

# **Genetic and Epigenetic Landscapes of Pediatric Acute Lymphoblastic Leukemia Subtypes**

Inaugural dissertation

for the attainment of the title of doctor  
in the Faculty of Mathematics and Natural Sciences  
at the Heinrich Heine University Düsseldorf

presented by

**Mareike Dörrenberg**  
from Recklinghausen

Düsseldorf, September 2019

from the Department of Pediatric Oncology, Hematology and Clinical Immunology at the  
Heinrich Heine University Düsseldorf

Published by permission of the  
Faculty of Mathematics and Natural Sciences at  
Heinrich Heine University Düsseldorf

Supervisor: Prof. Dr. Arndt Borkhardt  
Co-supervisor: Prof. Dr. Holger Gohlke

Date of the oral examination: 16.12.2019

## Contents

1.	Introduction.....	1
1.1	Lymphopoiesis.....	1
1.2	Acute Lymphoblastic Leukemia.....	2
1.2.1	Precursor B-cell ALL.....	4
1.2.2	Pro-B ALL.....	7
1.2.3	Infant T-cell Acute Lymphoblastic Leukemia.....	9
1.3	miRNAs in Leukemia.....	10
1.4	Aim of the Thesis.....	13
2.	Materials.....	14
2.1	Samples.....	14
2.1.1	Patient Samples.....	14
2.1.2	Human Cell Lines.....	14
2.2	Chemicals.....	16
2.3	Nucleic Acids.....	17
2.3.1	Oligonucleotides.....	17
2.3.2	Nucleotides.....	19
2.3.3	Vectors and Vector Cards.....	19
2.3.4	siRNAs (Small Interfering RNAs).....	21
2.3.5	Northern Blot Probes (DNA):.....	21
2.4	Enzymes.....	22
2.5	Bacteria.....	23
2.6	Culture Media, HDAC-Inhibitor and Buffer.....	23
2.6.1	Media and Additives for Cell Cultivation.....	23
2.6.2	HDAC (Histone Deacetylase) Inhibitors.....	23
2.7	Kits.....	23

2.8	Software.....	24
2.8	Hardware.....	25
2.9	Other Materials.....	25
3.	Methods.....	26
3.1	Cell Culture.....	26
3.1.1	Cultivation of Human Suspension Cell Lines.....	26
3.1.2	Cultivation of Human Adherent Cell Lines.....	26
3.1.3	Cryopreservation of Human Cell Lines.....	26
3.1.4	Treatment with Inhibitors.....	27
3.2	Methods of Molecular Biology.....	27
4.2.1	Extraction of Nucleic Acids.....	27
3.2.3	Extraction of Total RNA by TRIzol Reagent.....	27
3.2.2	Determination of Concentration and Purity of Nucleic Acids.....	28
3.2.3	Reverse Transcription.....	28
3.2.4	Agarose Gel Electrophoresis.....	28
3.2.5	Restriction Digest.....	28
3.2.6	Gel Extraction.....	29
3.2.7	Ligation.....	29
3.2.8	Transformation of One Shot™ Top10 Competent E.coli.....	29
3.2.9	Plasmid Extraction.....	29
3.2.10	Cloning of RNAi Expression Vectors.....	30
3.2.11	Lipofection.....	30
3.2.12	Luciferase Assay.....	31
3.2.13	Whole Genome Amplification.....	31
3.2.14	Sanger Sequencing.....	32
3.2.15	Northern Blot.....	32
3.2.16	Western Blot.....	33

3.3	Polymerase Chain Reaction (PCR).....	34
3.3.1	Standard PCR.....	34
3.3.2	Quantitative Real Time PCR (qRT-PCR).....	34
3.4	Next Generation Sequencing.....	34
3.4.1	Quality Control of Extracted RNA.....	34
3.4.2	miRNA Library Preparation.....	35
3.4.3	Total RNA Library Preparation.....	35
3.4.4	Library Validation.....	35
3.4.5	Loading of the Flow Cell.....	35
3.4.6	Deep Sequencing.....	35
3.5	Bioinformatic Methods.....	36
3.5.1	Analysis of Transcriptome Data.....	36
3.5.2	Pathway Analysis.....	36
3.5.3	Correlation Analysis.....	36
3.5.4	Analysis of Exome Data.....	37
4.	Results.....	38
4.1	Differences of the mRNA/miRNA expression pattern between t(1;19)-ALL and t(17;19)-ALL.....	38
4.1.1	Patient Cohort.....	38
4.1.2	Target Prediction of Differentially Expressed Genes.....	40
4.1.3	Novel miRNAs.....	60
4.1.4	Mutated miRNAs.....	62
4.1.4	Treatment of <i>TCF3/HLF</i> -positive cell line with HDAC6 Inhibitor KSK.....	63
4.2	Differences between Infant and Childhood T-ALL.....	65
4.2.1	Patient Cohort.....	65
4.2.2	Mutations in iT-ALL Samples.....	66
4.2.3	Differences in Gene Expression.....	68

4.2.4	Upstream Regulator Analysis.....	70
4.2.5	Druggable Targets.....	72
4.2.6	Target Prediction of DE miRNAs.....	73
5.	Discussion.....	79
5.1	Differences between mRNA and miRNA target pairs t(1;19) and t(17;19) positive ALL.....	79
5.2	Novel miRNAs in t(1;19) and t(17;19) ALL.....	82
5.3	Mutated miRNAs in t(1;19) and t(17;19).....	83
5.4	Treatment of t(17;19) ALL with HDAC6 inhibitor KSK.....	83
5.5	Outlook.....	85
5.6	Differences between Infant and Childhood T-ALL.....	85
6.	Literature.....	88
7.	List of Abbreviations.....	98
8.	Acknowledgements.....	104

## **Abstract**

Acute lymphoblastic leukemia (ALL) is a malignant disease of the hematopoietic system. It is distinguished in B-cell- and T-cell acute lymphoblastic leukemia depending on the affected cell type. Both leukemic types can be further divided into subgroups regarding genetic aberrations, which occur in the leukemic genomes and are mainly responsible for tumor development. One aspect of this thesis is the analysis of differences between two of these tumor subtypes of B-cell ALL. The first subtype is characterized by a translocation between chromosomes 1 and 19, leading to a fusion of the two transcription factors *TCF3* and *PBX1*. It is a frequently occurring leukemic subtype in childhood, inferring a good prognosis. The second subtype is characterized by a translocation between chromosomes 17 and 19, leading to a fusion of the two transcription factors *TCF3* and *HLF*. This leukemic subtype is very rare and associated with a poor prognosis.

Another aspect of leukemic grouping is the age of patients. Especially in T-cell ALL the development of leukemia in infancy is extremely rare compared to a development in childhood. Clinical and molecular knowledge about infant T-cell ALL is nearly lacking and it is not clear, if it represents a distinct disease compared to childhood T-cell ALL. Therefore, the second topic of this thesis deals with the analysis of differences between infant and childhood T-cell ALL.

To address both topics, patient material was sequenced on an Illumina next generation sequencing platform. Different filter settings of the data analysis pipeline were used to detect differentially expressed genes and miRNAs between the two B-cell ALL subtypes.

Correlation analysis and target prediction was performed to find potential mRNA-miRNA target pairs which, however, could not be validated *in-vitro* by qPCR and luciferase assay. Potential novel miRNAs were found using miRDeep, which –again- could not be validated by qPCR and Northern blot. Hsa-miR-7847 was found to be mutated in some t(1;19) patients, which has to be validated. A novel HDAC6 inhibitor was tested on a t(17;19) cell line, leading to differences in gene expression.

Exome sequencing was performed to identify differences between infant and childhood T-cell ALL and mutations in *NOTCH2*, *NOTCH3*, *PTEN* and *KRAS* were detected in infant patients. Analysis of transcriptomes revealed 760 differentially expressed genes and 58 differentially expressed miRNAs between both cohorts. Correlation analysis and target prediction revealed 47 potential mRNA-miRNA target pairs of which the most have an impact on tumor development. A pathway analysis revealed affected pathways in infant patients, which rely on immune functions or tumorigenesis like the activation of the *ERK5* pathway.

In summary, predicted interactions between mRNAs and miRNAs as well as predicted novel miRNAs could be disproved by different techniques of molecular biology by analyzing t(1;19)- and t(17;19) B-cell ALL. Nevertheless, we found distinct molecular features in infant compared to childhood T-cell ALL on a transcriptomic and epigenetic level, which have a potential impact on development and course of disease.



## Zusammenfassung

Akute lymphatische Leukämie ist eine maligne Krankheit des hämatopoetischen Systems. Je nachdem welcher Zelltyp betroffen ist, wird die akute lymphatische Leukämie (ALL) in B-Zell- und T-Zell ALL eingeteilt. Beide Leukämietypen können in Subtypen unterteilt werden. Diese werden über genetische Veränderungen im Patientengenom definiert, die für die Tumorentstehung hauptverantwortlich sind. Ein Teilgebiet dieser Arbeit ist die Analyse von Unterschieden zwischen zwei Tumorsubtypen der B-Zell ALL. Der erste Subtyp wird durch eine Translokation der beiden Chromosomen 1 und 19 charakterisiert. Diese Translokation führt zu einem Fusionsprotein aus den beiden Transkriptionsfaktoren *TCF3* und *PBX1*. Dieser leukämische Subtyp tritt häufig im Kindesalter auf und besitzt eine hohe Heilungschance. Der zweite Subtyp wird durch die Translokation der beiden Chromosomen 17 und 19 charakterisiert. Durch diese entsteht ein Fusionsprotein der beiden Transkriptionsfaktoren *TCF3* und *HLF*. Dieser leukämische Subtyp kommt sehr selten vor und ist schwer zu behandeln.

Ein weiterer Aspekt der Gruppierung von Leukämietypen ist das Alter der Patienten. Im Vergleich zu der Entstehung einer T-Zell Leukämie im Kindesalter, ist die Entstehung einer Leukämie im Säuglingsalter sehr selten.

Kenntnisse über klinische und molekulare Grundlagen der T-Zell akuten lymphatischen Leukämie im Säuglingsalter ist kaum vorhanden. Außerdem ist nicht klar, ob eine T-Zell ALL im Säuglingsalter im Vergleich zur T-Zell ALL im Kindesalter ein eigenes Krankheitsbild darstellt. Daher beschäftigt sich das zweite Thema dieser Arbeit mit der Analyse von Unterschieden zwischen einer T-Zell ALL im Säuglings- und im Kindesalter.

Für beide Analysen wurde Patientenmaterial auf einer Illuminaplattform sequenziert. Um differentiell exprimierte Gene und miRNAs zwischen den beiden B-Zell Leukämietypen zu erfassen, wurden verschiedene Filter der Datenanalyse-Pipeline verwendet. Eine Korrelationsanalyse sowie eine *target prediction* wurde durchgeführt, um potentielle mRNA-miRNA *target-pairs* zu entdecken, die jedoch *in-vitro* durch qPCR und einen Luciferase-Assay nicht validiert werden konnten. Mit miRDeep wurden potentielle neue miRNAs gefunden, die jedoch ebenfalls nicht durch qPCR und *Northern blot* validiert werden konnten. Anhand der Sequenzierdaten wurde aufgedeckt, dass hsa-miR-7847 in einigen Patienten mit t(1;19) mutiert im Genom vorliegt, was zukünftig validiert werden wird. Ein neuer HDAC6 Inhibitor wurde an einer Zelllinie mit t(17;19) getestet. Dieses führte nachweislich zu einer Veränderung in der Genexpression.

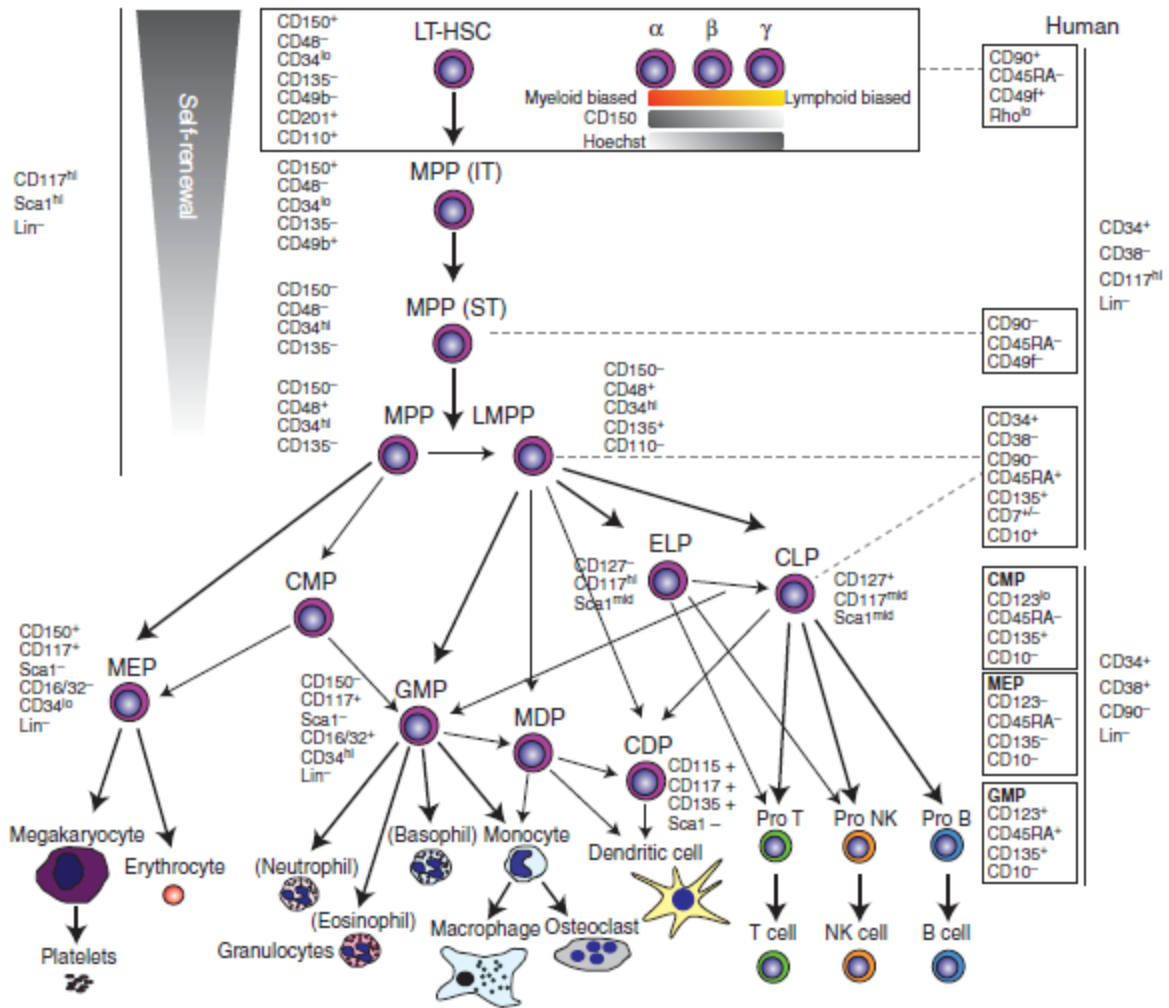
Um Unterschiede zwischen einer T-Zell-Leukämie im Säuglings- und im Kindesalter zu analysieren, wurde eine Exomsequenzierung durchgeführt. Dabei wurden Mutationen in den Genen *NOTCH2*, *NOTCH3*, *PTEN* und *KRAS* in den Säuglingspatienten gefunden. Die Analyse der Transkriptome zeigte 760 differentiell exprimierte Gene und 58 differentiell exprimierte miRNAs zwischen beiden Kohorten. Über eine Korrelationsanalyse und eine *target prediction* wurden 47 potentielle mRNA-miRNA *target pairs* gefunden, von denen die meisten einen Einfluss auf die Tumorentstehung haben. Eine *pathway*-Analyse deckte betroffene *pathways* in der Säuglings-T-Zell ALL auf, die im Zusammenhang mit Funktionen des Immunsystems und der Tumorentwicklung stehen, wie z.B. die Aktivierung des *ERK5 pathways*.

Bei Analyse von t(1;19)- und t(17;19)-ALL konnten vorhergesagte Interaktionen zwischen mRNAs und miRNAs sowie das Vorhandensein potentiell neuer miRNAs durch Anwendung verschiedener molekularbiologischer widerlegt werden. Trotzdem wurden deutliche molekulare Merkmale in Transkription und Epigenom in der Säuglings-T-Zell Leukämie im Vergleich zur T-Zell Leukämie im Kindesalter entdeckt, welche einen potentiellen Einfluss auf die Entstehung und Entwicklung einer Säuglings-T-Zell Leukämie haben können.

# 1. Introduction

## 1.1 Lymphopoiesis

Mammalian blood consists of many different kinds of cells. All of them arise constantly from multipotent hematopoietic stem cells (Figure 1). This process is called hematopoiesis and it takes place in the bone marrow and lymphatic organs<sup>1</sup>. The hematopoietic stem cell can undergo asymmetric cell division, producing a hematopoietic stem cell with the full potential of differentiation and a progenitor cell with less capability to differentiate. The latter can be a myeloid or lymphoid progenitor cell. The myeloid progenitor cell can divide and form megakaryocytes, which can further differentiate to thrombocytes, to monoblasts, which can differentiate to monocytes and further on to dendritic cells, to myeloblasts, which can differentiate to different types of granulocytes, to mast cells or to proerythroblasts, which can differentiate to erythrocytes. The lymphoid progenitor can divide and build B-lymphocytes and T-lymphocytes, dendritic cells, natural killer cells or innate lymphoid cells. Different sets of growth factors and transcription factors as well as stem cell niches determine the development of progenitor and stem cells ADDIN EN.CITE <sup>2,3</sup>. Aberrations of these factors can lead to severe lymphocytic diseases. Chromosomal translocations or other mutations are often found in transcription factors in blasts of patients suffering from chronic myeloid leukemia (CML), chronic lymphoblastic leukemia (CLL), acute myeloid leukemia (AML) or acute lymphoblastic leukemia (ALL) ADDIN EN.CITE <sup>4</sup>.



