration-free iPSC line AD-TREM2-1 from a 67year-old Alzheimer's disease patient expressing the TREM2 p.R47H variant.

Martins S(1), Yigit H(1), Bohndorf M(1), Graffmann N(1), Fiszl AR(1), Wruck W(1), Sleegers K(2), Van Broeckhoven C(2), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine (ISRM), Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. (2)Neurodegenerative Brain Disease Groups, Department of Molecular Genetics, VIB, Antwerp, Belgium; Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Belgium. (3)Institute for Stem Cell Research and Regenerative Medicine (ISRM), Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. Electronic address: james.adjaye@med.uni-duesseldorf.de.

Human lymphoblast cells from a male diagnosed with Alzheimer's disease (AD) expressing the TREM2 p.R47H variant were used to generate integration-free induced pluripotent stem cells (iPSCs) by over-expressing episomal-based plasmids harbouring OCT4, SOX2, NANOG, LIN28, c-MYC and L-MYC. AD-TREM2-1 was defined as pluripotent based on (i) expression of pluripotency-associated markers (ii) embryoid body-based differentiation into cell types representative of the three germ layers and (iii) the similarity between the transcriptome of the iPSC line and the human embryonic stem cell line H1 with a Pearson correlation of 0.947.

DOI: 10.1016/j.scr.2018.03.011 PMID: 29602048 [Indexed for MEDLINE]

A.19. Stem Cell Res. 2018 Jan;26:76-79. doi: 10.1016/j.scr.2017.12.007. Epub 2017 Dec 13.

Lymphoblast-derived integration-free ISRM-CON9 iPS cell line from a 75year old female.

Martins S(1), Bohndorf M(1), Schröter F(1), Assar F(1), Wruck W(1), Sleegers K(2), Van Broeckhoven C(2), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstrasse 5, Düsseldorf 40225, Germany. (2)Neurodegenerative Brain Disease Groups, Department of Molecular Genetics, VIB, Antwerp, Belgium; Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Belgium. (3)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstrasse 5, Düsseldorf 40225, Germany. Electronic address: james.adjaye@med.uni-duesseldorf.de.

Human lymphoblast cells were used to generate integration-free induced pluripotent stem cells (iPSCs) employing episomal-based plasmids expressing OCT4, SOX2, NANOG, LIN28, c-MYC and L-MYC. The derived iPSCs were defined as pluripotent based on (i) expression of pluripotency-associated markers, (ii) embryoid body-based differentiation into cell types representative of the three germ layers and (iii) the similarity between the transcriptomes of the iPSC line and the human embryonic stem cell line H1 with a Pearson correlation of 0.95.

DOI: 10.1016/j.scr.2017.12.007 PMID: 29268155 [Indexed for MEDLINE]

A.20. Stem Cells Int. 2017;2017:5932706. doi: 10.1155/2017/5932706. Epub 2017 Oct 31.

Isolation and Molecular Characterization of Amniotic Fluid-Derived Mesenchymal Stem Cells Obtained from Caesarean Sections.

Spitzhorn LS(#)(1), Rahman MS(#)(1), Schwindt L(1), Ho HT(1), Wruck W(1), Bohndorf M(1), Wehrmeyer S(1), Ncube A(1), Beyer I(2), Hagenbeck C(2), Balan P(2), Fehm T(2), Adjaye J(1).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. (2)Department of Obstetrics and Gynaecology, Heinrich Heine University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. (#)Contributed equally

Human amniotic fluid cells are immune-privileged with low immunogenicity and anti-inflammatory properties. They are able to self-renew, are highly proliferative, and have a broad differentiation potential, making them amenable for cell-based therapies. Amniotic fluid (AF) is routinely obtained via amniocentesis and contains heterogeneous populations of foetal-derived progenitor cells including mesenchymal stem cells (MSCs). In this study, we isolated human MSCs from AF (AF-MSCs) obtained during Caesarean sections (C-sections) and characterized them. These AF-MSCs showed typical MSC characteristics such as morphology, in vitro differentiation potential, surface marker expression, and secreted factors. Besides vimentin and the stem cell marker CD133, subpopulations of AF-MSCs expressed pluripotency-associated markers such as SSEA4, c-Kit, TRA-1-60, and TRA-1-81. The secretome and related gene ontology (GO) terms underline their immune modulatory properties. Furthermore, transcriptome analyses revealed similarities with native foetal bone marrow-derived MSCs. Significant KEGG pathways as well as GO terms are mostly related to immune function, embryonic skeletal system, and TGFβ-signalling. An AF-MSC-enriched gene set included putative AF-MSC markers PSG5, EMX-2, and EVR-3. In essence, C-section-derived AF-MSCs can be routinely obtained and are amenable for personalized cell therapies and disease modelling.

DOI: 10.1155/2017/5932706 PMCID: PMC5684599 PMID: 29225627

A.21. Stem Cell Res. 2017 Dec;25:18-21. doi: 10.1016/j.scr.2017.10.004. Epub 2017 Oct 7.

Derivation and characterization of integration-free iPSC line ISRM-UM51 derived from SIX2-positive renal cells isolated from urine of an African male expressing the CYP2D6 *4/*17 variant which confers intermediate drug metabolizing activity.

Bohndorf M(1), Ncube A(1), Spitzhorn LS(1), Enczmann J(2), Wruck W(1), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany. (2)Institute of Transplantation Diagnostics and Cell Therapeutics, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany. (3)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany. Electronic address: james.adjaye@med.uni-duesseldorf.de.

SIX2-positive renal cells isolated from urine from a 51year old male of African origin bearing the CYP2D6 *4/*17 variant were reprogrammed by nucleofection of a combination of two episomal-based plasmids omitting pathway (TGFβ, MEK and GSK3β) inhibition. The induced pluripotent stem cells (iPSCs) were characterized by immunocytochemistry, embryoid body formation, DNA-fingerprinting and karyotype analysis. Comparative transcriptome analyses with human embryonic stem cell lines H1 and H9 revealed a Pearson correlation of 0.9243 and 0.9619 respectively.

DOI: 10.1016/j.scr.2017.10.004 PMID: 29035842 [Indexed for MEDLINE]

A.22. FEBS Lett. 2017 Aug;591(15):2226-2240. doi: 10.1002/1873-3468.12716. Epub 2017 Jul 2.

New insights into human primordial germ cells and early embryonic development from single-cell analysis.

Otte J(1), Wruck W(1), Adjaye J(1).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, Duesseldorf, Germany.

Human preimplantation developmental studies are difficult to accomplish due to associated ethical and moral issues. Preimplantation cells are rare and exist only in transient cell states. From a single cell, it is very challenging to analyse the origination of the heterogeneity and complexity inherent to the human body. However, recent advances in single-cell technology and data analysis have provided new insights into the process of early human development and germ cell specification. In this Review, we examine the latest single-cell datasets of human preimplantation embryos and germ cell development, compare them to bulk cell analyses, and interpret their biological implications.