**Figure 1: Adult hematopoietic differentiation hierarchy.** Long-term self-renewing hematopoietic stem cells have the potential to differentiate to all blood cell lineages. Hematopoietic progenitor cell stages are defined by distinct surface marker expression in the murine and human system. This picture shows the murine differentiation hierarchy. Corresponding human progenitor cell populations and their markers are indicated. Bold arrows: long described differentiation routes; thin arrows: more recently described differentiation routes; HSC: hematopoietic stem cell; MPP: multipotent progenitor; LT-: long-term repopulating; IT-: intermediate-term repopulating; ST-: short-term repopulating; LMPP: lymphoid-primed MPP; ELP: early lymphoid progenitor; CLP: common lymphoid progenitor; CMP: common myeloid progenitor; GMP: granulocyte-macrophage progenitor; MEP: megakaryocyte-erythrocyte progenitor; CDP: common dendritic progenitor; MDP: monocyte-dendritic cell progenitor; NK: natural killer cell. Figure taken from Rieger and Schroeder, 2012<sup>2</sup>.

## 1.2 Acute Lymphoblastic Leukemia

Leukemia is a malignant disease of the hematopoietic system and the most frequent cancer type in children. About 30% of all pediatric cancer patients are diagnosed with leukemia (Deutsches Kinderkrebsregister, version 2016 based on 14791 patients). ALL is the most common type of pediatric leukemia.

ALL is characterized by an increased production of aberrant lymphoblastic cells in the bone marrow and peripheral blood. Clinically, it is often accompanied by fatigue, bleeding and recurrent infections<sup>5</sup>.

ALL was first classified by the French-American-British (FAB) Co-Operative Group in 1976 based on morphological criteria<sup>6</sup>. The latest version of classification originates from the World Health Organization (WHO) in 2008, detailing morphological, cytochemical, immunophenotypical, genetic and clinical features ADDIN EN.CITE <sup>7</sup>. ALL classification allows differentiating between different risk-classes. According to them, different therapy strategies are applied. Patients with a good prognosis receive less intensive medication to avoid undue toxicity. Patients with a poorer prognosis receive a more aggressive therapeutic approach to increase the probability of cure and long-time remission ADDIN EN.CITE <sup>8</sup>.

In 1975 the Berlin-Frankfurt-Münster (BFM) study group was founded to coordinate multicentric clinical trials of leukemias and lymphomas. The BFM study group developed the BFM treatment backbone, a chemotherapeutic treatment approach of eight different drugs. Today, this treatment backbone is the foundation of many risk-based therapy approaches<sup>9</sup>.

In 1993 the Childrens' Cancer Group (CCG), the Pediatric Oncology Group (POG), the Dana-Farber Cancer Institute (DFCI), the St Jude Children's Research Hospital (SJCRH) and the Cancer Therapy Evaluation Program (CTEP) defined uniform criteria for risk-directed treatment assignment. Initial white blood cell count (WBC) at diagnosis, central nervous system (CNS) status, age, ALL subtype, chromosomal abnormalities, cytogenetics, immunophenotype and early response to treatment are factors, which determine the risk-class. In general, patients with a lower WBC ( $< 50.000/\mu\text{L}$ ) have a better prognosis than patients with a high WBC ( $> 50.000/\mu\text{L}$ ). Infants and adolescents have a poorer prognosis than children with intermediate age (12 months until ten years of age) ADDIN EN.CITE <sup>8</sup> and patients with hyperdiploid blast cells also have a better outcome ADDIN EN.CITE <sup>10</sup>. Also, an early therapy response is a promising indicator of a good prognosis. Patients with CNS-invasion are treated with high risk protocols with many of them receiving craniospinal irradiation ADDIN EN.CITE <sup>11</sup>. ALL patients with a T-cell immunophenotype often present with a higher WBC at initial diagnosis compared to ALL patients with a B-cell immunophenotype and are therefore usually also treated with high risk protocols<sup>12</sup>. Chromosomal translocations also contribute to the development of ALL – some are accompanied with a poor prognosis, like t(4;11), some are accompanied with a better

prognosis, like the t(12;21) ADDIN EN.CITE <sup>13</sup>.

The use of Next Generation Sequencing (NGS) brought deeper insight into the development and progression of ALL. Detection of chromosomal translocations and other mutations affecting genes important for apoptosis, proliferation, differentiation and drug resistance allows a more detailed classification of risk classes and application of individualized treatment strategies. In the last decades, a five and more year event free survival (EFS) rate increased to approximately 90% of all children with ALL in developed countries because of improved therapeutic strategies<sup>14</sup>. However, there are still some classes of leukemia with high rates of treatment failure. NGS will likely further improve therapeutic strategies by elucidation of disease patterns of these kinds of leukemias<sup>14</sup>.

### **1.2.1 Precursor B-cell ALL**

The differentiation of lymphoid cells is accompanied by the production of a set of markers, specific to their lineage and their differentiation stage. Precursor B-cells are large cells with a small amount of cytoplasmic immunoglobulin M (cIgM) and no detectable surface immunoglobulin (sIg), but with a high proliferation rate. Pre-B-ALL is a malignant disease with an increased proliferation of aberrant precursor B-cells. The pre-B immunophenotype was first described in 1978 by Vogler *et al.* ADDIN EN.CITE <sup>15</sup>. The leukemic cells express cytoplasmic  $\mu$  heavy chain (C $\mu$ ), terminal deoxynucleotidyl transferase (TdT), BA-1-antigens and are CD10<sup>+</sup>/CD5<sup>-</sup>/CD7<sup>-</sup>/CD13<sup>-</sup>/CD19<sup>+</sup>/CD20<sup>+</sup>/CD21<sup>-</sup>/CD22<sup>+</sup>/CD24<sup>+</sup>/HLA-DR<sup>+</sup> ADDIN EN.CITE <sup>16,17</sup>. An expression of  $\kappa$ - and  $\lambda$ -light chains is missing. ADDIN EN.CITE <sup>18</sup> Pre-B-cell ALL is the most common ALL type. It is divided into subtypes regarding genetic aberrations in the blast cells. The most common structural aberrations are t(9;22) (*BCR/ABL1*), t(v;11q23) (v = variable) (*KMT2A* rearrangement), t(12;21) (*ETV6/RUNX1*) and t(1;19) (*TCF3/PBX1*)<sup>19</sup>. A less common chromosomal translocation is t(17;19) (*TCF3/HLF*) ADDIN EN.CITE <sup>20</sup>.

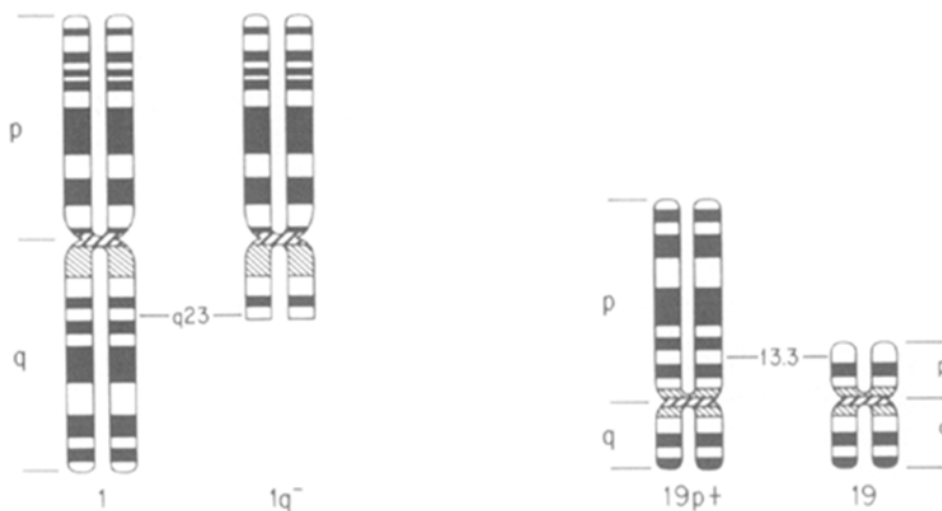
#### **1.2.1.1 Pre-B-cell ALL with t(1;19)**

Around 5% of all pediatric ALL patients and 20% of all pre-B-cell ALL patients harbor the translocation t(1;19)(q23;p13.3) in their blast cells<sup>21</sup>, ADDIN EN.CITE <sup>22,23</sup>. Also, 3% of adult patients exhibit this translocation ADDIN EN.CITE <sup>24,25</sup>. Gender ratio of t(1;19) shows

a higher occurrence in woman than in men<sup>26</sup>.

The t(1;19) positive ALL is accompanied with a moderate organomegaly, frequent CNS involvement and a high WBC<sup>27</sup>. With intensive treatment protocols, a good prognosis is achievable with a median five year EFS-rate of around 85% ADDIN EN.CITE <sup>28</sup>.

t(1;19) involves the long arm of chromosome 1 and the short arm of chromosome 19, with breakpoints at 1q23 and 19p13.3 (Figure 2). It was first described in 1984 by A. J. Carroll et al.<sup>29</sup>. The immunophenotype of t(1;19) blasts differs from the immunophenotype of B-ALL blasts without this translocation (CD19<sup>+</sup>/CD10<sup>+/-</sup>/CD34<sup>+</sup>/CD20<sup>+/-</sup>) ADDIN EN.CITE <sup>17,30</sup>. Two types of t(1;19) exist – the balanced t(1;19) and the unbalanced t(1;19). More than 80% of all patients have an unbalanced chromosomal translocation (-19, +der(19) t(1;19)(q23;p13)) in which the derivative chromosome 1 is lost and just the derivative chromosome 19 exists. Less than 20% of all t(1;19) patients have a balanced chromosomal translocation, with both derivative chromosomes still present<sup>31</sup>. Pre-B ALL lymphoblasts with both, t(1;19)(q23;p13) or its derivate der(19)t(1;19)(q23;p13), have the same immunophenotype TdT<sup>+</sup>/CD19<sup>+</sup>/CD10<sup>+</sup>/CD22<sup>+</sup>/CD34<sup>-</sup>/CD20<sup>+/-</sup>/CD9<sup>+</sup> and are mostly hyperdiploid. There is also no difference between both translocations regarding patients' age, sex, race, leukocyte and platelet count, hemoglobin level, liver and spleen size or in treatment outcome ADDIN EN.CITE <sup>32</sup>.



Adopted from Williams *et al.*, 1984

**Figure 2: The t(1;19) translocation.** G-Band pattern of the chromosomes 1, 1q-, 19+ and 19. The breakpoint in 1 is q23. The breakpoint in 19 is p13.3.<sup>33</sup>

t(1;19) leads to a fusion product with the two activation domains (AD) of the helix-loop-helix transcription factor 3 (*TCF3*) at the N terminus and the homeodomain (HD) of DNA binding protein Pre-B cell leukemic homeobox 1 (*PBX1*) (formerly designated pre B-cell leukemia (prl)) at the C terminus ADDIN EN.CITE <sup>34</sup>, ADDIN EN.CITE <sup>35</sup>.

*TCF3*, also known as *E2A*, encodes the two proteins E12 and E47. In a dimerized state, both proteins bind to the kappa E-box site kE2 by a helix-loop-helix region and regulate the expression of immunoglobulin kappa light chain<sup>36</sup>. Both proteins are involved in developmental pathways including B- and T-cell differentiation ADDIN EN.CITE <sup>37,38</sup>.

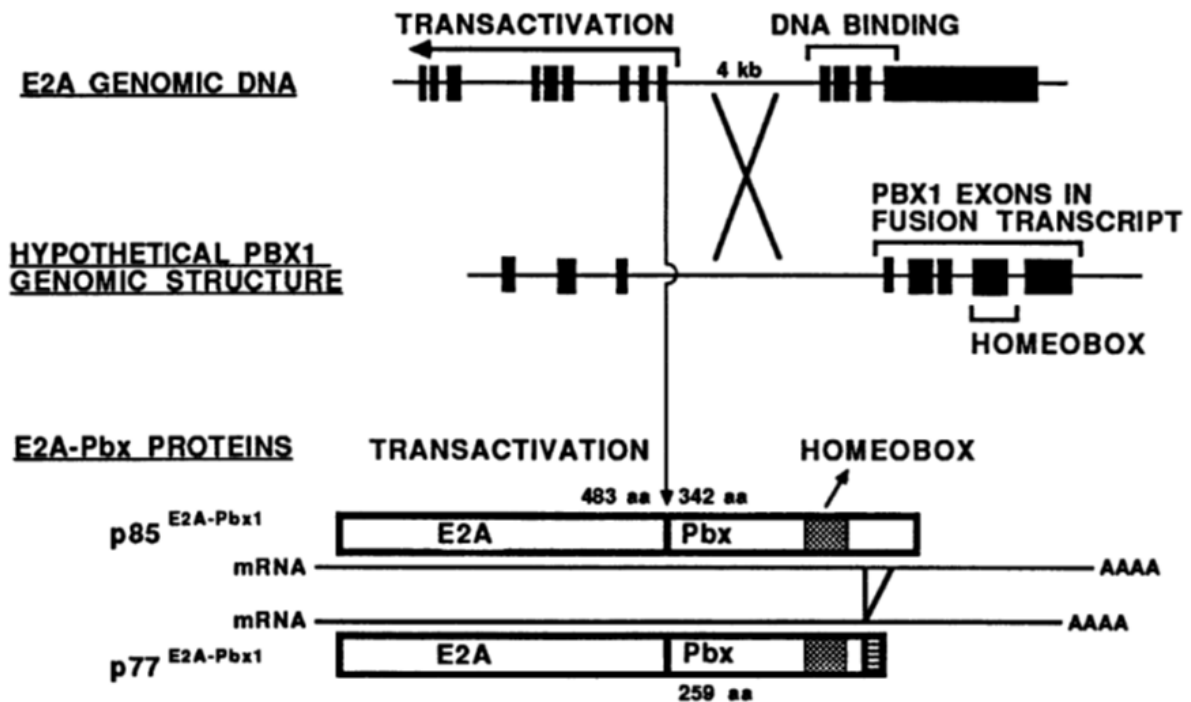
*PBX1* is part of a gene family (*PBX1*, *PBX2* and *PBX3*), whose amino acid sequence is highly conserved within as well as flanking their homeodomains ADDIN EN.CITE <sup>39</sup>. It has two different splicing forms varying in the C-terminal sequence which also exist in the chimeric *TCF3/PBX1* protein ADDIN EN.CITE <sup>35</sup>. The whole *PBX* family as well as the fusion protein bind to the *PBX1* recognition sequence (PRS) ADDIN EN.CITE <sup>40</sup>. *PBX2* and *PBX3* have a ubiquitous expression pattern while *PBX1* expression is lacking in B- and T-lymphoblasts ADDIN EN.CITE <sup>39</sup>.

In most t(1;19)-patients the breakpoints occur within the same introns of *TCF3* and *PBX1* leading to fusion of *TCF3* and *PBX1* in the same locus<sup>41</sup> (Figure 3). Around 15% of patients with t(1;19) express another variant of *TCF3/PBX1*-mRNA in addition to the expected one. This mRNA variant has a 27 base pair (bp) insertion, leading to a replacement of valine with ten amino acids in the protein ADDIN EN.CITE <sup>42</sup>. Five different protein species of the fusion protein were found in patient lymphoblasts and 697, a cell line also harboring the t(1;19) translocation. Two species arise by alternative splicing and the other ones by posttranslational modifications. The two *TCF3/PBX1* splicing forms p85<sup>E2A-Pbx1</sup> and p77<sup>E2A-Pbx1</sup> differ in carboxyterminal amino acids of *PBX1* sequence leading to proteins with different transforming capacity ADDIN EN.CITE <sup>43</sup>. Experiments with mice showed the emergence of leukemia after intravenous injection of patient leukemic blasts with t(1;19) ADDIN EN.CITE <sup>16</sup>.

Although t(1;19) is considered as a pre-B ALL chromosomal translocation<sup>33</sup>, it also appears in other types of ALL with low frequency<sup>44</sup>, ADDIN EN.CITE <sup>23</sup>. While *TCF3/PBX1* RNA and protein are detectable in cIg<sup>+</sup> lymphoblasts (pre-B ALL), this phenotype is often absent in cIg<sup>-</sup> lymphoblasts with t(1;19), which represent an early pre-B ALL ADDIN EN.CITE <sup>45,46</sup>. The surface phenotype of t(1;19) blasts with *TCF3/PBX1*<sup>+</sup> shows homogeneous expression of



CD19, CD10 and CD9 with lack of CD34 and full or partial lack of CD20, whereas *TCF3/PBX1*-t(1;19) blasts do not show this surface phenotype<sup>17</sup>.



Taken from Kamps et al., 1993

**Figure 3: Chromosomal translocation t(1;19) with effects on mRNA and protein.** N terminus of *TCF3* (*E2A*) with its transactivation domain and C terminus of *PBX1* with its homeobox domain are fused together. Two protein variants arise (p88<sup>E2A-Pbx1</sup> and p77<sup>E2A-Pbx1</sup>) by differential splicing. [ADDIN EN.CITE](#) <sup>47</sup>

## 1.2.2 Pro-B ALL

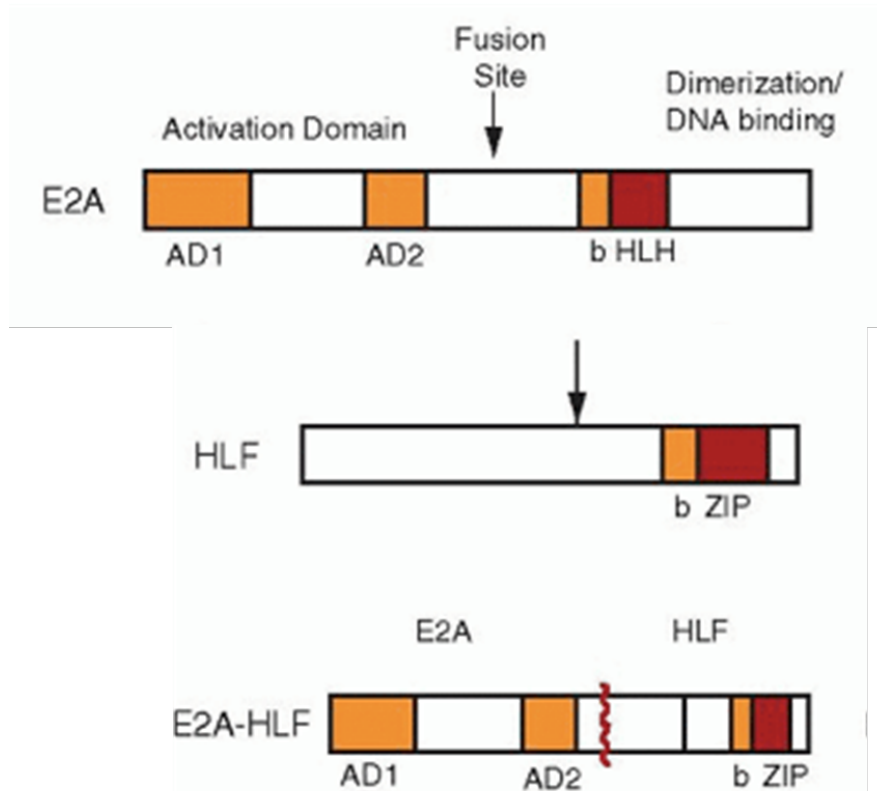
Early pre-B cell ALL or pro-B-ALL describes a leukemic B-cell subtype, in which B-cells in a very early differentiation state proliferate. These cells are identified by the specific surface phenotype CD10<sup>+</sup>/CD19<sup>+</sup>/CD22<sup>+</sup>/CD34<sup>+</sup>/CD79a<sup>+</sup><sup>48</sup> first described in 1982 by Nadler et. al.<sup>49</sup>. Translocations which often accompany this leukemic subtype are t(4;11)(q21;q23) with an expression of *KMT2A/AFF1* fusion gene [ADDIN EN.CITE](#) <sup>50</sup>. Another translocation is the rare t(17;19) [ADDIN EN.CITE](#) <sup>20</sup>.

### B-cell ALL with t(17;19)

A t(17;19) (q21-22;p13) in childhood pro-B cell or early B-cell ALL is a very rare event with an epidemiologic rate of less than 1% of all pediatric ALL cases<sup>51</sup>. The prognosis for patients with t(17;19) is extremely poor. It is associated with relapse and death within 2 years [ADDIN](#)

EN.CITE <sup>52-55</sup>.

t(17;19) positive ALL was first described in 1991 by Raimondi et al. ADDIN EN.CITE <sup>56</sup>. It involves chromosomal bands 17q21-q22 and the short arm of chromosome 19 on band q13.3, which is the same as in t(1;19) ADDIN EN.CITE <sup>56</sup>. It leads to a fusion product of the two activation domains of *TCF3* at the N terminus and the DNA binding domain basic leucine zipper (bZIP) of hepatic leukemic factor (*HLF*) at the C terminus (Figure 4). The chimeric protein contains the identical proportion of *TCF3* compared to the chimeric protein arisen of t(1;19). An insertion often occurs in the breakpoint region ADDIN EN.CITE <sup>20</sup>.



Taken and modified from Thompson and Davé, 2016

**Figure 4: Chimeric protein of TCF3 (E2A) and HLF.** The N terminus of TCF3 (E2A) with its two activation domains is fused to the C terminus of HLF with its basic leucine zipper domain. Arrows mark the fusion sites. AD: activation domain, bHLH: basic helix-loop-helix, bZIP: basic leucine zipper<sup>57</sup>.

HLF is a proline and acidic amino-acid-rich region (PAR) protein, which binds to the thyrotrophic embryonic factor (TEF) growth hormone (GH) site as a homodimer or a heterodimer with TEF or D-box binding PAR BZIP transcription factor (*DBP*). HLF and *TEF* have high structural similarity and are closely related to each other. Transcriptional activation sites and bZIP domains show a similarity of 70% to 80% and both proteins share the same DNA binding consensus sequence ADDIN EN.CITE <sup>58</sup>. HLF is primarily expressed in

hepatocytes. In normal peripheral blood mononuclear cells and hematolymphoid cell lines without t(17;19), no HLF expression is detectable ADDIN EN.CITE <sup>20</sup>.

The *HLF* gene spans about 40 kb genomic DNA and consists of four exons, while *TCF3* gene spans about 19 exons. *HLF* exon number four contains the information for the bZIP domains, which is completely present in the *TCF3/HLF* fusion construct. The extended basic region and proline and acidic amino acid-rich region are contained in exon three. Two different chromosomal rearrangements occur to build the *TCF3/HLF* fusion construct. In the first the breakpoint region contains a part of *TCF3* intron 13, an insertion of 27 non-template nucleotides and a part of *HLF* intron three. The second breakpoint region leads to a construct of *TCF3* exon 12 directly fused to *HLF* exon 4. Both constructions ensure the frame of *HLF* exon four in the fusion constructs ADDIN EN.CITE <sup>20</sup>.

Binding site analysis of chimeric *TCF3/HLF* showed a weak binding to GH promotor region in a monomeric state. The binding increased in a homo- or heterodimerized state with the related PAR proteins TEF or DBP, but it is still weaker than binding of wild type (WT) HLF ADDIN EN.CITE <sup>20</sup>.

### **1.2.3 Infant T-cell Acute Lymphoblastic Leukemia**

In T-ALL an uncontrolled expansion of T-lymphoblasts leads to ALL. Around 15% of all pediatric ALL patients suffer from T-ALL ADDIN EN.CITE <sup>59</sup>, which represents a more aggressive subtype of leukemia than B-ALL with a higher risk of relapse ADDIN EN.CITE <sup>60</sup>.<sup>61</sup> It is accompanied with a high WBC at initial diagnosis and is usually considered a high risk leukemia. Therefore an intensive chemotherapy treatment regimen is used. This often leads to severe acute toxicities and long-term side effects like development of secondary tumors<sup>62</sup>. T-ALL is classified into four differentiation stages, namely: pro-T-ALL, pre-T-ALL, cortical T-ALL and mature T-ALL<sup>63</sup>.

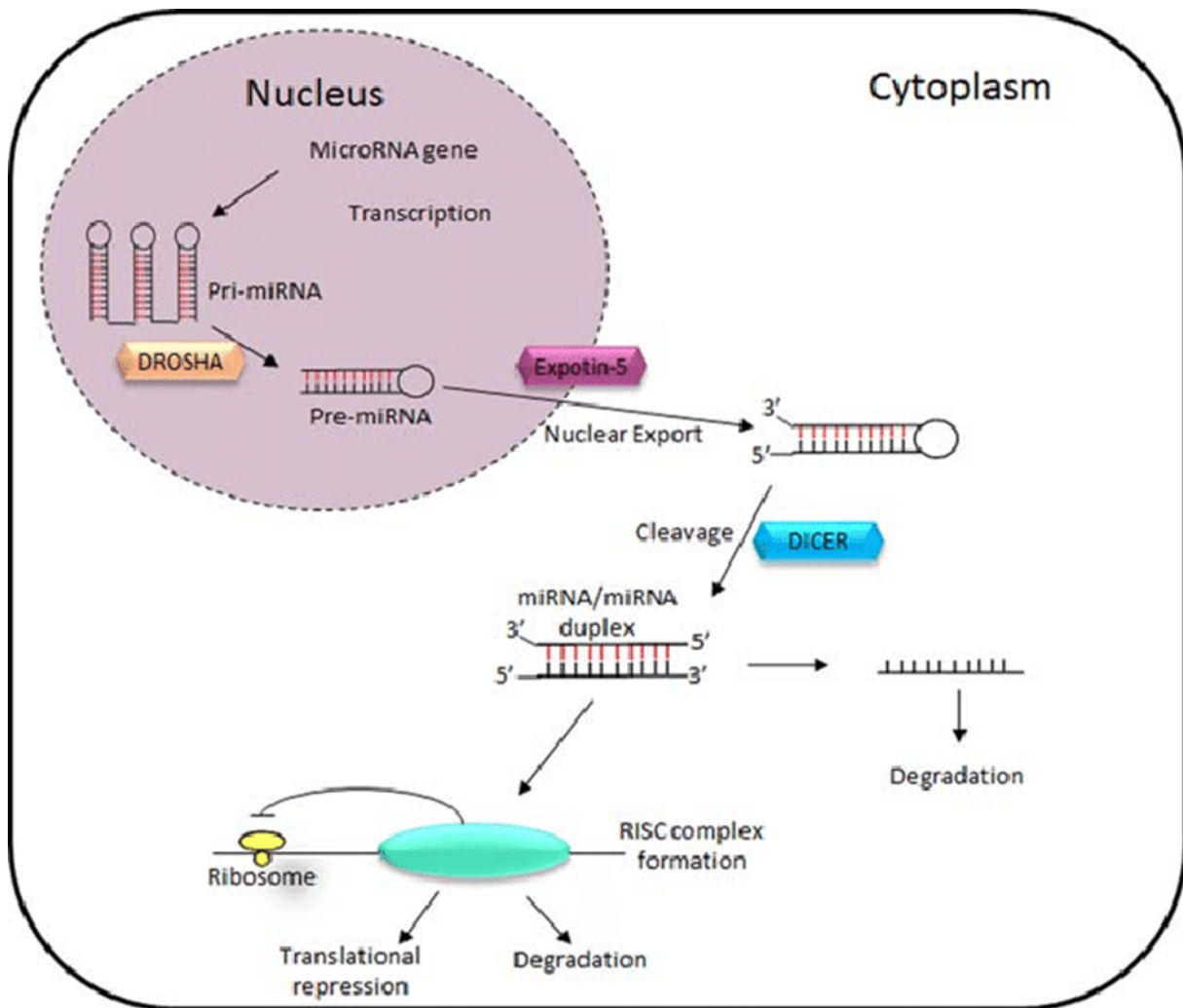
Primary mutations leading to the development of T-ALL are often chromosomal translocations involving the T-cell receptor genes A/B/D where the *TCR* enhancer is fused to different transcription factors<sup>64</sup>. Other driving mutations occur in the *NOTCH1* gene ADDIN EN.CITE <sup>65</sup>.

A T-ALL in infancy (under one year of age) ranks among the extremely rare and hard to cure subtypes of leukemia. Pediatric T-ALL patients of older age seem to have a better outcome

than infants ADDIN EN.CITE <sup>66</sup>. It is not clear, weather iT-ALL (infant T-ALL) represents a distinct disease compared to childhood ALL.

## **1.2 miRNAs in Leukemia**

Many diseases, like leukemia, arise by disturbed cellular functions like apoptosis, cell-cycle control or growth signaling pathways. Many factors are important to maintain cellular physiology, like the epigenetic regulation performed by microRNAs (miRNAs, miRs). miRNAs were first described in 1993 by Lee et al.<sup>67</sup>. miRNAs are small non-coding RNAs ( 18-22 nucleotides), which regulate gene expression by targeting the corresponding messenger RNA (mRNA). mRNA and miRNA form a duplex, on the basis of the complementarity between parts of the mRNA untranslated region and the miRNA seed region, which is mostly located at positions 2-7 of the miRNA 5'-end. Thereby the mRNA is degraded or translational inhibited. Primary-miRNAs (pri-miRNAs) are transcribed from DNA by RNA polymerase II. Further on the pri-miRNA is processed by RNase Drosha into precursor miRNAs (pre-miRNAs) which are translocated into the cytoplasm via the transporter protein Exportin-5. Pre-miRNAs consist of 60-90 nucleotides which build a secondary stem loop-structure. The pre-miRNA is processed into its functional form, the mature miRNA, by cleaving off the loop structure by the RNase Dicer. The mature double-stranded miRNA interacts with argonaute proteins to build the RNA induced silencing complex (RISC), which induced specific mRNA degradation or translational repression (Figure 5) reviewed by Nelson et al, 2003<sup>68</sup>.

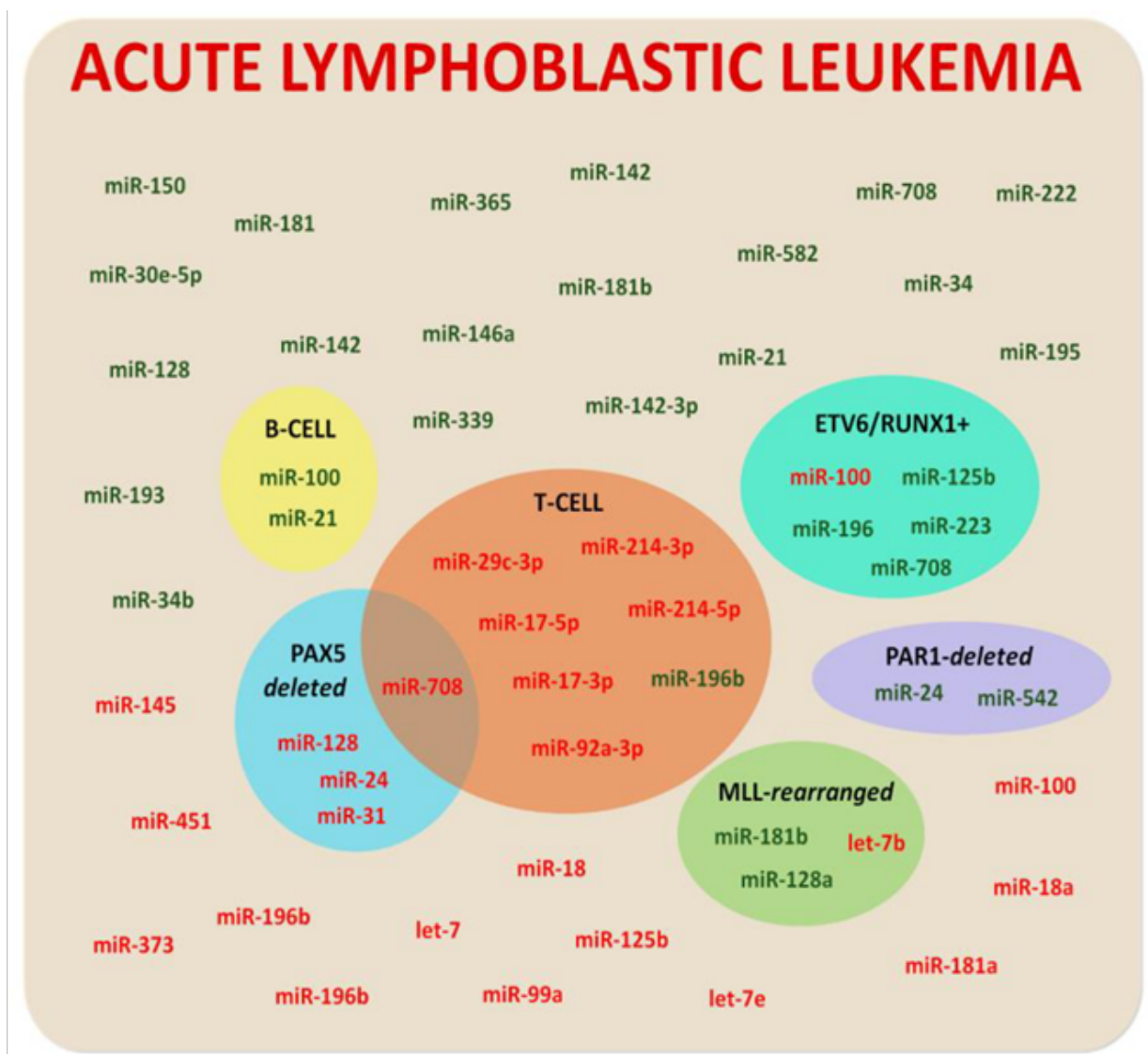


Taken from Ballantyne et al, 2016

**Figure 5: Pathway of miRNA processing.** Pri-miRNAs are transcribed from DNA and cleaved by Droscha, to build pre-miRNA. These pre-miRNA molecules are translocated from the nucleus into the cytoplasm by Exportin-5 for further processing. By cleaving of the stem loop by Dicer, the mature miRNA arises. Mature miRNA builds a complex with Argonaute proteins, to build the RISC, which inhibits mRNA translation or performs mRNA degradation [ADDIN EN.CITE](#) <sup>69</sup>.