DOI: 10.1002/1873-3468.12716 PMID: 28627120 [Indexed for MEDLINE]

A.23. Stem Cell Res. 2017 Apr;20:50-53. doi: 10.1016/j.scr.2017.02.007. Epub 2017 Feb 24.

Generation and characterization of two iPSC lines from human epicardium-derived cells.

Paulitschek C(1), Schulze-Matz P(2), Hesse J(1), Schmidt T(1), Wruck W(2), Adjaye J(2), Schrader J(3).

Author information: (1)Department of Molecular Cardiology, Medical Faculty, Heinrich Heine University, 40225 Duesseldorf, Germany. (2)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225 Duesseldorf, Germany. (3)Department of Molecular Cardiology, Medical Faculty, Heinrich Heine University, 40225 Duesseldorf, Germany. Electronic address: schrader@uni-duesseldorf.de.

Human epicardium-derived cells (EPDC) were reprogrammed to generate two iPSC lines, MCDU1i-EPDC and MCDU2i-EPDC, by nucleofection of episomal-based plasmids expressing the reprogramming factors OCT4, SOX2, KLF4, c-MYC, NANOG and LIN28. Pluripotency was confirmed in vitro by immunofluorescence analysis and embryoid body formation. The iPSC lines and the human embryonic stem cell line H1 show a Pearson correlation co-efficient of 0.951 (MCDU1i-EPDC) and 0.937 (MCDU2i-EPDC) as assessed by comparative transcriptome profiling.

DOI: 10.1016/j.scr.2017.02.007 PMID: 28395740 [Indexed for MEDLINE]

A.24. Mol Ther. 2017 Feb 1;25(2):427-442. doi: 10.1016/j.ymthe.2016.11.014.

Human Amniocytes Are Receptive to Chemically Induced Reprogramming to Pluripotency.

Hawkins KE(1), Moschidou D(1), Faccenda D(2), Wruck W(3), Martin-Trujillo A(4), Hau KL(5), Ranzoni AM(1), Sanchez-Freire V(6), Tommasini F(7), Eaton S(8), De Coppi P(8), Monk D(3), Campanella M(9), Thrasher AJ(8), Adjaye J(3), Guillot PV(10).

Author information: (1)Institute for Women's Health, Maternal and Fetal Medicine Department, University College London (UCL), London WC1E 6HX, UK. (2)Department of Comparative Biomedical Sciences, The Royal Veterinary College (RVC), Royal College Street, London NW1 0TU, UK. (3)Institute for Stem Cell Research and Regenerative

Medicine, Heinrich Heine University Dusseldorf, Dusseldorf 40225, Germany. (4)Imprinting and Cancer Group, Cancer Epigenetic and Biology Program, Bellvitge Institute for Biomedical Research (IDIBELL), Hospital Duran i Reynals, Barcelona 08908, Spain. (5)Institute for Women's Health, Maternal and Fetal Medicine Department, University College London (UCL), London WC1E 6HX, UK; Imperial College London, National Heart and Lung Institute, Hammersmith Campus, Du Cane Road, London W12 0NN, UK. (6)Departments of Medicine and Radiology, Stanford University, Stanford, CA 94305-5454, USA. (7)Institute for Women's Health, Maternal and Fetal Medicine Department, University College London (UCL), London WC1E 6HX, UK; Institute for Child Health, University College London, London WC1N 1EH, UK. (8)Institute for Child Health, University College London, London WC1N 1EH, UK. (9)Department of Comparative Biomedical Sciences, The Royal Veterinary College (RVC), Royal College Street, London NW1 0TU, UK; Consortium for Mitochondrial Research, University College London, Royal College Street, London NW1 0TU, UK. (10)Institute for Women's Health, Maternal and Fetal Medicine Department, University College London (UCL), London WC1E 6HX, UK. Electronic address: p.guillot@ucl.ac.uk.

Restoring pluripotency using chemical compounds alone would be a major step forward in developing clinical-grade pluripotent stem cells, but this has not yet been reported in human cells. We previously demonstrated that VPA_AFS cells, human amniocytes cultivated with valproic acid (VPA) acquired functional pluripotency while remaining distinct from human embryonic stem cells (hESCs), questioning the relationship between the modulation of cell fate and molecular regulation of the pluripotency network. Here, we used single-cell analysis and functional assays to reveal that VPA treatment resulted in a homogeneous population of self-renewing non-transformed cells that fulfill the hallmarks of pluripotency, i.e., a short G1 phase, a dependence on glycolytic metabolism, expression of epigenetic modifications on histones 3 and 4, and reactivation of endogenous OCT4 and downstream targets at a lower level than that observed in hESCs. Mechanistic insights into the process of VPA-induced reprogramming revealed that it was dependent on OCT4 promoter activation, which was achieved independently of the PI3K (phosphatidylinositol 3-kinase)/AKT/mTOR (mammalian target of rapamycin) pathway or GSK3β inhibition but was concomitant with the presence of acetylated histones H3K9 and H3K56, which promote pluripotency. Our data identify, for the first time, the pluripotent transcriptional and molecular signature and metabolic status of human chemically induced pluripotent stem cells.

DOI: 10.1016/j.ymthe.2016.11.014 PMCID: PMC5368475 PMID: 28153093 [Indexed for MEDLINE]

A.25. Stem Cell Res. 2016 Nov;17(3):597-599. doi: 10.1016/j.scr.2016.10.002. Epub 2016 Oct 19.

Characterization of iPSCs derived from dermal fibroblasts from a healthy 19year old female.

Kawala MA(1), Bohndorf M(1), Graffmann N(1), Wruck W(1), Zatloukal K(2), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany. (2)Institute of Pathology, Medical University of Graz, 8036 Graz, Austria. (3)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany. Electronic address: James.Adjaye@med.uni-duesseldorf.de.

Primary fibroblasts from a healthy 19years old female were reprogrammed by transduction of retroviruses OCT4, SOX2, c-MYC and KLF4. iPSCs were characterized by immunocytochemistry, embryonic body-formation, DNA-fingerprint and karyotype analysis and comparative transcriptome analyses with the human embryonic stem cell line H1 revealed a Pearsons correlation coefficient of 0.8952.

DOI: 10.1016/j.scr.2016.10.002 PMID: 27934589 [Indexed for MEDLINE]

A.26. Stem Cell Res. 2016 Nov;17(3):568-571. doi: 10.1016/j.scr.2016.10.007. Epub 2016 Oct 19.

Characterization of dermal fibroblast-derived iPSCs from a patient with high grade steatosis.

Kawala MA(1), Bohndorf M(1), Graffmann N(1), Wruck W(1), Zatloukal K(2), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany. (2)Institute of Pathology, Medical University of Graz, 8036 Graz, Austria. (3)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany. Electronic address: James.Adjaye@med.uni-duesseldorf.de.