Over 50% of miRNA genes are located at fragile sites and cancer-associated regions [ADDIN EN.CITE](#) <sup>70</sup>. Profiling of miRNAs showed a distinct pattern in leukemic cells different from healthy cells, hinting an influence of miRNA expression in tumor development and progression [ADDIN EN.CITE](#) <sup>71</sup>. By influencing proliferation, metastasis, apoptosis, angiogenesis or drug resistance, many miRNAs can act as oncomiRs or tumor suppressor miRNAs [ADDIN EN.CITE](#) <sup>72,73</sup>. Analyzing the expression pattern of miRNAs in pediatric leukemia revealed a wide spectrum of miRNAs differentially expressed between leukemic and healthy cells as well as between different leukemic subtypes (Figure 6) [ADDIN EN.CITE](#) <sup>74,75</sup>. Differences in miRNA expression between the leukemic subtypes *TCF3/PBX1* and *TCF3/HLF* as well as between infant and childhood T-cell ALL are still unknown, but

important for clinical use. Because of the unique miRNA expression pattern in leukemic subtypes, miRNAs can be used as biomarkers in diagnosis [ADDIN EN.CITE 76](#) or for potential therapeutic options. Until now two miRNA therapy strategies entered clinical trials. Targeting miRNA-122 in patients with chronic hepatitis C virus (HCV) by a modified oligonucleotide (serving as an antisense oligonucleotide) showed reductions in HCV RNA levels [ADDIN EN.CITE 77](#). Treating patients with MRX34®, a liposomal miR-34a mimic, which acts as a tumor-suppressor miRNA, showed antitumor-activity in patients with refractory advanced solid tumors [ADDIN EN.CITE 78](#). miRNA therapy for hematological diseases are not yet in use.



Taken from Carvalho de Oliveira, 2018

**Figure 6: Differentially expressed miRNAs in pediatric ALL.** Green: upregulated miRNAs, red: downregulated miRNAs in childhood ALL and certain ALL subgroups [ADDIN EN.CITE 79](#).

### **1.3 Aim of the Thesis**

t(17;19) B-ALL and infant T-ALL are two hard to cure types of pediatric leukemia. The identification of targetable genes or miRNAs in these leukemic subtypes could greatly impact on the development of novel therapy strategies.

In this thesis, by using next generation sequencing, miRNA expression landscapes of both types of leukemia should be analyzed and compared to similar leukemia subtypes with a better treatment outcome. To this end, t(17;19) B-ALL should be compared to t(1;19) B-ALL and infant T-ALL to childhood T-ALL.

The aim of the comparison between t(1;19)- and t(17;19)-positive ALL was to find novel and mutated miRNAs in one of the cohorts, whose expression could explain the clinical differences between both B-ALL subtypes. Additionally, there was a focus on differentially expressed miRNAs and mRNAs, which could act as mRNA-miRNA target pairs and have a potential impact on disease development. Also, changes of gene expression should be analyzed after treatment of a model of t(17;19) B-ALL with the Histone-Deacetylase (HDAC) inhibitor KSK, a promising novel drug for treatment of the t(17;19)-positive leukemic subtype.

Likewise, differences in infant and childhood T-ALL should be analyzed on a genetic and epigenetic level. Exomes of iT-ALL patients should be sequenced to identify mutations in tumor relevant genes. Transcriptomes and miRNomes of infant and childhood T-ALL patients should be sequenced to identify differences in expression levels. Computational evaluation should be used to analyze the impact of aberrant gene expression patterns.

## 2. Materials

### 2.1 Samples

#### 2.1.1 Patient Samples

Blood or bone marrow samples were obtained from pediatric ALL patients of the international Berlin-Frankfurt-Münster (i-BFM) study group. Mononuclear cells of blood or bone marrow from initial diagnosis and remission (if possible) of patients diagnosed with leukemia were isolated. Written informed consent was obtained from the parents and/or legal guardians. Analysis of patient material with B-ALL was part of the treatment study approval as add-on studies approved by the ethic committee of the medical faculty of Christian-Albert-University in Kiel and all participating treatment centers. Analysis of patient material with T-ALL was approved by local ethics committee (study numbers 3432, 4769, 5036).

#### 2.1.2 Human Cell Lines

**Cell line: HAL-01**

DMSZ no.: ACC 610

Species: human (*Homo sapiens*)

Cell type: B-cell precursor leukemia

Origin: HAL-01 was established in 1990 from peripheral blood (PB) of a 17-year-old woman suffering from ALL. The cell line contains t(17;19)(q22;p13) leading to the *TCF3/HLF* fusion gene ADDIN EN.CITE<sup>80,81</sup>.

**Cell line: HSB-2**

DMSZ no.: ACC 435

Species: human (*Homo sapiens*)

Cell type: T-cell leukemia



Origin: HSB-2 was established in 1966 from PB of an 11-year-old boy whose lymphosarcoma progressed to ALL. The cell line contains t(1;7)(p34;q34) leading to Lymphocyte Cell-

Specific Protein-Tyrosine Kinase (*LCK*)-T-cell receptor beta (*TCR-B*) fusion gene. It also contains del(1)(p32) leading to the *STIL/TAL1* fusion gene. ADDIN EN.CITE <sup>82,83</sup>

**Cell line: REH**

DMSZ no.: ACC 22

Species: human (*Homo sapiens*)

Cell type: B-cell precursor leukemia

Origin: REH was established in 1973 from peripheral blood of a 15-year-old North African girl with ALL at relapse. The cells contain t(12;21) leading to ETS Translocation Variant 6/ Runt-related transcription factor 1 (*ETV6/RUNX1*) fusion gene ADDIN EN.CITE <sup>81,84</sup>.

**Cell line: SEM**

DMSZ no.: ACC 546

Species: human (*Homo sapiens*)

Cell type: B-cell precursor leukemia

Origin: SEM was established in 1990 from peripheral blood of a 5-year-old girl with ALL at relapse. The cells contain t(4;11) leading to Mixed Lymphoid Leukemia fusion gene (*KMT2A/AFF1*) ADDIN EN.CITE <sup>85-87</sup>.

**Cell line: 697**

DMSZ no.: ACC 42

Species: human (*Homo sapiens*)

Cell type: B-cell precursor leukemia

Origin: 697 was established in 1979 from peripheral blood of a 12-year-old boy with ALL at relapse. The cells contain t(1;19) leading to *TCF3/PBX1* fusion gene<sup>88</sup>.

**Cell line: HEK293T**

DMSZ no.: ACC 635

Species: human (*Homo sapiens*)

Cell type: embryonal kidney

Origin: HEK293T are a highly transfectable derivate of the cell line 293, which was established from human embryonal kidney cells. The cells contain a mutant of SV-40 large T-antigen ADDIN EN.CITE <sup>89-91</sup>.

All information regarding used cell lines are according to DMSZ.

## 2.2 Chemicals

Product	Company	Order number
Ampicillin	Sigma-Aldrich, St. Louis, MO, USA	69-53-4
AMPure XP Beads	Beckmann Coulter, Krefeld, Germany	A83881
1-Methylimidazole	Sigma-Aldrich, St. Louis, MO, USA	A3678
1,4.dithio-DL-threitol (DTT)	Roth, Karlsruhe, Germany	6908.2
Ammoniumpersulfat (APS)	Merck, Darmstadt, Germany	7727-54-0
Boric Acid	Merck, Darmstadt, Germany	005-007-00-2
Bromophenol blue	Merck, Darmstadt, Germany	115-39-9
Chloroform	Merck, Darmstadt, Germany	602-006-00-4
Complete, EDTA-free Protease Inhibitor Cocktail Tablets	Roche Diagnostics, Mannheim, Germany	11873580001
EDTA (disodium salt)	Sigma-Aldrich, St. Louis, MO, USA	E5134
Ethanol	VWR, Langenfeld, Germany	20821.330
Ethidium bromide 10 mg/mL	Sigma Aldrich	E1510
Ficoll type 400	Sigma-Aldrich, St. Louis, MO, USA	17-0300-15
Formamide deionized	ThermoFischer, Waltham, Massachusetts, USA	AM9342
GlycoBlue 300 µL (15mg/mL)	Ambion, Huntington, Cambridgeshire, UK	AM9515
Isoamylalkohol	Merck, Darmstadt, Germany	100979
Isopropyl alcohol	VWR, Langenfeld, Germany	20842.300

<b>Lipofectamine 2000</b>	Invitrogen, Carlsbad, CA, USA	11668-019
<b>microRNA Marker</b>	NEB, Frankfurt a. M., Germany	N2102S
<b>Natriumdihydrogenphosphat</b>	Roth, Karlsruhe, Germany	T879.1
<b>NEBuffer™ 3.1</b>	NEB, Frankfurt a. M., Germany	B7203S
<b>Phenol</b>	Merck, Darmstadt, Germany	100206
<b>Phenol acid</b>	Sigma-Aldrich, St. Louis, MO, USA	P4628-400ML
<b>Phusion HF Buffer 5x</b>	NEB, Frankfurt a. M., Germany	B0518S
<b>Quick Start Bradford Dye Reagent</b>	Bio-Rad, München, Germany	500-0205
<b>Sodiumdodecylsulfate (SDS) Pellets</b>	Roth, Karlsruhe, Germany	CN30.3
<b>RNAClean XP Beads</b>	Beckmann Coulter, Krefeld, Germany	A63987
<b>SequaGel-UreaGel Buffer</b>	National Diagnostics, Atlanta, GA, USA	EC-835
<b>SequaGel-UreaGel Concentrate</b>	National Diagnostics, Atlanta, GA, USA	EC-830
<b>SequaGel-UreaGel Diluent</b>	National Diagnostics, Atlanta, GA, USA	EC-840
<b>3m Sodium acetate buffer solution, pH 5.2</b>	Sigma-Aldrich, St. Louis, MO, USA	S7899
<b>SYBR Green PCR Master Mix</b>	ABI, Carlsbad, CA, USA	4309155
<b>T4 DNA Ligase Reaction Buffer</b>	NEB, Frankfurt a. M., Germany	B0202
<b>TEMED</b>	Bio-Rad, München, Germany	110-18-9
<b>Tris base</b>	Sigma-Aldrich, St. Louis, MO, USA	T-1053
<b>TRIzol reagent</b>	Invitrogen, Carlsbad, CA, USA	15596-018
<b>Xylene cyanol FF</b>	Sigma-Aldrich, St. Louis, MO, USA	X-4126

## 2.3 Nucleic Acids

### 2.3.1 Oligonucleotides

<b>Name</b>	<b>Company</b>	<b>Length</b>	<b>Sequence 5' -&gt; 3'</b>
<b>NOTCH2 SNP for</b>	MWG, Ebersberg, Germany	20 nt	CAAGATGCCAAACAACCAGA
<b>NOTCH2 SNP rev</b>	MWG, Ebersberg, Germany	20 nt	TTTGTTTCCTCATGGCAGTG
<b>NOTCH3 for</b>	MWG, Ebersberg, Germany	20 nt	GTCAGGGGTCAGAGGAGACA
<b>NOTCH3 rev</b>	MWG, Ebersberg, Germany	20 nt	AGGTCAGGAGTTCAACAGCA
<b>IL7R for</b>	MWG, Ebersberg, Germany	20 nt	TGGCTATGCTCAAAATGGTG
<b>IL7R rev</b>	MWG, Ebersberg, Germany	22 nt	TGAATCCAGTTTGATCTCCTG A
<b>PTEN1for</b>	MWG, Ebersberg, Germany	20 nt	TCTTTTCAGGCAGGTGTCAA
<b>PTEN1rev</b>	MWG, Ebersberg, Germany	20 nt	TCTGCAGGAAATCCCATAGC
<b>PTEN2for</b>	MWG, Ebersberg, Germany	20 nt	ACCGCCAAATTTAATTGCAG
<b>PTEN2rev</b>	MWG, Ebersberg, Germany	20 nt	GCTTTCTCCCTGTGATTGCT
<b>PTEN3for</b>	MWG, Ebersberg, Germany	22 nt	TCGTTTTTGACAGTTTGACAG T
<b>PTEN3rev</b>	MWG, Ebersberg, Germany	20 nt	CACCAATGCCAGAGTAAGCA
<b>PTEN4for</b>	MWG, Ebersberg, Germany	20 nt	TCCGAAGGGTTTTGCTACAT
<b>PTEN4rev</b>	MWG, Ebersberg, Germany	20 nt	TCAAGCCATTCTTTGTTGA
<b>PTEN5for</b>	MWG, Ebersberg, Germany	21 nt	CCACCCTTTTGACCTTACACA
<b>PTEN5rev</b>	MWG, Ebersberg, Germany	20 nt	TGCAGTCTGGGCATATCAAA

<b>PTEN6for</b>	MWG, Ebersberg, Germany	20 nt	GCCTCATCCCAATCAGATGT
<b>PTEN6rev</b>	MWG, Ebersberg, Germany	23 nt	TGGACTTTTTCAGGACTAGAA CG
<b>PTEN7for</b>	MWG, Ebersberg, Germany	22 nt	CACCTTTAGGATTTTCTGCCT A
<b>PTEN7rev</b>	MWG, Ebersberg, Germany	20 nt	TGCCAACTTTGGTTTAATGC
<b>KRAS for</b>	MWG, Ebersberg, Germany	20 nt	GGCACTCAAAGGAAAAATGC
<b>KRAS rev</b>	MWG, Ebersberg, Germany	23 nt	TGCATTGAGAACTGAATAGC TG
<b>TLR2delfor</b>	MWG, Ebersberg, Germany	20 nt	GTAATTCCGGATGGTTGTGC
<b>TLR2delrev</b>	MWG, Ebersberg, Germany	20 nt	CTTCCTTGGAGAGGCTGATG
<b>TLR2SNPfor</b>	MWG, Ebersberg, Germany	20 nt	TCCATTGAAAAGAGCCACAA
<b>TLR2SNPprev</b>	MWG, Ebersberg, Germany	21 nt	TCCTCAAATGACGGTACATCC
<b>TLR4for</b>	MWG, Ebersberg, Germany	20 nt	TCAGAACTGCTCGGTCAGA
<b>TLR4rev</b>	MWG, Ebersberg, Germany	20 nt	GCCCCTGTTAGCACTCAAAA
<b>Ren-Luc for</b>	MWG, Ebersberg, Germany	22 nt	CGAGGTCCGAAGACTCATTTA G
<b>SOX4 for</b>	MWG, Ebersberg, Germany	24 nt	ACGCCTCGAGGGTGTTAATCG GTG
<b>SOX4 rev</b>	MWG, Ebersberg, Germany	31 nt	ATAGTTTACGGCAGTGAAC TGACATG
<b>Actin for</b>	MWG, Ebersberg, Germany	20 nt	AGAGCTACGAGCTGCCTGAC
<b>Actin rev</b>	MWG, Ebersberg, Germany	20 nt	AGCACTGTGTTGGCGTACAG
<b>Oligo(dT)</b>	MWG, Ebersberg, Germany	20 nt	TTTTTTTTTTTTTTTTTTTT
<b>miR-199aTop</b>	MWG, Ebersberg, Germany	68 nt	TGCTGCCCAGTGTTTCAGACTA CCTGTTTCGTTTTGGCCACTGA CTGACGAACAGGTTCTGAACA CTGGGG
<b>miR-199aBottom</b>	MWG, Ebersberg, Germany	68 nt	CCTGCCCAGTGTTTCAGAACCT GTTTCGTCAGTCAGTGGCCAAA ACGAACAGGTAGTCTGAACA CTGGGC
<b>miR-199bTop</b>	MWG, Ebersberg, Germany	68 nt	TGCTGCCCAGTGTTTACTACTA TCTGTTTCGTTTTGGCCACTGA CTGACGAACAGATTCTAAACA CTGGG
<b>miR-199bBottom</b>	MWG, Ebersberg, Germany	68 nt	CCTGCCCAGTGTTTACTAATCT GTTTCGTCAGTCAGTGGCCAAA ACGAACAGATAGTCTAAACA CTGGGC
	MWG, Ebersberg, Germany	20 nt	TCCAAGCTGGCTAGTTAAG

<b>miRNA forward sequencing primer</b>			
<b>miRNA reversed sequencing primer</b>	Thermo Fisher Scientific, Waltham, MA, USA	20 nt	CTCTAGATCAACCACTTTGT

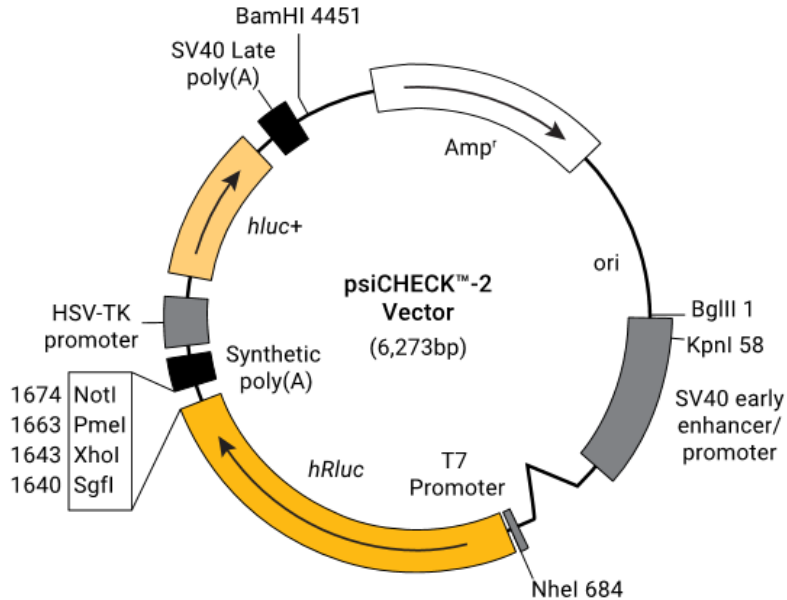
### 2.3.2 Nucleotides

Identifier	Company	Order number
dATP, PCR Grade 100 nM	Qiagen, Hilden, Germany	1039397
dCTP, PCR Grade, 100 nM	Qiagen, Hilden, Germany	1039396
dGTP, PCR Grade, 100 nM	Qiagen, Hilden, Germany	1039395
dTTP, PCR Grade, 100 nM	Qiagen, Hilden, Germany	1039394