Primary fibroblasts from a high grade steatosis patient were reprogrammed by transduction of retroviruses OCT4, SOX2, c-MYC and KLF4. IPSCs were characterized by immunocytochemistry, embryoid body-formation, DNA-fingerprint, karyotype analysis and comparative transcriptome analyses with the human embryonic stem cell line H1 revealed a Pearsons correlation coefficient of 0.9287. Resource table.

DOI: 10.1016/j.scr.2016.10.007 PMID: 27789412 [Indexed for MEDLINE]

A.27. Stem Cell Res. 2016 Nov;17(3):560-563. doi: 10.1016/j.scr.2016.10.003. Epub 2016 Oct 19.

Lymphoblast-derived integration-free iPSC lines from a female and male Alzheimer's disease patient expressing different copy numbers of a coding CNV in the Alzheimer risk gene CR1.

Schröter F(1), Sleegers K(2), Van Cauwenberghe C(2), Bohndorf M(1), Wruck W(1), Van Broeckhoven C(2), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. (2)Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, VIB, Antwerp, Belgium; Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Belgium. (3)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. Electronic address: James.Adjaye@med.uni-duesseldorf.de.

Human lymphoblast cells from a female and male patient diagnosed with Alzheimer's disease (AD) with different genotypes of a functional copy number variation (CNV) in the AD risk gene CR1 were used to generate integration-free induced pluripotent stem cells (iPSCs) employing episomal plasmids expressing OCT4, SOX2, NANOG, LIN28, c-MYC and L-MYC. The iPSCs retained the CR1 CNV, and comparative transcriptome analyses with the human embryonic stem cell line H1 revealed a Pearson correlation of 0.956 for AD1-CR10 and 0.908 for AD1-CR14.

DOI: 10.1016/j.scr.2016.10.003 PMID: 27789410 [Indexed for MEDLINE]

A.28. Stem Cell Res. 2016 Nov;17(3):553-555. doi: 10.1016/j.scr.2016.10.005. Epub 2016 Oct 20.

Lymphoblast-derived integration-free iPS cell line from a female 67-year-old Alzheimer's disease patient with TREM2 (R47H) missense mutation.

Schröter F(1), Sleegers K(2), Cuyvers E(2), Bohndorf M(1), Wruck W(1), Van Broeckhoven C(2), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. (2)Neurodegenerative Brain Disease Group, Department of Molecular

Genetics, VIB, Antwerp, Belgium; Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Belgium. (3)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. Electronic address: James.Adjaye@med.uni-duesseldorf.de.

Human lymphoblast cells from a female patient diagnosed with Alzheimer's disease (AD) possessing the missense mutation TREM2 p.R47H were used to generate integration-free induced pluripotent stem cells (iPSCs) employing episomal plasmids expressing OCT4, SOX2, NANOG, LIN28, c-MYC and L-MYC. The iPSCs retained the TREM2 mutation, and were defined as pluripotent based on (i) expression of pluripotent-associated markers, (ii) embryoid body-based differentiation into cell types representative of the three germ layers and (iii) the similarity between the transcriptomes of the iPSC line and the human embryonic stem cell line H1 with a Pearson correlation of 0.961.

DOI: 10.1016/j.scr.2016.10.005 PMID: 27789408 [Indexed for MEDLINE]

A.29. Stem Cell Res. 2016 Nov;17(3):547-549. doi: 10.1016/j.scr.2016.10.004. Epub 2016 Oct 20.

Characterization of dermal fibroblast-derived iPSCs from a patient with low grade steatosis.

Kawala MA(1), Bohndorf M(1), Graffmann N(1), Wruck W(1), Zatloukal K(2), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany. (2)Institute of Pathology, Medical University of Graz, 8036 Graz, Austria. (3)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich Heine University, 40225 Düsseldorf, Germany. Electronic address: James.Adjaye@med.uni-duesseldorf.de.

Primary fibroblasts from a low grade steatosis patient were reprogrammed by transduction of a combination of two episomal-based plasmids OCT4,SOX2, c-MYC and KLF4. iPSCs were characterized by immunocytochemistry, embryonic body-formation, DNA-fingerprint karyotype analysis and comparative transcriptome analyses with the human embryonic stem cell line H1 revealed a Pearsons correlation of 0.9251.

DOI: 10.1016/j.scr.2016.10.004 PMID: 27789406 [Indexed for MEDLINE]

A.30. Stem Cell Res. 2016 Jan;16(1):113-5. doi: 10.1016/j.scr.2015.12.017. Epub 2015 Dec 28.

Lymphoblast-derived integration-free iPS cell line from a 65-year-old Alzheimer's disease patient expressing the TREM2 p.R47H variant.

Schröter F(1), Sleegers K(2), Cuyvers E(2), Bohndorf M(1), Wruck W(1), Van Broeckhoven C(2), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstr, 5, 40225 Düsseldorf, Germany. (2)Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, VIB, Antwerp, Belgium; Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Belgium. (3)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstr, 5, 40225 Düsseldorf, Germany. Electronic address: James.Adjaye@med.uni-duesseldorf.de.

Human lymphoblast cells from a male patient diagnosed with Alzheimer's disease (AD) expressing the TREM2 p.R47H variant were used to generate integration-free induced pluripotent stem (iPS) cells employing episomal plasmids expressing OCT4, SOX2, NANOG, LIN28, c-MYC and L-MYC. The iPS cells retained the TREM2 mutation, and were defined as pluripotent based on (i) expression of pluripotent-associated markers, (ii) embryoid body-based differentiation into cell types representative of the three germ layers and (iii) the similarity between the transcriptomes of the iPS cell line and the human embryonic stem cell line H1 with a Pearson correlation of 0.966.

DOI: 10.1016/j.scr.2015.12.017 PMID: 27345793 [Indexed for MEDLINE]

A.31. Stem Cell Res. 2016 Jan;16(1):29-31. doi: 10.1016/j.scr.2015.11.016. Epub 2015 Dec 1.

Lymphoblast-derived integration-free iPS cell line from a 69-year-old male.

Schröter F(1), Sleegers K(2), Bohndorf M(1), Wruck W(1), Van Broeckhoven C(2), Adjaye J(1).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Medical Faculty, Heinrich-Heine-University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf,

Germany. (2)Neurodegenerative Brain Disease Groups, Department of Molecular Genetics, VIB, Antwerp, Belgium; Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Belgium.

Human lymphoblast cells were used to generate integration-free induced pluripotent stem (iPS) cells employing episomal plasmids expressing OCT4, SOX2, NANOG, LIN28, C-MYC and L-MYC. The derived iPS cells were defined as pluripotent based on (i) expression of pluripotent-associated markers, (ii) embryoid body-based differentiation into cell types representative of the three germ layers and (iii) the similarity between the transcriptomes of the iPS cell line and the human embryonic stem cell line H1 with a Pearson correlation of 0.95.

DOI: 10.1016/j.scr.2015.11.016 PMID: 27345781 [Indexed for MEDLINE]

A.32. Genom Data. 2016 Apr 26;8:131-3. doi: 10.1016/j.gdata.2016.04.014. eCollection 2016 Jun.

Combined sequencing of mRNA and DNA from human embryonic stem cells.

Mertes F(1), Kuhl H(2), Wruck W(3), Lehrach H(1), Adjaye J(3).

Author information: (1)Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany. (2)Next Generation Sequencing Group, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany. (3)Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, Moorenstr. 5, 40225 Düsseldorf, Germany.

Combined transcriptome and whole genome sequencing of the same ultra-low input sample down to single cells is a rapidly evolving approach for the analysis of rare cells. Besides stem cells, rare cells originating from tissues like tumor or biopsies, circulating tumor cells and cells from early embryonic development are under investigation. Herein we describe a universal method applicable for the analysis of minute amounts of sample material (150 to 200 cells) derived from sub-colony structures from human embryonic stem cells. The protocol comprises the combined isolation and separate amplification of poly(A) mRNA and whole genome DNA followed by next generation sequencing. Here we present a detailed description of the method developed and an overview of the results obtained for RNA and whole genome sequencing of human embryonic stem cells, sequencing data is available in the Gene Expression Omnibus (GEO) database under accession number GSE69471.

DOI: 10.1016/j.gdata.2016.04.014 PMCID: PMC4880790 PMID: 27275414

A.33. BMC Genomics. 2015 Nov 12;16:925. doi: 10.1186/s12864-015-2025-z.

Combined ultra-low input mRNA and whole-genome sequencing of human embryonic stem cells.

Mertes F(1)(2), Lichtner B(3), Kuhl H(4), Blattner M(5), Otte J(6), Wruck W(7), Timmermann B(8), Lehrach H(9), Adjaye J(10)(11).

Author information: (1)Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195, Berlin, Germany. mertes@molgen.mpg.de. (2)Molecular Exposomics, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764, Neuherberg, Germany. mertes@molgen.mpg.de. (3)Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195, Berlin, Germany. lichtner@molgen.mpg.de. (4)Next Generation Sequencing Group, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195, Berlin, Germany. kuhl@molgen.mpg.de. (5)Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195, Berlin, Germany. blattner@molgen.mpg.de. (6)Institute for stem cell research and regenerative medicine, Medical Faculty, Heinrich Heine University, Moorenstr. 5, 40225, Düsseldorf, Germany. joerg.otte@med.uni-duesseldorf.de. (7)Institute for stem cell research and regenerative medicine, Medical Faculty, Heinrich Heine University, Moorenstr. 5, 40225, Düsseldorf, Germany. wasco.wruck@med.uni-duesseldorf.de. (8)Next Generation Sequencing Group, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195, Berlin, Germany. timmerma@molgen.mpg.de. (9)Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195, Berlin, Germany. lehrach@molgen.mpg.de. (10)Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195, Berlin, Germany. james.adjaye@med.uni-duesseldorf.de. (11)Institute for stem cell research and regenerative medicine, Medical Faculty, Heinrich Heine University, Moorenstr. 5, 40225, Düsseldorf, Germany. james.adjaye@med.uni-duesseldorf.de.

BACKGROUND: Next Generation Sequencing has proven to be an exceptionally powerful tool in the field of genomics and transcriptomics. With recent development it is nowadays possible to analyze ultra-low input sample material down to single cells. Nevertheless, investigating such sample material often limits the analysis to either the genome or transcriptome. We describe here a combined analysis of both types of nucleic acids from the same sample material. METHODS: The method described enables the combined preparation of amplified cDNA as well as amplified whole-genome DNA from an ultra-low input sample material derived from a sub-colony of in-vitro cultivated human embryonic stem cells. cDNA is prepared by the application of oligo-dT coupled magnetic beads for mRNA capture, first strand synthesis and 3'-tailing followed by PCR. Whole-genome amplified DNA is prepared by Phi29 mediated amplification. Illumina sequencing is applied to short fragment libraries prepared from the amplified samples. RESULTS: We developed a protocol which enables the combined analysis of the genome as well as the transcriptome by Next Generation Sequencing from ultra-low input samples. The protocol was evaluated by sequencing sub-colony structures from human embryonic stem

cells containing 150 to 200 cells. The method can be adapted to any available sequencing system. CONCLUSIONS: To our knowledge, this is the first report where sub-colonies of human embryonic stem cells have been analyzed both at the genomic as well as transcriptome level. The method of this proof of concept study may find useful practical applications for cases where only a limited number of cells are available, e.g. for tissues samples from biopsies, tumor spheres, circulating tumor cells and cells from early embryonic development. The results we present demonstrate that a combined analysis of genomic DNA and messenger RNA from ultra-low input samples is feasible and can readily be applied to other cellular systems with limited material available.

DOI: 10.1186/s12864-015-2025-z PMCID: PMC4643517 PMID: 26564201 [Indexed for MEDLINE]

A.34. Int J Dev Biol. 2015;59(4-6):211-6. doi: 10.1387/ijdb.150172mz.

3D culture of ovarian follicles: a system towards their engineering?

Zuccotti M(1), Merico V, Rebuzzini P, Belli M, Vigone G, Mulas F, Fassina L, Wruck W, Adjaye J, Bellazzi R, Garagna S.

Author information: (1)Sezione di Anatomia, Istologia ed Embriologia, Dipartimento di Scienze Biomediche, Biotecnologiche e Traslazionali (S.Bi.Bi.T.), Università degli Studi di Parma, Italy. maurizio.zuccotti@unipr.it.

Infertility in women is a health priority. Designing a robust culture protocol capable of attaining complete follicle growth is an exciting challenge, for its potential clinical applications, but also as a model to observe and closely study the sequence of molecular events that lie behind the intricate relationship existing between the oocyte and surrounding follicle cells. Here, we describe the procedures used to maintain the ovarian follicle 3D architecture employing a variety of in vitro systems and several types of matrices. Collagen and alginate are the matrices that led to better results, including proof-of-concept of full-term development. Pioneer in its kind, these studies underlie the drawbacks encountered and the need for a culture system that allows more quantitative analyses and predictions, projecting the culture of the ovarian follicle into the realm of tissue engineering.

DOI: 10.1387/ijdb.150172mz PMID: 26505254 [Indexed for MEDLINE]

A.35. Stem Cell Res Ther. 2015 Oct 8;6:196. doi: 10.1186/s13287-015-0188-9.

Association between in vivo bone formation and ex vivo migratory capacity of human bone marrow stromal cells.

Andersen RK(1), Zaher W(2)(3), Larsen KH(4), Ditzel N(5), Drews K(6), Wruck W(7), Adjaye J(8)(9), Abdallah BM(10)(11), Kassem M(12)(13)(14).