### 2.3.3 Vectors and Vector Cards

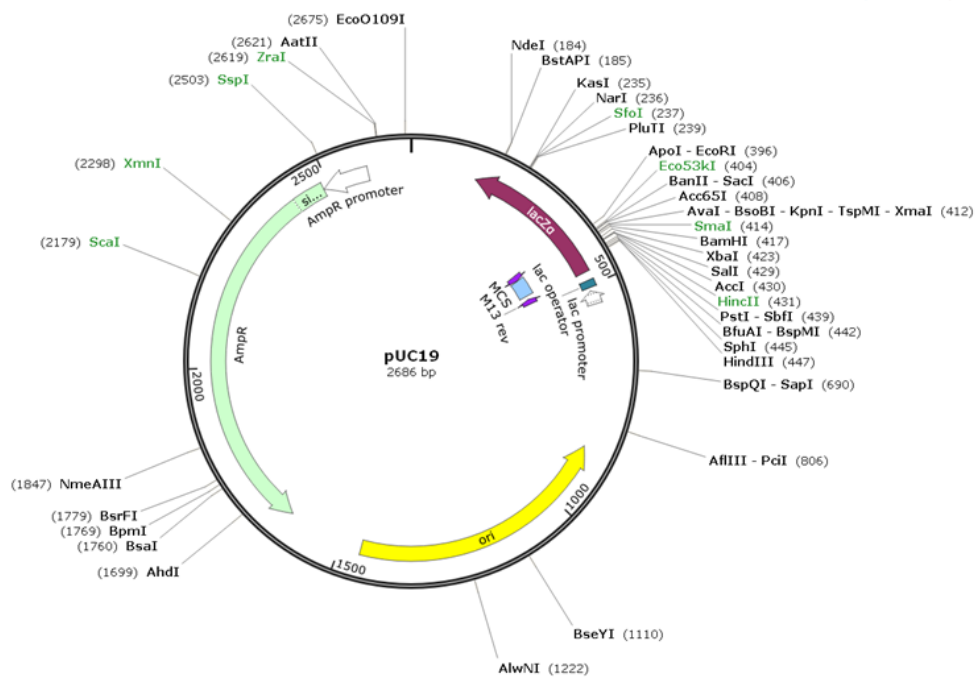
Identifier	Company	Order number
psiCHECK™-2 Vector	Promega, Fitchburg, WI, USA	C8021
pUC19	Invitrogen, Carlsbad, CA, USA	SD0061

#### psiCHECK™-2 Vector:

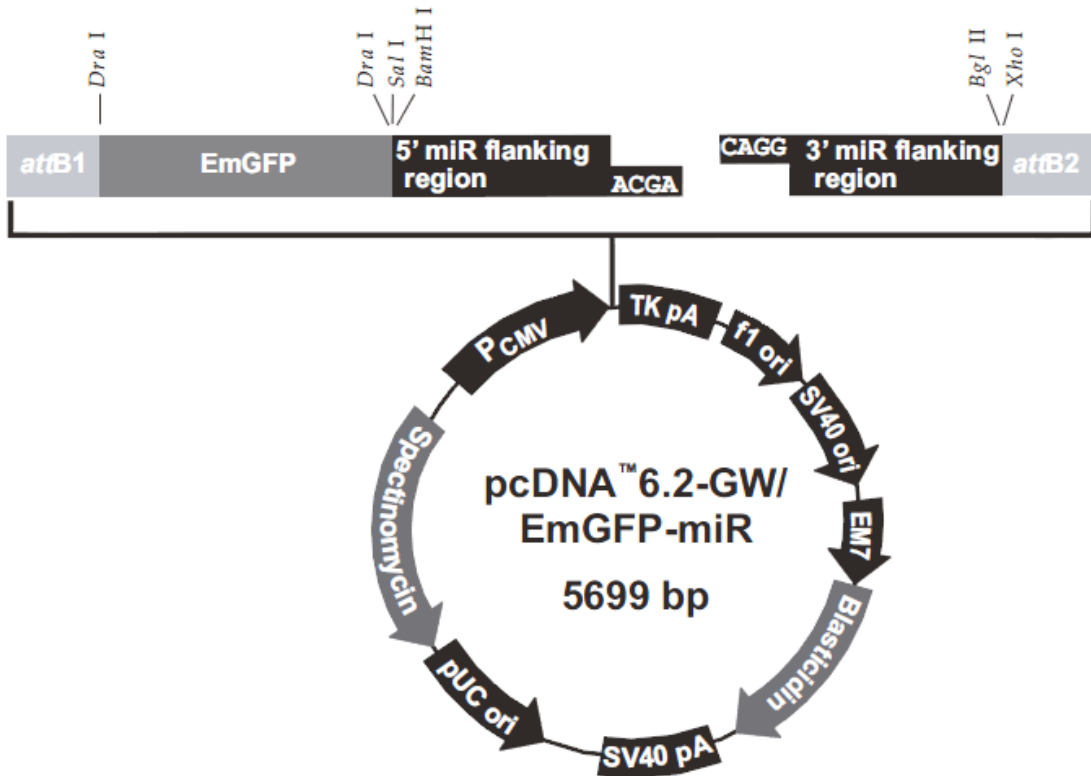


**Figure 7: psiCHECK-2 vector card.** Amp: gene for ampicillin resistance; ori: origin of replication; SV40: Simian-Virus 40; T7: Bacteriophage T7; hRluc: Renilla luciferase reporter gene; HSV: Herpes Simplex Virus; TK: Thymidine Kinase; hLuc: firefly luciferase gene. Vector and vector cards were received from promega.

# pUC19:



**Figure 8: pUC19 vector card.** lacZα: gene for β-galactosidase; MCS: multiple cloning site; ori: origin of replication; M13: Bacteriophage M13; AmpR: gene for ampicillin resistance. Vector card was taken from transomic.com.



**Figure 9: pcDNA6.2-GW/EmGFP-miR vector card.** pCMV: cytomegalovirus promotor; attB: bacterial attachment site; EmGFP: Emerald green fluorescent protein; TK pA: thymidine kinase polyadenylation signal sequence; fl ori: fl origin of replication; SV40 ori: Simian Virus 40 origin of replication; SV40 pA: Simian Virus 40 polyadenylation signal; EM7: synthetic modification of T7 promoter; Blastidicin: eukaryotic selection marker; Spectinomycin: prokaryotic selection marker.

### 2.3.4 siRNAs (Small Interfering RNAs)

Name	Company	Length	Sequence
<b>199a-5p siRNA</b>	MWG, Ebersberg, Germany	23 nt	CCCAGUGUUCAGACUACCUGUUC(dTdT)
<b>Renilla Luciferase siRNA</b>	MWG, Ebersberg, Germany	19 nt	GGCCUUUCACUAAACUCCUAC(dTdT)

### 2.3.5 Northern Blot Probes (DNA):

Name	Company	Length	Sequence
<b>miRNA-Marker ss-Probe (S1328S)</b>	NEB, Frankfurt a. M., Germany	21 nt	ATCTCAACCAGCCACTG
<b>Novel miR 1</b>	MWG, Ebersberg, Germany	18 nt	AGGTGGAGGAGCCGGGGG
<b>Novel miR 2</b>	MWG, Ebersberg, Germany	22 nt	TCGGGTGATCTGCCTGCCTCGG
<b>Novel miR 3</b>	MWG, Ebersberg, Germany	23 nt	AGCCTGGGAGGTCGAGTCTGCAG
<b>Novel miR 4</b>	MWG, Ebersberg, Germany	18 nt	ATTCAGAGTCCAGAGTGC
<b>Novel miR 5</b>	MWG, Ebersberg, Germany	23 nt	CCAGGATGGAGGGAACAGGGCTG
<b>Novel miR 6</b>	MWG, Ebersberg, Germany	22 nt	TGAGTGCCCCTTTAAATCCTGA
<b>Novel miR 7</b>	MWG, Ebersberg, Germany	24 nt	TCTCAGGGGAGCTGGCAGAACCAG
<b>Novel miR 8</b>	MWG, Ebersberg, Germany	22 nt	ACGCGAGCCTCGAGACGTCTGA
<b>Novel miR 9</b>	MWG, Ebersberg, Germany	23 nt	CTCAGGGGAGCTGGCAGAACCAG
<b>Novel miR 10</b>	MWG, Ebersberg, Germany	21 nt	GAAGGCCCGAACCCTACCGGA
<b>Novel miR 11</b>	MWG, Ebersberg, Germany	18 nt	CCGCGCCCGCCCGCCCGG
<b>Novel miR 12</b>	MWG, Ebersberg, Germany	23 nt	TCCAGGGAGCCCAGCCAGGCCTG
<b>Novel miR 13</b>	MWG, Ebersberg, Germany	22 nt	CCAATCCTTCAGATCCTCCTCA
<b>Novel miR 14</b>	MWG, Ebersberg, Germany	20 nt	CAGAAGCCGAGGGCGCATGT
<b>Novel miR 15</b>	MWG, Ebersberg, Germany	22 nt	AGCCTGGGAGGTCGAGTCTGCA
<b>Novel miR 16</b>	MWG, Ebersberg, Germany	21 nt	ATTTCCCTCACAGTCTGCCAC

<b>Novel miR 17</b>	MWG, Ebersberg, Germany	18 nt	AAGCCTACAGCACCCGCT
<b>Novel miR 18</b>	MWG, Ebersberg, Germany	22 nt	TGTACGGAGGGGAACGATCAGA
<b>miR-21-5p probe</b>	MWG, Ebersberg, Germany	22 nt	TCAACATCAGTCTGATAAGCTA

## 2.4 Enzymes

<b>Identifier</b>	<b>Company</b>	<b>Odernumber</b>
<b>BamHI</b>	NEB, Frankfurt a. M., Germany	R0136
<b>HindIII</b>	NEB, Frankfurt a. M., Germany	R0104
<b>NcoI</b>	NEB, Frankfurt a. M., Germany	R0193
<b>NotI</b>	NEB, Frankfurt a. M., Germany	R0189
<b>Phusion High-Fidelity Polymerase</b>	NEB, Frankfurt a. M., Germany	M0530L
<b>RNA Ligase 2 Rnl2 (1-249) K227Q 1µg/µL</b>	NEB, Frankfurt a. M., Germany	M0236L
<b>SuperScriptII Reverse Transcriptase 10,000 U</b>	Invitrogen, Carlsbad, CA, USA	18064014
<b>SuperScriptIII Reverse Transcriptase 10,000 U</b>	Invitrogen, Carlsbad, CA, USA	18080044
<b>T4 DNA ligase</b>	NEB, Frankfurt a. M., Germany	M0202
<b>XhoI</b>	NEB, Frankfurt a. M., Germany	R0146

## 2.5 Bacteria

<b>Identifier</b>	<b>Company</b>	<b>Odernumber</b>
<b>One Shot™ Top 10 Competent E. coli</b>	Invitrogen, Carlsbad, CA, USA	C404006

## 2.6 Culture Media, HDAC-Inhibitor and Buffer

### 2.6.1 Media and Additives for Cell Cultivation

<b>Identifier</b>	<b>Company</b>	<b>Odernumber</b>
<b>DMEM-GlutaMAX™-I</b>	Gibco, Invitrogen, Carlsbad, CA, USA	31966-021
<b>Dulbecco's Phosphate Buffered Saline (PBS)</b>	Sigma-Aldrich, St. Louis, MO, USA	D8537
<b>Fetal Bovine Serum (FBS) heat inactivated</b>	Biowest, Nuaille, France	S181H-500
<b>IMDM</b>	Gibco, Invitrogen, Carlsbad, CA, USA	31980030
<b>LB-Agar</b>	Roth, Karlsruhe, Germany	X965.2
<b>LB-Medium</b>	Roth, Karlsruhe, Germany	X964.3
<b>L-Glutamine</b>	Gibco, Invitrogen, Carlsbad, CA, USA	25030081
<b>Penicillin-Streptomycin 10,000 U/mL</b>	Gibco, Invitrogen, Carlsbad, CA, USA	15140122



<b>RPMI 1640 Medium, GlutaMAX</b>	Gibco, Invitrogen, Carlsbad, CA, USA	61870-010
<b>Trypsin-EDTA 0.05%</b>	Gibco, Invitrogen, Carlsbad, CA, USA	25300-054

## 2.6.2 HDAC (Histone Deacetylase) Inhibitors

<b>KSK</b>	Institute for pharmaceutical and medical chemistry – Heinrich-Heine-University	
<b>Ricolinostat</b>	Selleckchem	ACY-1215

## 2.7 Kits

<b>Identifier</b>	<b>Company</b>	<b>Ordernumber</b>
<b>Agilent DNA 1000 Kit</b>	Agilent, Santa Clara, CA, USA	5067-1504
<b>Agilent High Sensitivity DNA Kit</b>	Agilent, Santa Clara, CA, USA	5067-4626
<b>Agilent RNA 6000 Nano Kit</b>	Agilent, Santa Clara, CA, USA	5067-1511
<b>AllPrep DNA/RNA/miRNA Universal Kit</b>	Qiagen, Hilden, Germany	80224
<b>BLOCK-iT Pol II miR RNAi Expression Vector Kit with EmGFP</b>	Thermo Fisher Scientific, Waltham, MA, USA	K4939-00
<b>Fast Plasmid Mini-Kit</b>	5 Prime, Hilden, Germany	2300010
<b>HighSpeed Plasmid Maxi Kit</b>	Qiagen, Hilden, Germany	12663
<b>QIAquick Gel Extraction Kit</b>	Qiagen, Hilden, Germany	28706
<b>REPLI-g Ultra Fast mini Kit</b>	Qiagen, Hilden, Germany	150033
<b>TruSeq SBS Kit v3-HS</b>	Illumina, San Diego, CA, USA	FC-401-3002
<b>TruSeq Small RNA Library Prep Kit</b>	Illumina, San Diego, CA, USA	RS-200-0012
<b>TruSeq SR Cluster Kit v3-cBot-HS</b>	Illumina, San Diego, CA, USA	GD-401-3001
<b>TruSeq Stranded Total RNA Prep Kit with Ribo-Zero Gold</b>	Illumina, San Diego, CA, USA	RS-122-2301

## 2.8 Software

<b>Identifier</b>	<b>Company/available at</b>
<b>2100 BioAnalyzer Expert</b>	ABI, Carlsbad, CA, USA
<b>ApE – A Plasmid Editor</b>	<a href="http://biologylabs.utah.edu/jorgensen/wayned/ape/">http://biologylabs.utah.edu/jorgensen/wayned/ape/</a>
<b>Burrows-Wheeler Aligner (BWA)</b>	<a href="http://bio-bwa.sourceforge.net">http://bio-bwa.sourceforge.net</a>
<b>cBot software v1.5</b>	Illumina, San Diego, CA, USA
<b>Cutadapt</b>	<a href="https://github.com/marcelm/cutadapt">https://github.com/marcelm/cutadapt</a>

<b>dbSNP</b>	Broad Institute, Cambridge, MA, USA
<b>edgeR</b>	<a href="https://bioconductor.org/packages/release/bioc/html/edgeR.html">https://bioconductor.org/packages/release/bioc/html/edgeR.html</a>
<b>GATK</b>	Broad Institute, Cambridge, MA, USA
<b>HapMap3</b>	Broad Institute, Cambridge, MA, USA
<b>HTseq</b>	<a href="https://github.com/simon-anders/htseq">https://github.com/simon-anders/htseq</a>
<b>Ingenuity Pathway Analysis</b>	Qiagen, Hilden, Germany
<b>Intas GDS</b>	Intas, Göttingen, Germany
<b>miRanda August 2010</b>	<a href="http://www.microrna.org/microrna/getDownloads.do">http://www.microrna.org/microrna/getDownloads.do</a>
<b>miRDB</b>	<a href="http://mirdb.org/download.html">http://mirdb.org/download.html</a>
<b>miRTarBase</b>	<a href="http://mirtarbase.mbc.nctu.edu.tw/php/download.php">http://mirtarbase.mbc.nctu.edu.tw/php/download.php</a>
<b>MuTect</b>	<a href="http://archive.broadinstitute.org/cancer/cga/mutect">http://archive.broadinstitute.org/cancer/cga/mutect</a>
<b>OligoAnalyzer 3.1</b>	IDT, Coralville, USA – <a href="https://eu.idtdna.com/calc/analyzer">https://eu.idtdna.com/calc/analyzer</a>
<b>OmniArray</b>	Broad Institute, Cambridge, MA, USA
<b>R2 MegaSampler</b>	<a href="https://hgserver1.amc.nl/cgi-bin/r2/main.cgi">https://hgserver1.amc.nl/cgi-bin/r2/main.cgi</a>
<b>SAMtools</b>	<a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a>
<b>seqtk</b>	<a href="https://github.com/lh3/seqtk">https://github.com/lh3/seqtk</a>
<b>Sequencher 4.8</b>	Gene Codes, Ann Arbor, US
<b>Sequencing Analysis Viewer 1.8.46</b>	Illumina, San Diego, CA, USA
<b>STAR</b>	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
<b>TarBase V7</b>	<a href="http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=tarbasev8%2Findex">http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=tarbasev8%2Findex</a>
<b>TargetScan V6.2</b>	<a href="http://www.targetscan.org/vert_61/">www.targetscan.org/vert_61/</a>

## 2.8 Hardware

<b>Hardware</b>	<b>Company</b>
<b>Agilent 2100 Bioanalyzer</b>	ABI, Carlsbad, CA, USA
<b>cBot</b>	Illumina, San Diego, CA, USA
<b>Centrifuge 5403</b>	Eppendorf, Hamburg, Germany
<b>Centrifuge 5417R</b>	Eppendorf, Hamburg, Germany
<b>CFX Connect real-time system</b>	Bio-Rad, München, Germany
<b>HeroLab UVT 2035 UV lamp</b>	HeroLab, Wiesloch, Germany
<b>HiSeq 2500</b>	Illumina, San Diego, CA, USA
<b>Intas UV lamp</b>	Intas, Göttingen, Germany
<b>Luminoskan Ascent Microplate Luminometer</b>	Thermofischer, Waltham, Massachusetts, USA
<b>Milli-Q Integral 15</b>	Millipore, Billerica, MA, USA
<b>Nanodrop 1000</b>	Peqlab, Erlangen, Germany
<b>Neubauer counting chamber</b>	Brand, Wertheim, Germany
<b>Thermomixer Comfort</b>	Eppendorf, Hamburg, Germany

<b>Trans-Blot Semi-Dry (SD) Transfer Blot</b>	Bio-Rad, München, Germany
<b>Vi-Cell XR</b>	Beckmann Coulter, Krefeld, Germany
<b>Vortex2 Genie</b>	Scientific Industries, NY, USA

## 2.9 Other Materials

<b>Identifier</b>	<b>Company</b>	<b>Odernumber</b>
<b>Amersham Protran Supported 0.45 NC 300mmx4m roll</b>	GE Healthcare Life Sciences, Pittsburgh, PA, USA	10600016
<b>Corning Costar Spin-X</b>	Sigma-Aldrich, St. Louis, MO, USA	CLS8163
<b>Cryo freezing container, 18 vials</b>	Nalgene, NY, USA	5100-0001
<b>Culture Dish</b>	Thermofischer, Waltham, Massachusetts, USA	140-675
<b>Hard-Shell PCR plates 96-well, thin wall</b>	Bio-Rad, München, Germany	HSP9601
<b>Microseal 'B' seal</b>	Bio-Rad, München, Germany	MSB1001

## **3. Methods**

### **3.1 Cell Culture**

#### **3.1.1 Cultivation of Human Suspension Cell Lines**

Cell lines 697 and REH were cultivated in RPMI 1640 with 20% (v/v) FBS and 1% (v/v) 10,000 U/mL Penicillin-Streptomycin. Cell lines HAL-01 and HSB-2 were cultivated in RPMI 1640 with 10% (v/v) FBS and 1% (v/v) 10,000 U/mL Penicillin-Streptomycin. Cell line SEM was cultivated in IMDM with 10% (v/v) FBS, 1% (v/v) 10,000 U/mL Penicillin-Streptomycin and 1% (v/v) L-glutamine (200 mM). 697 and HAL-01 were maintained in a concentration of 0.5 up to  $1.5 \times 10^6$  cells/mL. REH was maintained in a concentration of 0.5 till  $2 \times 10^6$  cells/mL. SEM was maintained in a concentration of 0.2 to  $2 \times 10^6$  cells/mL. HSB-2 was maintained in a concentration of 0.3 up to  $1.3 \times 10^6$  cells/mL. All cell lines were grown at 37°C in a 5% CO<sub>2</sub> enriched atmosphere. Cells were split 1:3 two times a week. Cell counting and viability check was performed with Neubauer chamber or Vi-Cell XR by Beckmann Coulter.

#### **3.1.2 Cultivation of Human Adherent Cell Lines**

HEK293T were cultivated in DMEM with 10% (v/v) FBS and 1% (v/v) 10,000 U/mL Penicillin-Streptomycin. The cells were split 1:10 two times a week and cultivated at 37°C in a 5% CO<sub>2</sub> enriched atmosphere. Cell counting and viability check was performed with Neubauer chamber or Vi-Cell XR by Beckmann Coulter.

#### **3.1.3 Cryopreservation of Human Cell Lines**

Cells were transferred in Eppendorf tubes and pelleted for five minutes at 500 x g.  $1 \times 10^6$  cells were resuspended in 1 mL of freezing medium (DMEM, 10% FBS, 10% DMSO) and transferred into cryotubes. Cryotubes were placed in a pre-chilled Mr. Frosty freezing container filled with isopropanol and frozen at -80°C. After 24 h, cryotubes were stored in the gas phase of a liquid nitrogen tank.

For recultivation of cells, a cryotube was thawed in a water bath at 37°C. Cell suspension was transferred into pre-warmed RPMI 1640 for DMSO dilution. Freezing medium was removed by centrifugation at 500 x g for 5 min. Afterwards the cell pellet was resuspended in fresh culture medium and transferred to a culture dish.

### **3.1.4 Treatment with Inhibitors**

1 x 10<sup>6</sup> cells (HAL-01)/6-well were transferred into 6-well plates and inhibitors (KSK concentration: 100 nM, Ricolinostat concentration: 2.8 µM) diluted in DMSO were added. Ricolinostat was used as a positive control. Pure DMSO was used as a negative control. After 24 h cells were harvested by centrifugation at 1200 rpm for ten minutes, washed with PBS and centrifuged at 3.500 rpm for ten minutes at 4°C. Cell pellets were shock frozen in liquid nitrogen and stored at -20°C.

## **3.2 Methods of Molecular Biology**

### **4.2.1 Extraction of Nucleic Acids**

DNA, RNA and miRNA were extracted using AllPrep DNA/RNA/miRNA Universal Kit from Qiagen following the manufacturer's protocol.

### **3.2.3 Extraction of Total RNA by TRIzol Reagent**

Extraction of RNA was performed with TRIzol reagent for Northern blot analysis as downstream application. Cells were harvested, pelleted and 0.75 mL of TRIzol were added to 5 to 10 x 10<sup>6</sup> cells. Cells were homogenized and lysed by resuspension and incubated at room temperature (RT) for 5 min. One hundred fifty µL chloroform were added to 750 µL of TRIzol. The tubes were shaken by hand for 15 seconds and incubated for another 3 min. In order to obtain a clear phase separation, samples were centrifuged at 12,000 x g at 4° C for 15 min. The upper phase, containing total RNA, was transferred into a new tube. 357 µL of 100% isopropanol were added per 750 µL of applied TRIzol in the initial homogenization and samples were incubated for 10 min. To pellet total RNA, samples were centrifuged at 12,000 x g at 4° C for 10 min. Supernatant was discarded and RNA pellet was washed with 750 µL of

85% ethanol per 750  $\mu\text{L}$  applied TRIzol in initial homogenization. Samples were vortexed and centrifuged at 7,500 x g at 4°C for 5 min. Supernatant was discarded and RNA pellets were air dried for 10 min. RNA pellets were resuspended in 20  $\mu\text{L}$  of RNase-free water and incubated in a heat block at 55°C for 15 min. RNA was stored at -80°C.

### **3.2.2 Determination of Concentration and Purity of Nucleic Acids**

The concentration of nucleic acids was measured using NanoDrop 1000 spectrophotometer measuring the absorption at 260 nm. An  $\text{OD}_{260} = 1$  has a concentration of 50  $\mu\text{g}/\text{mL}$  of double stranded DNA and a concentration of 40  $\mu\text{g}/\text{mL}$  of RNA. Purity of the solution was determined by measuring the absorption at 260 nm and 280 nm and computing the ratio 260/280. A pure DNA solution has a ratio of 1.8. A pure RNA solution has a ratio of 2.0.

### **3.2.3 Reverse Transcription**

Reverse transcription of RNA was performed with Superscript III reverse transcriptase (Invitrogen). 1.5  $\mu\text{g}$  of total RNA were mixed with 1  $\mu\text{L}$  of oligo(dT)<sub>20</sub> primer (50  $\mu\text{M}$ ) and 1  $\mu\text{L}$  of dNTP mix (10 mM each) in a 13  $\mu\text{L}$  mixture. The tube was incubated at 65°C for 5 min and placed on ice for at least one minute. Afterwards 4  $\mu\text{L}$  of 5x First-Strand buffer, 1  $\mu\text{L}$  of 0.1 M DTT, 1  $\mu\text{L}$  of 40 U/ $\mu\text{L}$  RNase OUT and 1  $\mu\text{L}$  of 200 U/ $\mu\text{L}$  Superscript III RT per reaction were added. After gently mixing, an incubation at 50°C for one hour followed. Subsequently, inactivation of the enzyme took place at 70°C for 15 min.

### **3.2.4 Agarose Gel Electrophoresis**

For separation of nucleic acids agarose gel electrophoresis was performed. Separation was done in a 1% agarose gel with 1:10.000 ethidium bromide for 45 min at 120 V. Visualization of DNA bands was performed with Intas UV lamp.

### **3.2.5 Restriction Digest**

Restriction enzyme double digest was performed in a 30  $\mu\text{L}$  volume. The digestion mixture contained plasmid DNA (up to 1.5  $\mu\text{g}$ ) or PCR product (up to 500 ng), restriction buffer and

compatible restriction enzymes (each 1  $\mu$ L). Digestion was performed for two hours at 37°C.

### **3.2.6 Gel Extraction**

Extraction of DNA after gel electrophoresis was done with QIAquick Gel Extraction Kit from Qiagen. For elution, 30  $\mu$ L of buffer EB were used.

### **3.2.7 Ligation**

Ligation was performed in 20  $\mu$ L mixture. It contained 50 ng of vector, a molar insert:vector ratio of 3, T4 DNA ligase reaction buffer and T4 DNA ligase from NEB. Insert:vector ratio was calculated as follows:

.

The mixture was incubated at RT for one hour. Afterwards it was incubated at 16°C overnight (ON).

### **3.2.8 Transformation of One Shot™ Top10 Competent E.coli**

Chemically-competent Escherichia coli (E.coli) cells were thawed on ice and 1  $\mu$ L of ligation mixture was added. The vial was mixed gently and incubated on ice for 30 minutes. Heat-shock took place at 42°C for 30 s. Afterwards cells were incubated on ice for 2 min. Two hundred fifty  $\mu$ L of pre-warmed LB medium was added and cells were incubated for one hour at 37°C with 350 rpm in a shaking incubator. Twenty  $\mu$ L and 200  $\mu$ L of cell suspension were plated on ampicillin (100  $\mu$ g/mL) containing LB agar plates. Plates were inverted and incubated at 37°C ON for colony selection. pUC19 vector served as a positive control and H<sub>2</sub>O as a negative control. Clones were picked and cultivated in 10 mL LB media with ampicillin (100  $\mu$ g/mL) at 37°C ON. Cells were harvested and plasmid DNA was analyzed by restriction digest for the correct insert. Next, 10  $\mu$ L of cell suspension were used to inoculate 100 mL of LB-medium with ampicillin (100  $\mu$ g/mL) and incubated at 37°C ON.

### **3.2.9 Plasmid Extraction**

Extraction of plasmid DNA was performed with Fast Plasmid Mini-Kit from 5 Prime for a small scale approach (2 mL bacterial culture) according to manufacturer's instructions.

Extraction of plasmid DNA was performed with HighSpeed Plasmid Maxi Kit from Qiagen for a large scale approach (100 mL bacterial culture) according to manufacturer's instructions.

### **3.2.10 Cloning of RNAi Expression Vectors**

For establishing RNAi expression vectors the BLOCK-iT Pol II miR RNAi Expression vector kit with EmGFP was used. First double stranded oligos of desired shRNA were generated. At first, single stranded oligos were diluted to a concentration of 200  $\mu$ M. Next, a 20  $\mu$ L annealing reaction was prepared: 5  $\mu$ L of 200  $\mu$ M top strand oligo, 5  $\mu$ L of 200  $\mu$ M bottom strand oligo, 2  $\mu$ L of 10x Oligo Annealing buffer and 8  $\mu$ L of nuclease free water. The mixture was incubated at 95°C for 4 min. Afterwards it was incubated at room temperature for 10 min to cool down. In this time top and bottom strand annealed to each other. 1  $\mu$ L of the mixture was diluted to 10 nM in two steps: first a 1:100 dilution in nuclease free water was performed to receive a 500 nM solution, second, 1  $\mu$ L of the dilution was mixed with 5  $\mu$ L of 10x Oligo Annealing buffer and 44  $\mu$ L of nuclease free water to receive a 10 nM solution. To check the integrity, single strands and annealed strands were analyzed by agarose gel electrophoresis. Samples were run on a 2% agarose gel at 100 V for 45 min. Afterwards double-stranded oligos were ligated into linearized pcDNA™ 6.2-GW/EmGFP-miR. A 20  $\mu$ L ligation mixture was prepared with 4  $\mu$ L of 5x ligation buffer, 2  $\mu$ L of pcDNA™ 6.2-GW/EmGFP-miR (5 ng/ $\mu$ L), 4  $\mu$ L of miR-double strand oligos, 9  $\mu$ L of nuclease free water and 1  $\mu$ L of T4 DNA ligase (1 U/ $\mu$ L). A negative control contained the same ingredients but no miR-double strand oligos. The ligation mixture was resuspended and incubated at RT for 1 h. Two  $\mu$ L of the ligation mixture were used to transform OneShot TOP10 competent E.coli. Transformed E.coli were plated out on LB agar plates containing 50  $\mu$ g/mL spectinomycin. miRNA forward and reverse sequencing primers were used for Sanger sequencing.

### **3.2.11 Lipofection**

One day before transfection 1 x 10<sup>5</sup> HEK293T cells were seeded out per well on a 24-well plate so that they were 90% confluent at the day of transfection. One  $\mu$ L of lipofectamine 2000 was diluted in 25  $\mu$ L of Opti-MEM media. One  $\mu$ g of plasmid DNA plus 100 pM siRNA were diluted in 50  $\mu$ L Opti-MEM media. Diluted DNA and diluted lipofectamine 2000 were mixed



in a 1:1 ratio (50  $\mu$ L each). The mixture was incubated at RT for five min. Fifty  $\mu$ L of DNA-lipid complex solution were added to each well of the 24-well plate. Cells were incubated at 37°C. Twenty-four hours after transfection the medium was replaced with fresh culture medium. Transfected cells were harvested between one and three days after transfection.

### 3.2.12 Luciferase Assay

For luciferase assay, the following reagents were prepared:

**1x Passive Lysis Buffer (PLB):** 1 volume of 5x PLB and 4 volumes of H<sub>2</sub>O dest.

**LAR II:** resuspension of lyophilized Luciferase Assay Substrate in Luciferase Assay Buffer II

**Stop & Glo reagent:** dilution of 50x to 1x Stop & Glo substrate with Stop & Glo buffer

From day one until day three after transfection of HEK293T cells with psiCHECK2 containing the gene of interest, growth medium was removed and cells were washed twice with 1x PBS. For lysis of the cells 100  $\mu$ L PLB were added per well of the 24 well plate. The plate was shaken at RT for 15 min. Lysates were transferred to a tube and centrifuged at 14,000 rpm for 30 s to remove cell debris. Twenty mL of supernatant were transferred into a white 96-well plate. One hundred  $\mu$ L of LAR II were added and firefly luciferase activity was measured on a Luminoskan Ascent Microplate Luminometer. Immediately, 100  $\mu$ L of Stop & Glo reagent were added and renilla luciferase activity was measured on a Luminoskan Ascent Microplate Luminometer. The luminometer protocol was as follows: 1) shake at 600 rpm for 3 s, 2) LARII measurement integration time: 10 s, lag time: 2 s, 3) manually add renilla reagent, 4) shake at 600 rpm for 3 s, 5) renilla measurement time 10 s, lag time: 2 s. For normalization renilla luciferase activity was divided by firefly luciferase activity.

### 3.2.13 Whole Genome Amplification

To amplify genomic DNA (gDNA) of patient material, whole genome amplification (WGA) was performed with REPLI-g UltraFast Mini Kit from Qiagen following the manufacturer's instructions.

### **3.2.14 Sanger Sequencing**

Amplified genomic patient DNA was used for amplification of mutated regions. PCR primers for PCR amplification flanked the mutated sites. PCR was performed with High Fidelity DNA polymerase from NEB in 50  $\mu$ L reactions with HF buffer. Cycling was performed at 98°C for 30 s as an initial denaturation step, 30 cycles of 98°C for 5 s, 60°C for 30 s, 72°C for 20 s/kbase and a final extension at 72°C for 10 min. Sanger sequencing of amplicons was performed at Biologisch-Medizinisches Forschungszentrum (BMFZ) Düsseldorf.

### **3.2.15 Northern Blot**

RNA for Northern blot analysis was isolated using the TRIzol reagent protocol. Northern Blot probes were diluted and labeled as follows: a 20  $\mu$ L mixture of 20  $\mu$ M oligonucleotide, 6000Ci/mmol gamma-32p-ATP, 1 x T4 PNK buffer, ddH<sub>2</sub>O and 2.4 U T4 PNK was prepared. The reaction mixture was incubated at 37°C for 15 min. To stop the reaction, 30  $\mu$ L of 30 mM EDTA (pH 8.0) were added. To remove unbound gamma-32p-ATP, the mixture was transferred to the top-center of the resin of a prepared sephadex G-25 spin column and centrifuged at 735 x g for 2 min. Probes were stored at -20° C.