Author information: (1)Department of Endocrinology and Metabolism, University Hospital of Odense, Odense, Denmark. rka@biomed.au.dk. (2)Department of Endocrinology and Metabolism, University Hospital of Odense, Odense, Denmark. wzaher@health.sdu.dk. (3)Stem Cell Unit, Department of Anatomy, College of Medicine, King Saud University, Riyadh, Saudi Arabia. wzaher@health.sdu.dk. (4)Department of Endocrinology and Metabolism, University Hospital of Odense, Odense, Denmark. khaubergs@gmail.com. (5)Department of Endocrinology and Metabolism, University Hospital of Odense, Odense, Denmark. nditzel@health.sdu.dk. (6)Department of Vertebrate Genomics, Molecular Embryology and Aging Group, Max Planck Institute for Molecular Genetics, Berlin, Germany. katharina_drews@gmx.de. (7)Research and Regenerative Medicine, Faculty of Medicine, Heinrich Heine University, Duesseldorf, Germany. Wasco.Wruck@med.uni-duesseldorf.de. (8)Department of Vertebrate Genomics, Molecular Embryology and Aging Group, Max Planck Institute for Molecular Genetics, Berlin, Germany. james.adjaye@med.uni-duesseldorf.de. (9)Research and Regenerative Medicine, Faculty of Medicine, Heinrich Heine University, Duesseldorf, Germany. james.adjaye@med.uni-duesseldorf.de. (10)Department of Endocrinology and Metabolism, University Hospital of Odense, Odense, Denmark. babdallah@health.sdu.dk. (11)Faculty of Science, Helwan University, Cairo, Egypt. babdallah@health.sdu.dk. (12)Department of Endocrinology and Metabolism, University Hospital of Odense, Odense, Denmark. mkassem@health.sdu.dk. (13)Stem Cell Unit, Department of Anatomy, College of Medicine, King Saud University, Riyadh, Saudi Arabia. mkassem@health.sdu.dk. (14)Danish Stem Cell Center (DanStem), Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. mkassem@health.sdu.dk.

INTRODUCTION: There is a clinical need for developing systemic transplantation protocols for use of human skeletal stem cells (also known bone marrow stromal stem cells) (hBMSC) in tissue regeneration. In systemic transplantation studies, only a limited number of hBMSC home to injured tissues suggesting that only a subpopulation of hBMSC possesses "homing" capacity. Thus, we tested the hypothesis that a subpopulation of hBMSC defined by ability to form heterotopic bone in vivo, is capable of homing to injured bone. METHODS: We tested ex vivo and in vivo homing capacity of a number of clonal cell populations derived from telomerized hBMSC (hBMSC-TERT) with variable ability to form heterotopic bone when implanted subcutaneously in immune deficient mice. In vitro transwell migration assay was used and the in vivo homing ability of transplanted hBMSC to bone fractures in mice was visualized by bioluminescence imaging (BLI). In order to identify the molecular phenotype associated with enhanced migration, we carried out comparative DNA microarray analysis of gene expression of hBMSC-derived high bone forming (HBF) clones versus low bone forming (LBF) clones. RESULTS: HBF clones were exhibited higher ex vivo transwell migration and following intravenous injection, better in vivo homing ability to bone fracture when compared to LBF clones. Comparative microarray analysis of HBF versus LBF clones identified

enrichment of gene categories of chemo-attraction, adhesion and migration associated genes. Among these, platelet-derived growth factor receptor (PDGFR) α and β were highly expressed in HBF clones. Follow up studies showed that the chemoattractant effects of PDGF in vitro was more enhanced in HBF compared to LBF clones and this effect was reduced in presence of a PDGFRβ-specific inhibitor: SU-16 f. Also, PDGF exerted greater chemoattractant effect on PDGFRβ(+) cells sorted from LBF clones compared to PDGFRβ(-) cells. CONCLUSION: Our data demonstrate phenotypic and molecular association between in vivo bone forming ability and migratory capacity of hBMSC. PDGFRβ can be used as a potential marker for the prospective selection of hBMSC populations with high migration and bone formation capacities suitable for clinical trials for enhancing bone regeneration.

DOI: 10.1186/s13287-015-0188-9 PMCID: PMC4599318 PMID: 26450135 [Indexed for MEDLINE]

A.36. PLoS One. 2014 Nov 4;9(11):e111637. doi: 10.1371/journal.pone.0111637. eCollection 2014.

miR-27 negatively regulates pluripotency-associated genes in human embryonal carcinoma cells.

Fuchs H(1), Theuser M(2), Wruck W(1), Adjaye J(3).

Author information: (1)Institute for Stem Cell Research and Regenerative Medicine, Faculty of Medicine, Heinrich Heine University, Duesseldorf, Germany. (2)Department of Vertebrate Genomics, Molecular Embryology and Aging Group, Max Planck Institute for Molecular Genetics, Berlin, Germany. (3)Institute for Stem Cell Research and Regenerative Medicine, Faculty of Medicine, Heinrich Heine University, Duesseldorf, Germany; Department of Vertebrate Genomics, Molecular Embryology and Aging Group, Max Planck Institute for Molecular Genetics, Berlin, Germany.

Human embryonic stem cells and human embryonal carcinoma cells have been studied extensively with respect to the transcription factors (OCT4, SOX2 and NANOG), epigenetic modulators and associated signalling pathways that either promote self-renewal or induce differentiation in these cells. The ACTIVIN/NODAL axis (SMAD2/3) of the TGFß signalling pathway coupled with FGF signalling maintains self-renewal in these cells, whilst the BMP (SMAD1,5,8) axis promotes differentiation. Here we show that miR-27, a somatic-enriched miRNA, is activated upon RNAi-mediated suppression of OCT4 function in human embryonic stem cells. We further demonstrate that miR-27 negatively regulates the expression of the pluripotency-associated ACTIVIN/NODAL axis (SMAD2/3) of the TGFß signalling pathway by targeting ACVR2A, TGFßR1 and SMAD2. Additionally, we have identified a number of pluripotency-associated genes

such as NANOG, LIN28, POLR3G and NR5A2 as novel miR-27 targets. Transcriptome analysis revealed that miR-27 over-expression in human embryonal carcinoma cells leads indeed to a significant up-regulation of genes involved in developmental pathways such as TGFß- and WNT-signalling.

DOI: 10.1371/journal.pone.0111637 PMCID: PMC4219743 PMID: 25369332 [Indexed for MEDLINE]

A.37. PLoS One. 2014 May 5;9(5):e92596. doi: 10.1371/journal.pone.0092596. eCollection 2014.

The nerve growth factor receptor CD271 is crucial to maintain tumorigenicity and stem-like properties of melanoma cells.