A 15 x 17 cm polyacrylamide gel with 1.5 mm thickness was cast. Thirty  $\mu$ g of RNA were mixed with 2x loading buffer. Samples were heated at 92°C for 1 min and then placed on ice. The size marker was heated to 92°C for 5 min and placed on ice. Size marker and samples were loaded on the gel. The gel ran at 10 W for 10 min and then at 30 W for 45 min. For visualization, the gel was removed from the glass plates and slowly shaken in ethidium bromide solution for 10 min. Samples and size marker were transferred to Amersham hybond-NX membrane, a charged nylon membrane, in a semi-dry transfer cell from BIO-RAD in 0.5x TBE-buffer flanked from blotting paper. Transfer was performed at 300 mA for 1.5 h. Afterwards, RNA was UV-crosslinked to the membrane at 1200 J for 30 s with UV Stratalinker 2400. Another crosslinking step was performed by incubation in 24 mL EDC crosslinking solution at 60°C for 1 h. Afterwards, membrane was rinsed with ddH<sub>2</sub>O and incubated in 30 mL pre-hybridization solution (without probes) at 50° C for 1 h. After that, labeled probes were added to the pre-hybridization solution and the membrane was incubated at 50°C for 12 h. Washing solution was pre-warmed to 50°C and then used for washing the membrane after hybridization at 50°C for 5 min. For exposition to a phosphorimager screen,

the membrane was wrapped in saran wrap and exposed to the screen. A picture was taken with FLA-7000 from Fujifilm. Stripping of the membrane was performed by washing it with 100 mL 0.1 % SDS twice at 85°C. Stripping was checked by exposing the membrane to the phosphorimager screen. The exposed screen was erased by exposing it to white light for 45 min with Fujifilm IP Eraser 3.

### **3.2.16 Western Blot**

Cells were harvested and centrifuged at 2000 rpm at 4°C for 3 min. The pellet was washed with 1 mL PBS and again centrifuged at 2000 rpm at 4°C for 3 min. PBS was discarded and the pellet was stored at -80°C or directly lysed. For cell lysis 50 mL NP40 buffer were added with 0.5 mL of 1 M DTT and one tablet of protease inhibitor. To 1 mL of cell pellet, 3 mL of lysis buffer were added. The solution was incubated on ice for 15 min and then centrifuged at 13000 rpm at 4°C for 30 min. Lysate was then transferred into a new tube. Concentration of protein was measured using the Bradford assay. Two µL of protein lysate were mixed with 1 mL of 1x Bradford reagent and measured with Eppendorf biophotometer with an OD of 600 nm in duplicates. Protein concentrations were normalized and 1x SDS loading dye was added. Proteins were denatured at 95°C for 5 min. Ten x SDS running buffer was freshly diluted to 1x. 12% protein gels were prepared. Six µL of loading buffer were loaded on the gel. Samples were loaded and the gel did run at 120 V for 90 min. After separation of proteins the gel was put into transfer buffer. Blotting membrane and Whatman paper were also moistened with transfer buffer. Gel and membrane were put into the semidry transfer chamber surrounded by three Whatman paper at the bottom and on top. Air bubbles were removed by pressing the layers with a pipette. Transfer of proteins to the membrane was performed at 80 mA for 2 h for one membrane. Afterwards the membrane was blocked in 5% milk-TBST or 5% BSA-TBST solution for 1 h. After blocking, the first antibody was diluted in blocking solution and incubated at RT for 3 h. Then the membrane was washed three times with 1x TBST for 10 min and was incubated with secondary antibody diluted in blocking solution at RT for 1 h. Again the membrane was washed three times with 1x TBST. Chemiluminescence detection solution was prepared and added to the blot. It was incubated at RT for 5 min. Chemiluminescence signal was detected using LAS 3000 mini.

### **3.3 Polymerase Chain Reaction (PCR)**

#### **3.3.1 Standard PCR**

cDNA was used to amplify certain gene regions flanked by specific primers. Phusion High Fidelity Polymerase from NEB was used for amplification. Standard PCR mixture contained Phusion HF buffer, dNTPs (200  $\mu$ M), specific forward and reverse primers (each 0.5  $\mu$ M), template DNA (50-250 ng) and the polymerase (1 U/50  $\mu$ l mixture). The PCR reaction was performed in 50  $\mu$ L PCR mixture. PCR was performed at an initial denaturation temperature of 98°C for 30 s, 35 cycles of denaturation at 98°C for 10 s, annealing at 3°C above the melting temperature ( $T_m$ ) of lower  $T_m$  primer and extension at 72°C for 15 s per kb and a final extension at 72°C for 5 min. Actin primers were used as positive control. H<sub>2</sub>O instead of template DNA was used as negative control.

#### **3.3.2 Quantitative Real Time PCR (qRT-PCR)**

For quantification of a specific transcript, a qRT-PCR was performed. A master mix of 12.5  $\mu$ L SYBR Green, 1  $\mu$ L forward primer (10 pmol), 1  $\mu$ L reversed primer (10 pmol) and 9.5  $\mu$ L H<sub>2</sub>O were prepared and distributed on a PCR plate (24  $\mu$ L/well). One  $\mu$ L of cDNA was added per well. The PCR plate was covered with a microseal. The following program was used for amplification: 95°C for 10 min, 40 cycles of 95°C for 15 s and 60°C for 1 min. Fluorescence was measured with CFX Connect real-time system (Bio-Rad).

### **3.4 Next Generation Sequencing**

#### **3.4.1 Quality Control of Extracted RNA**

Quantity and quality of total RNA were analyzed with Agilent BioAnalyzer 2100 using the Nano RNA 6000 Kit. RNA integrity number (RIN) scores and sample profiles were visually inspected to evaluate sample quality. Samples with RIN-scores from 8 to 10 were used for further processing.

#### **3.4.2 miRNA Library Preparation**

miRNA library preparation was performed using TruSeq Small RNA Sample prep kit from Illumina using a maximum amount of 1  $\mu$ g of total RNA per sample.

### **3.4.3 Total RNA Library Preparation**

Total RNA library preparation was performed using TruSeq Stranded Total RNA Library Prep Kit with Ribo Zero-Gold by Illumina with 300 ng total RNA per sample as input. The fragmentation step was adapted depending on the quality of the RNA sample.

### **3.4.4 Library Validation**

Total RNA and miRNA libraries were quantified by loading 1  $\mu$ L of cDNA on a DNA specific chip (DNA 1000 or High Sensitivity DNA). The concentration was measured with Agilent Bioanalyzer 2100.

### **3.4.5 Loading of the Flow Cell**

Denaturation of template DNA was performed by adding 0.1 N NaOH to a final DNA concentration of 20 pM (10  $\mu$ L 2 nM template DNA + 10  $\mu$ L 0.1 N NaOH). The solution was vortexed and spun down at 280 x g for 1 min. The mixture was incubated at RT for 5 min. Afterwards, 20  $\mu$ L of denatured DNA were transferred to 980  $\mu$ L of pre-chilled Hybridization buffer (HT1). Next, template DNA was diluted to 6.5 pM with HT1. Samples, flow cell and cBot reagents were loaded on cBot and the S.R.Amp-Lin/Block/Hyb.V8 program was started to load samples on the flow cell.

### **3.4.6 Deep Sequencing**

High-throughput sequencing was performed on a HiSeq 2500 for 51 cycles plus seven for miRNA library sequencing and for 101 plus seven cycles for total RNA library sequencing according to standard protocols.

## **3.5 Bioinformatic Methods**

### 3.5.1 Analysis of Transcriptome Data

After next generation sequencing adapter sequences and low quality ends were trimmed off from reads of mRNA and miRNA datasets with seqtk (<https://github.com/lh3/seqtk>) and cutadapt<sup>92</sup>. Remaining reads were aligned against the human reference genome GRCh38 with STAR<sup>93</sup> and BWA<sup>94</sup> for RNA-seq and miRNA-seq, respectively. BWA was applied with -n 0.04 option to accommodate for sequencing errors in the miRNA-Seq. For read count calculation HTseq<sup>95</sup> was used for mRNA and miRNA datasets. Annotations from Ensemble Genes V82 ADDIN EN.CITE <sup>96</sup> were used for mRNAs and annotations from miRbase V21<sup>97</sup> were used for miRNAs. EdgeR<sup>98</sup> was used to inter-sample normalize expression counts (leading to counts per million (CPM values)) and to perform a differential gene expression between the cohorts.

### 3.5.2 Pathway Analysis

Pathway analysis and upstream regulator analysis were performed using Ingenuity® Pathway Analysis (IPA) from Qiagen. Upstream regulator analysis takes differentially expressed genes as input and queries literature entries for upstream regulators of the respective genes. Depending on the regulatory effect of the potential upstream regulator and the actual expression of its targets, it is either considered as inhibited or activated. Expression of upstream regulators in different cell types were analyzed with R2 MegaSampler (<https://hgserver1.amc.nl/cgi-bin/r2/main.cgi>).

### 3.5.3 Correlation Analysis

To find mRNA-miRNA target pairs expressed in a negative correlation fashion, a correlation analysis based on five different public miRNA databases (miRanda August 2010<sup>97</sup>, miRDB V5 ADDIN EN.CITE <sup>99</sup>, TarBase V7<sup>100</sup>, TargetScan V6.2<sup>101</sup>, miRTarBase V4.5 ADDIN EN.CITE <sup>102</sup>) was performed. A target pair was used for downstream analysis, if it was either reported as experimentally validated (by miRTarBase V4.5) or predicted by at least two of the computational prediction algorithms. All pairs without a significant differential expression for any miRNA or mRNA between the cohorts were discarded. The spearman's rho was calculated for the resulting mRNA-miRNA target pairs.

### **3.5.4 Analysis of Exome Data**

Alignment of read sequences to the human reference genome (GRCh37) was performed with BWA version 0.6.1-r104<sup>103</sup>. Conversion steps were carried out using Samtools<sup>104</sup> followed by removal of duplicated reads by mapping positions ADDIN EN.CITE <sup>105</sup>. All reads aligned with at least one insertion/deletion (indels) compared to the genome were locally realigned by GATK ADDIN EN.CITE <sup>106</sup>. GATK was also used for SNV calling, annotation, and recalibration. For recalibration, data from HapMap, OmniArray and dbSNP V135 provided by the Broad Institute were used. Resulting variation calls were annotated by Variant Effect Predictor<sup>107</sup> using data from Ensembl V70 and imported into an in-house MySQL database to facilitate automatic and manual annotation, reconciliation and data analysis. Predictions for loss of function were provided by PolyPhen2<sup>108</sup> and SIFT<sup>109</sup>. Sequence variants with less than 15% frequency in the 1000 genomes and hapmap projects were considered for further analysis

## 4. Results

### 4.1 Differences of the mRNA/miRNA expression pattern between t(1;19)-ALL and t(17;19)-ALL

#### 4.1.1 Patient Cohort

To find differences between the frequently occurring t(1;19)-positive B-ALL (*TCF3/PBX1*) and the rare and aggressive t(17;19)-positive B-ALL (*TCF3/HLF*), blast material of five patients of each cohort was analyzed by NGS. The aim was to find novel and mutated miRNAs in one of the cohorts, whose expression could explain the clinical differences between both B-ALL subtypes. Additionally, there was a focus on differentially expressed miRNAs and mRNAs, which could act as mRNA-miRNA target pairs and have a potential impact on disease development.

RNA of blast and remission material from five pediatric B-ALL patients with t(1;19) (Table 1) and five pediatric B-ALL patients with t(17;19) (\* MERGEFORMAT WBC: white blood cell count; CNS: central nervous system; CNS status: 1 – lumbar puncture nontraumatic with no leukemic blasts in the cerebrospinal fluid (CSF) after cytocentrifugation; 2 – puncture nontraumatic with  $\leq 5$  leukocytes/ $\mu\text{L}$  in the CSF with identifiable blasts. Prednisone response evaluated after 7 days induction with daily prednisone and a single intrathecal dose of methotrexate on treatment day 1; good:  $< 1000/\mu\text{L}$  blood blasts; poor  $\geq 1000/\mu\text{L}$  blood blasts. M1 -  $< 5\%$  blasts; CCR: continuous complete remission; HSCT: hematopoietic stem cell transplantation; BM: bone marrow; PB: peripheral blood; BFM: Berlin-Frankfurt-Münster study group. Table taken from Fischer et al.<sup>110</sup>

Table 2) was extracted and used for mRNA and miRNA library preparation. Patients with t(1;19) are numbered from 1 - 5. Patients with t(17;19) are numbered from 6 - 9 and 11. Tumor samples named with a postpositive "a" and remission samples are named with a postpositive "b". Remission samples were used as germline controls and were taken from patients after induction treatment with minimal residual disease (MRD, maximum leukemic cell load cell load  $\leq 10^{-3}$ ). The median age of patients with t(1;19)-positive ALL, as well as the median age of patients with t(17;19)-positive ALL, was 14.5 years at diagnosis of disease. There was no gender preference regarding *TCF3-PBX1*-positive ALL samples. In the analyzed *TCF3/HLF* cohort, 4 out of 5 patients were male. Patients with t(1;19) achieved complete remission after



treatment. Most patients with t(17;19) responded to induction chemotherapy but remained MRD positive and died of relapse or due to treatment-related causes within an average of 2 years after initial diagnosis.

**Table 1: *TCF3-PBX1* Patient information.**

	<i>TCF3/PBX1</i>				
patient	1	2	3	4	5
sex	male	female	male	female	female
age at diagnosis (in years)	15.5	14.5	10.8	1.8	15.5
WBC/ $\mu$ L at diagnosis	26800	90600	10370	102000	15520
CNS status	CNS1	CNS1	CNS1	CNS2	CNS1
prednisone response	good	poor	poor	good	poor
morphological BM status after induction treatment	M1	M1	M1	M1	M1
MRD after induction treatment	negative	1,00E-04	negative	negative	negative
treatment outcome	CCR	CCR after HSCT	CCR	CCR	CCR
Leukemic specimen used for sequencing (% blasts)	BM (96)	PB (92)	BM (98)	BM (90)	BM (94)
germline reference tissue	BM	BM	BM	BM	BM
treatment protocol	BFM	BFM	BFM	BFM	BFM

WBC: white blood cell count; CNS: central nervous system; CNS status: 1 – lumbar puncture nontraumatic with no leukemic blasts in the cerebrospinal fluid (CSF) after cytocentrifugation; 2 – puncture nontraumatic with  $\leq 5$  leukocytes/ $\mu$ L in the CSF with identifiable blasts. Prednisone response evaluated after 7 days induction with daily prednisone and a single intrathecal dose of methotrexate on treatment day 1; good:  $<1000/\mu$ L blood blasts; poor  $\geq 1000/\mu$ L blood blasts. M1 -  $< 5\%$  blasts; CCR: continuous complete remission; HSCT: hematopoietic stem cell transplantation; BM: bone marrow; PB: peripheral blood; BFM: Berlin-Frankfurt-Münster study group. Table taken from Fischer et al. [ADDIN EN.CITE](#) <sup>110</sup>

**Table 2: *TCF3-HLF* Patient information.**

	<i>TCF3/HLF</i>				
patient	6	7	8	9	11
sex	male	female	male	male	male
age at diagnosis (in years)	16.5	14.5	14.1	13.8	15.5
WBC/ $\mu$ L at diagnosis	14300	2200	6300	5800	20300
CNS status	CNS1	CNS1	CNS1	CNS1	CNS1
prednisone response	good	good	poor	good	good
morphological BM status after induction treatment	M1	M1	M1	M1	M3
MRD after induction treatment	1,00E-03	1,00E-03	1,00E-01	1,00E-03	1,00E-02
treatment outcome	TRM	TRM	relapse and death	relapse and death	relapse and death

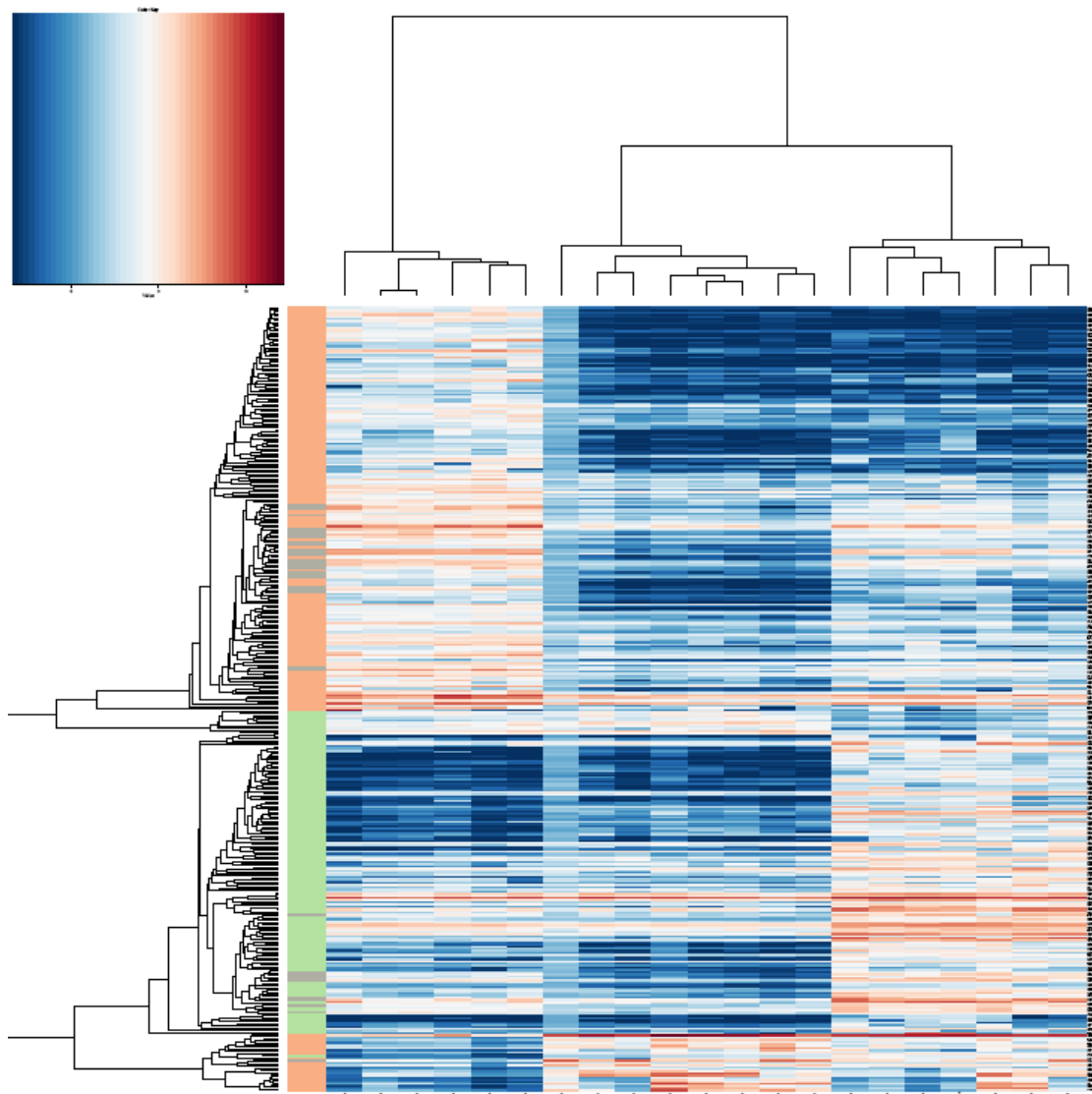
<b>Leukemic specimen used for sequencing (% blasts)</b>	BM(98)	BM(81)	BM(61)	BM(72)	BM(88)
<b>germline reference tissue</b>	BM	BM	BM	BM	BM
<b>treatment protocol</b>	BFM	BFM	BFM	FRALLE	BFM

WBC: white blood cell count; CNS: central nervous system; CNS status: 1 – lumbar puncture nontraumatic with no leukemic blasts in the CSF after cytocentrifugation. Prednisone response evaluated after 7 days induction with daily prednisone and a single intrathecal dose of methotrexate on treatment day 1; good: <1000/ $\mu$ L blood blasts; poor  $\geq$ 1000/ $\mu$ L blood blasts. M1 - < 5% blasts; M3 - > 5% blasts; TRM: treatment-related mortality; BM: bone marrow. BFM: Berlin-Frankfurt-Münster study group. Table taken from Fischer et al. [ADDIN EN.CITE](#) <sup>110</sup>

## 4.1.2 Target Prediction of Differentially Expressed Genes

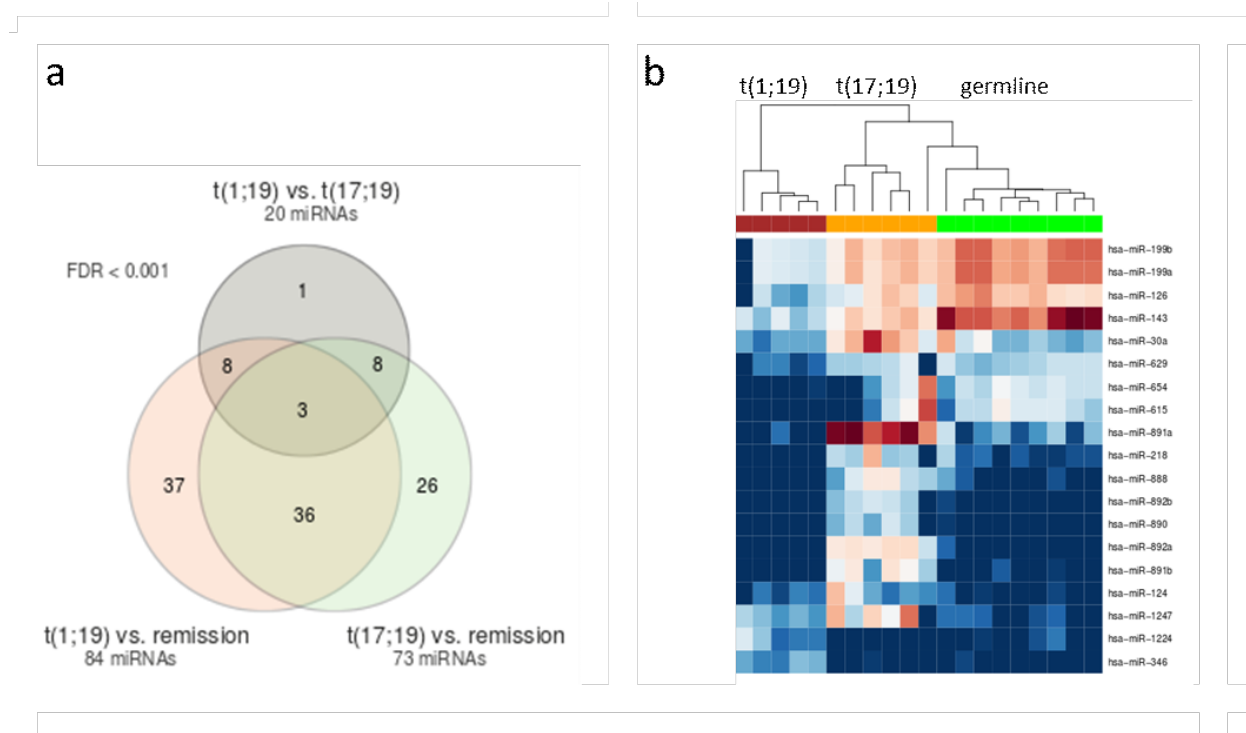
### 4.1.2.1 144 mRNAs and 19miRNAs are differentially expressed between t(1;19)- and t(17;19)-positive ALL in an initial analysis

cDNA libraries were sequenced with NGS and evaluated. Overall, 144 mRNAs were differentially expressed between t(1;19)-ALL and t(17;19)-ALL ( $|\log_{2}FC| \geq 1$ , p-value  $\leq 0.001$ ). Hierarchical cluster analysis revealed significant differences between t(1;19) and t(17;19) (Figure 10).



**Figure 10: Heatmap of differentially expressed genes.** On the left side patients with t(1;19), on the right side patients with t(17;19) and in the middle remission samples clustered.

Analysis of differentially expressed miRNAs revealed 84 miRNAs differentially expressed between t(1;19)-positive B-ALL and the corresponding germline samples and 73 differentially expressed miRNAs between t(17;19)-B-ALL and the corresponding germline samples ( $|\logFC| \geq 1.89$ ,  $FDR < 0.0001$ ). Twenty miRNAs were differentially expressed between t(1;19)- and t(17;19)-positive B-ALL. Out of these, 19 miRNAs were also differentially expressed between t(1;19)-positive B-ALL, t(17;19)-positive B-ALL and the germline samples (Figure 11a). Tumor and germline samples as well as t(1;19) and t(17;19) were separated by hierarchical clustering by these 19 miRNAs (Figure 11b).



**Figure 11: Analysis of differentially expressed miRNAs.** a) Venn diagram showing the numbers of miRNAs differentially expressed between analysed cohorts. b) Heatmap of 19 differentially expressed miRNAs between t(1;19)-positive B-ALL, t(17;19)-positive B-ALL and germline samples. Red: t(1;19)-ALL samples; yellow: t(17;19)-ALL samples, green: healthy samples.

For these 19 miRNAs target prediction and correlation analysis regarding the differentially expressed mRNAs were performed. The analysis revealed 14 potential mRNA-miRNA pairs with a correlation coefficient of less than -0.6 (Table 3). A negative correlation coefficient points to a negative correlation (i.e. high miRNA expression is correlated to low mRNA expression or vice versa). A numerically high correlation coefficient points to correlation between mRNA and miRNA expression.

**Table 3: Potential mRNA-miRNA target pairs.**

miRNA	mRNA	Correlation coefficient
hsa-miR-30a	<i>BaZ2B</i>	-0.7
hsa-miR-30a	<i>SEC23A</i>	-0.7
hsa-miR-124	<i>ABHD5</i>	-0.69
hsa-miR-124	<i>FAR1</i>	-0.69
hsa-miR-124	<i>NRG1</i>	-0.69
hsa-miR-124	<i>IL6R</i>	-0.7
hsa-miR-124	<i>SMOX</i>	-0.69
hsa-miR-124	<i>FCHO2</i>	-0.7
hsa-miR-124	<i>TWIST2</i>	-0.73
hsa-miR-124	<i>DNAJC1</i>	-0.85

hsa-miR-124	<i>ITGB1</i>	-0.78
hsa-miR-199a	<i>SOX4</i>	-0.69
hsa-miR-199b	<i>SOX4</i>	-0.72
hsa-miR-892c	<i>MCL1</i>	-0.72

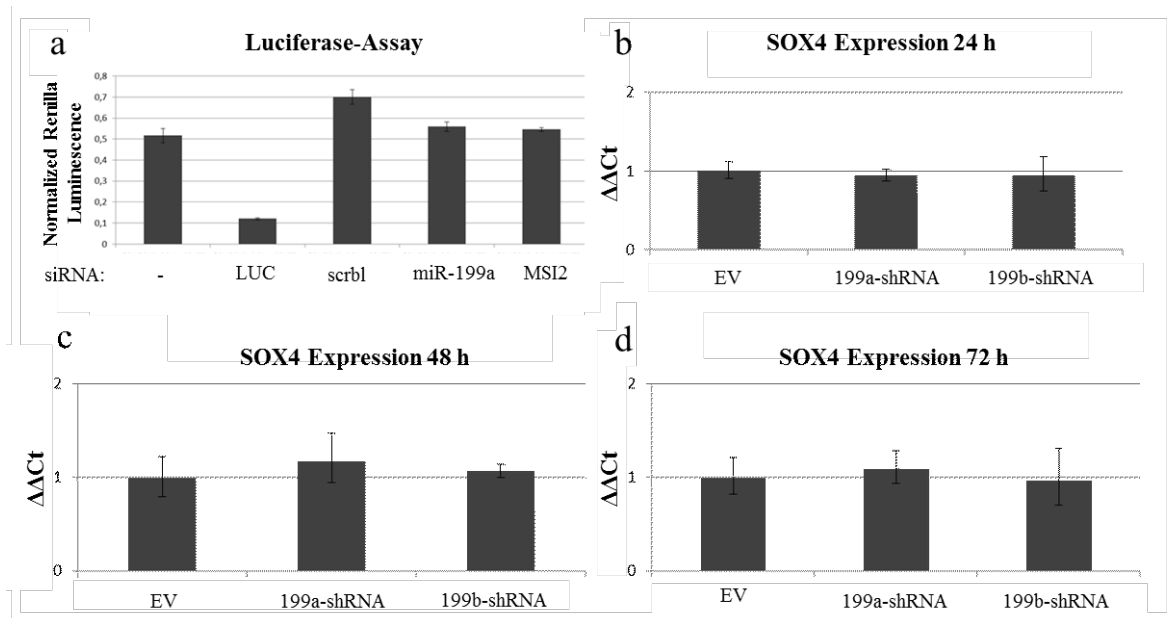
Target pair prediction of differentially expressed mRNAs and miRNAs revealed 14 potential pairs with a correlation coefficient <-0.6.

Two of these pairs involve hsa-miR-199a and hsa-miR-199b with their potential target Sex Determining Region Y-related Box4 (*SOX4*). *SOX4* act as an oncogene as well as a tumor suppressor dependent on the cellular context. Deregulated expression of *SOX4* is correlated with increased proliferation, cell survival, apoptosis inhibition, metastasis and tumor progression in different kinds of cancers. Knockout of *SOX4* in hematopoietic cells lead to a block in B-cell development at the pro-B-cell stage and to a decreased development of T cells (reviewed by Vervoort, Boxtel and Coffey, 2013)<sup>111</sup>. Hsa-miR-199a/b are downregulated in t(1;19)-ALL with a log fold change (FC) of -5.4 while *SOX4* expression is upregulated. Hsa-miR-199a and hsa-miR-199b belong to one sequence family due to their high sequence similarity (3p strands: 100% identity, 5p strands: 91.3% similarity; calculated by Smith-Waterman algorithm). The 5p strands are predicted to bind *SOX4*-3'-untranslated region (UTR) with their seed sequence (CCAGUG) (100% identity between miR-199a-5p and miR-199b-5p in seed region). Target scan predicts binding of hsa-miR-199a/b-5p to *SOX4*-3'-UTR at bp position 2135-2141. To validate the regulation of *SOX4* by hsa-miR-199a/b a luciferase assay was performed. *SOX4* 3'UTR was cloned into psiCHECK2 vector (Promega).

HEK293T cells were transfected with five different mixtures of nucleic acids: 1) psiCHECK2-*SOX4*-3'UTR only (negative control (CTRL)), 2) psiCHECK2-*SOX4*-3'UTR in combination with Luciferase (LUC)-siRNA (positive CTRL), 3) psiCHECK2-*SOX4*-3'UTR in combination with scrambled (scrbl)-siRNA (negative CTRL), 4) psiCHECK2-*SOX4*-3'UTR in combination with miR-199a-siRNA and 5) psiCHECK2-*SOX4*-3'UTR in combination with *MSI2*-siRNA (negative CTRL). Renilla luminescence was measured on Luminoskan Ascent Microplate Luminometer. However, no significant differences in renilla luminescence between samples miR-199a-siRNA and negative controls psiCHECK2-*SOX4*-3'UTR and psiCHECK2-*SOX4*-3'UTR with *MSI2*-siRNA were observed neither after 24 h (Figure 12a), 48 h nor 72 h.

To confirm these results, short hairpin (sh) RNA vectors were designed with shRNA-199a or shRNA-199b (pcDNA6.2GW/EmGFP-miR) with GFP as reporter gene. HEK293T cells were transfected with one of these vectors or with empty vector (EV, without any shRNA) as a control. After 24 h, 48 h and 72 h RNA of these cells was extracted and qPCR with *SOX4*

primers was performed. The expression of GFP was measured by qPCR as a control for positive transfection. No difference in *SOX4* expression could be observed between shRNA-miR199a/b vector transfected cells and EV transfected cells after 24 h (Figure 12b), 48 h (Figure 12c) and 72 h (Figure 12d).



**Figure 12: Interaction between miR-199 and *SOX4*-3'UTR is not confirmed by Luciferase assay (n=4) and qPCR.** a) Luciferase assay: All samples were co-transfected with the plasmid siCHECK2-*SOX4*-3'UTR and a siRNA. Positive controls: siCHECK2-*SOX4*-3'UTR without any siRNA, siCHECK2-*SOX4*-3'UTR + scrbl-siRNA, siCHECK2-*SOX4*-3'UTR + MSI2-siRNA. Negative control: siCHECK2-*SOX4*-3'UTR + LUC-siRNA. The experiment was repeated four times (3x 3 technical replicates, 1x 6 technical replicates). This bar plot is representative of all four bar plots at 24 h and also at 48 h and 72 h b-d) qPCR analysis of *SOX4* expression after shRNA199 treatment. ΔΔCt values after b) 24 h, c) 48 h and d) 72 h of lipofection of pcDNA6.2-GW/EmGFP-miR-199a/b. EV = empty vector control.

Both, luciferase-assay results and qPCR-results showed no difference in *SOX4* expression after treatment with miR-199 after 24 h, 48 h and 72 h. Therefore, the predicted interaction between miR-199a/b and *SOX4*-3'UTR could not be verified. Because of that, a new computational correlation analysis of differentially expressed and potentially interacting miRNA-mRNA target pairs was performed. This correlation analysis was based on five public databases for miRNA-mRNA targeting (miRanda August 2010, miRDB V5, TarBase V7, TargetScan V6.2, miRTarBase V4.5). miRNA-mRNA target pairs, which were experimentally validated or predicted by at least two different databases with a p-value of less than 0.5, are listed in Table 4. With the described stricter search parameters only four potential interaction pairs were identified.

**Table 4: List of other potential miRNA-mRNA target pairs.**

miRNA	Symbol	Gene logFC	Gene p-val	Gene FDR	Gene t(1,19) FPKM	Gene t(17,19) FPKM
hsa-miR-888-3p	<i>RORB</i>	-6.74	7.01E-19	1.03E-15	15.64042784	0.020011184
hsa-miR-888-3p	<i>IQCE</i>	-1.90	0.000584485	0.024462821	6.654619006	1.637912221
hsa-miR-892a	<i>MSI2</i>	-1.51	1.69E-06	0.000188507	35.38680755	12.02764037
hsa-miR-892a	<i>CARHSP1</i>	-0.68	0.008579933	0.197847681	25.77418171	15.55220108