Redmer T(1), Welte Y(2), Behrens D(3), Fichtner I(3), Przybilla D(4), Wruck W(5), Yaspo ML(6), Lehrach H(6), Schäfer R(4), Regenbrecht CR(7).

Author information: (1)Institute for Pathology, Charité - Universitätsmedizin Berlin, Berlin, Germany. (2)Institute of Pathology - University Hospital Berlin, Berlin, Germany. (3)Experimental Pharmacology & Oncology Berlin-Buch GmbH, Berlin, Germany. (4)Institute of Pathology - University Hospital Berlin, Berlin, Germany; Comprehensive Cancer Center Charité - University Hospital Berlin, Berlin, Germany. (5)Institute of Pathology - University Hospital Berlin, Berlin, Germany; Laboratory of Functional Genomics (LFGC) - University Hospital Berlin, Berlin, Germany. (6)Max-Planck Institute for Molecular Genetics, Berlin, Germany. (7)Institute of Pathology - University Hospital Berlin, Berlin, Germany; Laboratory of Functional Genomics (LFGC) - University Hospital Berlin, Berlin, Germany; Comprehensive Cancer Center Charité - University Hospital Berlin, Berlin, Germany.

Erratum in PLoS One. 2014;9(8):e105274.

BACKGROUND: Large-scale genomic analyses of patient cohorts have revealed extensive heterogeneity between individual tumors, contributing to treatment failure and drug resistance. In malignant melanoma, heterogeneity is thought to arise as a consequence of the differentiation of melanoma-initiating cells that are defined by cell-surface markers like CD271 or CD133. RESULTS: Here we confirmed that the nerve growth factor receptor (CD271) is a crucial determinant of tumorigenicity, stem-like properties, heterogeneity and plasticity in melanoma cells. Stable shRNA mediated knock-down of CD271 in patient-derived melanoma cells abrogated their tumor-initiating and colony-forming capacity. A genome-wide expression profiling and gene-set enrichment analysis revealed novel connections of CD271 with melanoma-associated genes like CD133 and

points to a neural crest stem cell (NCSC) signature lost upon CD271 knock-down. In a meta-analysis we have determined a shared set of 271 differentially regulated genes, linking CD271 to SOX10, a marker that specifies the neural crest. To dissect the connection of CD271 and CD133 we have analyzed 10 patient-derived melanoma-cell strains for cell-surface expression of both markers compared to established cell lines MeWo and A375. We found CD271+ cells in the majority of cell strains analyzed as well as in a set of 16 different patient-derived melanoma metastases. Strikingly, only 2/12 cell strains harbored a CD133+ sub-set that in addition comprised a fraction of cells of a CD271+/CD133+ phenotype. Those cells were found in the label-retaining fraction and in vitro deduced from CD271+ but not CD271 knock-down cells. CONCLUSIONS: Our present study provides a deeper insight into the regulation of melanoma cell properties and points CD271 out as a regulator of several melanoma-associated genes. Further, our data strongly suggest that CD271 is a crucial determinant of stem-like properties of melanoma cells like colony-formation and tumorigenicity.

DOI: 10.1371/journal.pone.0092596 PMCID: PMC4010406 PMID: 24799129 [Indexed for MEDLINE]

A.38. Front Physiol. 2012 Sep 3;3:339. doi: 10.3389/fphys.2012.00339. eCollection 2012.

A Systems Biology Approach to Deciphering the Etiology of Steatosis Employing Patient-Derived Dermal Fibroblasts and iPS Cells.

Jozefczuk J(1), Kashofer K, Ummanni R, Henjes F, Rehman S, Geenen S, Wruck W, Regenbrecht C, Daskalaki A, Wierling C, Turano P, Bertini I, Korf U, Zatloukal K, Westerhoff HV, Lehrach H, Adjaye J.

Author information: (1)Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics Berlin, Germany.

Non-alcoholic fatty liver disease comprises a broad spectrum of disease states ranging from simple steatosis to non-alcoholic steatohepatitis. As a result of increases in the prevalences of obesity, insulin resistance, and hyperlipidemia, the number of people with hepatic steatosis continues to increase. Differences in susceptibility to steatohepatitis and its progression to cirrhosis have been attributed to a complex interplay of genetic and external factors all addressing the intracellular network. Increase in sugar or refined carbohydrate consumption results in an increase of insulin and insulin resistance that can lead to the accumulation of fat in the liver. Here we demonstrate how a multidisciplinary approach encompassing cellular reprogramming, transcriptomics, proteomics, metabolomics, modeling, network reconstruction, and data management can be employed to unveil the mechanisms underlying the progression of steatosis. Proteomics revealed

reduced AKT/mTOR signaling in fibroblasts derived from steatosis patients and further establishes that the insulin-resistant phenotype is present not only in insulin-metabolizing central organs, e.g., the liver, but is also manifested in skin fibroblasts. Transcriptome data enabled the generation of a regulatory network based on the transcription factor SREBF1, linked to a metabolic network of glycerolipid, and fatty acid biosynthesis including the downstream transcriptional targets of SREBF1 which include LIPIN1 (LPIN) and low density lipoprotein receptor. Glutathione metabolism was among the pathways enriched in steatosis patients in comparison to healthy controls. By using a model of the glutathione pathway we predict a significant increase in the flux through glutathione synthesis as both gamma-glutamylcysteine synthetase and glutathione synthetase have an increased flux. We anticipate that a larger cohort of patients and matched controls will confirm our preliminary findings presented here.

DOI: 10.3389/fphys.2012.00339 PMCID: PMC3432516 PMID: 22969728

A.39. J Biol Chem. 2011 Jul 1;286(26):23521-32. doi: 10.1074/jbc.M111.220178. Epub 2011 May 9.

The BTB and CNC homology 1 (BACH1) target genes are involved in the oxidative stress response and in control of the cell cycle.

Warnatz HJ(1), Schmidt D, Manke T, Piccini I, Sultan M, Borodina T, Balzereit D, Wruck W, Soldatov A, Vingron M, Lehrach H, Yaspo ML.

Author information: (1)Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany.

The regulation of gene expression in response to environmental signals and metabolic imbalances is a key step in maintaining cellular homeostasis. BTB and CNC homology 1 (BACH1) is a heme-binding transcription factor repressing the transcription from a subset of MAF recognition elements at low intracellular heme levels. Upon heme binding, BACH1 is released from the MAF recognition elements, resulting in increased expression of antioxidant response genes. To systematically address the gene regulatory networks involving BACH1, we combined chromatin immunoprecipitation sequencing analysis of BACH1 target genes in HEK 293 cells with knockdown of BACH1 using three independent types of small interfering RNAs followed by transcriptome profiling using microarrays. The 59 BACH1 target genes identified by chromatin immunoprecipitation sequencing were found highly enriched in genes showing expression changes after BACH1 knockdown, demonstrating the impact of BACH1 repression on transcription. In addition to known and new BACH1 targets involved in heme degradation (HMOX1, FTL, FTH1, ME1, and SLC48A1) and redox regulation (GCLC, GCLM, and SLC7A11), we also discovered BACH1 target genes affecting cell cycle and apoptosis pathways (ITPR2,

CALM1, SQSTM1, TFE3, EWSR1, CDK6, BCL2L11, and MAFG) as well as subcellular transport processes (CLSTN1, PSAP, MAPT, and vault RNA). The newly identified impact of BACH1 on genes involved in neurodegenerative processes and proliferation provides an interesting basis for future dissection of BACH1-mediated gene repression in neurodegeneration and virus-induced cancerogenesis.

DOI: 10.1074/jbc.M111.220178 PMCID: PMC3123115 PMID: 21555518 [Indexed for MEDLINE]

A.40. Stem Cells. 2007 Feb;25(2):500-10. Epub 2006 Oct 26.

Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells.

Babaie Y(1), Herwig R, Greber B, Brink TC, Wruck W, Groth D, Lehrach H, Burdon T, Adjaye J.

Author information: (1)Roslin Institute, Department of Gene Function and Development, Roslin, Midlothian, United Kingdom.

The POU domain transcription factor OCT4 is a key regulator of pluripotency in the early mammalian embryo and is highly expressed in the inner cell mass of the blastocyst. Consistent with its essential role in maintaining pluripotency, Oct4 expression is rapidly downregulated during formation of the trophoblast lineage. To enhance our understanding of the molecular basis of this differentiation event in humans, we used a functional genomics approach involving RNA interference-mediated suppression of OCT4 function in a human ESC line and analysis of the resulting transcriptional profiles to identify OCT4-dependent genes in human cells. We detected altered expression of >1,000 genes, including targets regulated directly by OCT4 either positively (NANOG, SOX2, REX1, LEFTB, LEFTA/EBAF DPPA4, THY1, and TDGF1) or negatively (CDX2, EOMES, BMP4, TBX18, Brachyury [T], DKK1, HLX1, GATA6, ID2, and DLX5), as well as targets for the OCT4-associated stem cell regulators SOX2 and NANOG. Our data set includes regulators of ACTIVIN, BMP, fibroblast growth factor, and WNT signaling. These pathways are implicated in regulating human ESC differentiation and therefore further validate the results of our analysis. In addition, we identified a number of differentially expressed genes that are involved in epigenetics, chromatin remodeling, apoptosis, and metabolism that may point to underlying molecular mechanisms that regulate pluripotency and trophoblast differentiation in humans. Significant concordance between this data set and previous comparisons between inner cell mass and trophectoderm in human embryos indicates that the study of human ESC differentiation in vitro represents a useful model of early embryonic differentiation in humans.

DOI: 10.1634/stemcells.2006-0426 PMID: 17068183 [Indexed for MEDLINE]

A.41. J Bacteriol. 2005 Feb;187(4):1493-503.

Substrate-dependent regulation of anaerobic degradation pathways for toluene and ethylbenzene in a denitrifying bacterium, strain EbN1.

Kühner S(1), Wöhlbrand L, Fritz I, Wruck W, Hultschig C, Hufnagel P, Kube M, Reinhardt R, Rabus R.

Author information: (1)Max Planck Institute for Marine Microbiology, Celsiusstr. 1, 28359 Bremen, Germany.

Anaerobic biodegradation of toluene and ethylbenzene is of environmental concern and biochemical interest due to toxicity and novel reactions, respectively. The denitrifying strain EbN1 is unique in anaerobically degrading both alkylbenzenes via different pathways which converge at benzoyl coenzyme A. The organization of genes involved in both pathways was only recently determined for strain EbN1. In the present study, global expression analysis (DNA microarray and proteomics) indicated involvement of several thus-far-unknown proteins in the degradation of both alkylbenzenes. For example, orf68 and orf57, framing the ebd operon, are implicated in ethylbenzene degradation, and the ebA1932 and ebA1936 genes, located 7.2 kb upstream of the bbs operon, are implicated in toluene degradation. In addition, expression studies were now possible on the level of the complete pathways. Growth experiments demonstrated that degradative capacities for toluene and ethylbenzene could be simultaneously induced, regardless of the substrate used for adaptation. Regulation was studied at the RNA (real-time reverse transcription-PCR and DNA microarray) and protein (two-dimensional-difference gel electrophoresis) level by using cells adapted to anaerobic growth with benzoate, toluene, ethylbenzene, or a mixture of toluene and ethylbenzene. Expression of the two toluene-related operons (bss and bbs) was specifically induced in toluene-adapted cells. In contrast, genes involved in anaerobic ethylbenzene degradation were induced in ethylbenzene- and toluene-adapted cells, suggesting that toluene may act as a gratuitous inducer. In agreement with the predicted sequential regulation of the ethylbenzene pathway, Ebd proteins (encoding subunits of ethylbenzene dehydrogenase) were formed in ethylbenzene- but not in acetophenone-adapted cells, while Apc proteins (subunits of predicted acetophenone carboxylase) were formed under both conditions.

DOI: 10.1128/JB.187.4.1493-1503.2005 PMCID: PMC545613 PMID: 15687214 [Indexed for MEDLINE]

A.42. BMC Genomics. 2004 Oct 28;5:83.

Cross-species hybridisation of human and bovine orthologous genes on high density cDNA microarrays.

Adjaye J(1), Herwig R, Herrmann D, Wruck W, Benkahla A, Brink TC, Nowak M, Carnwath JW, Hultschig C, Niemann H, Lehrach H.

Author information: (1)Max Planck Institute for Molecular Genetics, Department of Vertebrate Genomics, Ihnestrasse 73, D-14195, Berlin, Germany. adjaye@molgen.mpg.de

BACKGROUND: Cross-species gene-expression comparison is a powerful tool for the discovery of evolutionarily conserved mechanisms and pathways of expression control. The usefulness of cDNA microarrays in this context is that broad areas of homology are compared and hybridization probes are sufficiently large that small inter-species differences in nucleotide sequence would not affect the analytical results. This comparative genomics approach would allow a common set of genes within a specific developmental, metabolic, or disease-related gene pathway to be evaluated in experimental models of human diseases. The objective of this study was to investigate the feasibility and reproducibility of cross-species analysis employing a human cDNA microarray as probe. RESULTS: As a proof of principle, total RNA derived from human and bovine fetal brains was used as a source of labelled targets for hybridisation onto a human cDNA microarray composed of 349 characterised genes. Each gene was spotted 20 times representing 6,980 data points thus enabling highly reproducible spot quantification. Employing high stringency hybridisation and washing conditions, followed by data analysis, revealed slight differences in the expression levels and reproducibility of the signals between the two species. We also assigned each of the genes into three expression level categories- i.e. high, medium and low. The correlation co-efficient of cross hybridisation between the orthologous genes was 0.94. Verification of the array data by semi-quantitative RT-PCR using common primer sequences enabled co-amplification of both human and bovine transcripts. Finally, we were able to assign gene names to previously uncharacterised bovine ESTs. CONCLUSIONS: Results of our study demonstrate the harnessing and utilisation power of comparative genomics and prove the feasibility of using human microarrays to facilitate the identification of co-expressed orthologous genes in common tissues derived from different species.

DOI: 10.1186/1471-2164-5-83 PMCID: PMC535340 PMID: 15511299 [Indexed for MEDLINE]

A.43. Plant J. 2002 Dec;32(5):845-57.

Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes.

Herwig R(1), Schulz B, Weisshaar B, Hennig S, Steinfath M, Drungowski M, Stahl D, Wruck W, Menze A, O'Brien J, Lehrach H, Radelof U.

Author information: (1)Max-Planck Institute for Molecular Genetics, Ihnestr. 73, D-14195 Berlin, Germany. herwig@molgen.mpg.de

Access to the complete gene inventory of an organism is crucial to understanding physiological processes like development, differentiation, pathogenesis, or adaptation to the environment. Transcripts from many active genes are present at low copy numbers. Therefore, procedures that rely on random EST sequencing or on normalisation and subtraction methods have to produce massively redundant data to get access to low-abundance genes. Here, we present an improved oligonucleotide fingerprinting (ofp) approach to the genome of sugar beet (Beta vulgaris), a plant for which practically no molecular information has been available. To identify distinct genes and to provide a representative 'unigene' cDNA set for sugar beet, 159 936 cDNA clones were processed utilizing large-scale, high-throughput data generation and analysis methods. Data analysis yielded 30 444 ofp clusters reflecting the number of different genes in the original cDNA sample. A sample of 10 961 cDNA clones, each representing a different cluster, were selected for sequencing. Standard sequence analysis confirmed that 89% of these EST sequences did represent different genes. These results indicate that the full set of 30 444 ofp clusters represent up to 25 000 genes. We conclude that the ofp analysis pipeline is an accurate and effective way to construct large representative 'unigene' sets for any plant of interest with no requirement for prior molecular sequence data.

DOI: 10.1046/j.1365-313x.2002.01457.x PMID: 12472698 [Indexed for MEDLINE]

A.44. Bioinformatics. 2001 Jul;17(7):634-41.

Automated image analysis for array hybridization experiments.

Steinfath M(1), Wruck W, Seidel H, Lehrach H, Radelof U, O'Brien J.

Author information: (1)Max-Planck-Institute for Molecular Genetics Berlin-Dahlem, Ihnestrasse 73, D-14195 Berlin, Germany. steinfat@molgen.mpg.de

MOTIVATION: Image analysis is a major part of data evaluation for array hybridization experiments in molecular biology. The program presented here is designed to analyze automatically images from hybridization experiments with various arrangements: different kinds of probes (oligonucleotides or complex probes), different supports (nylon filters or glass slides), different labeling of probes (radioactively or fluorescently). The program is

currently applied to oligonucleotide fingerprinting projects and complex hybridizations. The only precondition for the use of the program is that the targets are arrayed in a grid, which can be approximately transformed to an orthogonal equidistant grid by a projective mapping. RESULTS: We demonstrate that our program can cope with the following problems: global distortion of the grid, missing of grid nodes, local deviation of the spot from its specified grid position. This is checked by different quality measures. The image analysis of oligonucleotide fingerprint experiments on an entire genetic library is used, in clustering procedures, to group related clones together. The results show that the program yields automatically generated high quality input data for follow up analysis such as clustering procedures. AVAILABILITY: The executable files will be available upon request for academics.

DOI: 10.1093/bioinformatics/17.7.634 PMID: 11448881 [Indexed for MEDLINE]

CV

Professional career

since 2013 **Bioinformatician** at the Medical faculty of the Heinrich Heine University, Düsseldorf

- development of software for statistical evaluation of gene expression experiments and next-generation sequencing experiments in R, C++, Perl and Java
- publication of several research articles and review articles with the main focus on meta-analysis of biopsy-derived transcriptome data and stem-cell-derived disease models
- teaching bioinformatics in a master module

2010 - 2013 **Bioinformatician** at the Charité medical university, Berlin

- data management for the systems biology project livSYSiPS which explores non-alcoholic fatty liver diseases investigating liver cells derived from induced pluripotent stem cells
- chair of the data management group of 16 transnational ERASys-Bio+ projects
- development of software for statistical evaluation of gene expression experiments and next-generation sequencing experiments in R, C++, Perl and Java
- development of a web application for the evaluation of dose-response experiments for the determination of the IC50 in Perl/CGI
- development of data base and web applications (Perl/CGI, Solr/Lucene, XML, Apache, Wiki)
- administration of Linux servers and Windows computers

1999 - 2009 **Bioinformatician** at the Max Planck Institut for Molecular Genetics

- Bioinformatics: Statistics of microarray gene expressions experiments
- development of pipelines for the evaluation of gene expression experiments and second generation sequencing data using R, perl, python
- development of a JAVA application for statistical evaluation of experiments, integration of subsystems via CORBA (common object request broker architecture)
- Image analysis in C/C++

1998 - 1999 **software developper at the Bull AG**

- Development of software for the Berlin fire department
- Implementation of a GIS (geographical information system) subsystem
- construction of interfaces via object request broker architecture
- Software development in C++ under UNIX(AIX) using an Oracle data base, Ilog/views class libraries and Rational/Rose for the documentation of object-oriented design

1990-91,93-98 **software developper at the Siemens AG**

- Development of software for telecommunication applications
- Software development in C/C++ under UNIX(Solaris,HPUX) and Windows using proprietary software analysis and design tools

1989 - 1990 **student at the GMD**

- Development of software for the distributed SUPRENUM architecture (super-computer on a multi-processor base with the distributed operating system PEACE developped in the research project)

Studies

1991 - 1997	**Study of Architecture, Diploma 1997 at TU Berlin**
1990	**Diploma** at the **GMD (Gesellschaft für Mathematik und Datenverarbeitung) and the Technical University Berlin/Department of Computer science**: Subject of the Diploma work "Modellierung und Implementierung einer Objektverwaltung des Checkpoint/Backward-Recovery-Verfahrens für die verteilte SUPRENUM-Architektur" ("Modelling and Implementation of an object administration system of the checkpoint/backward recovery procedure for the distributed SUPRENUM architecture")
1982 - 1990	**Study of computer science** at the Technical University Berlin

School

1982	**Abitur**
1973 - 1982	**High school** Katharineum zu Lübeck
1969 - 1973	**Elementary school** Hardenberg-Schule, Kiel
1963	**born in Lübeck**

Erklärung

Ich, Wasco Wruck, versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

_____

Wasco Wruck

Datum: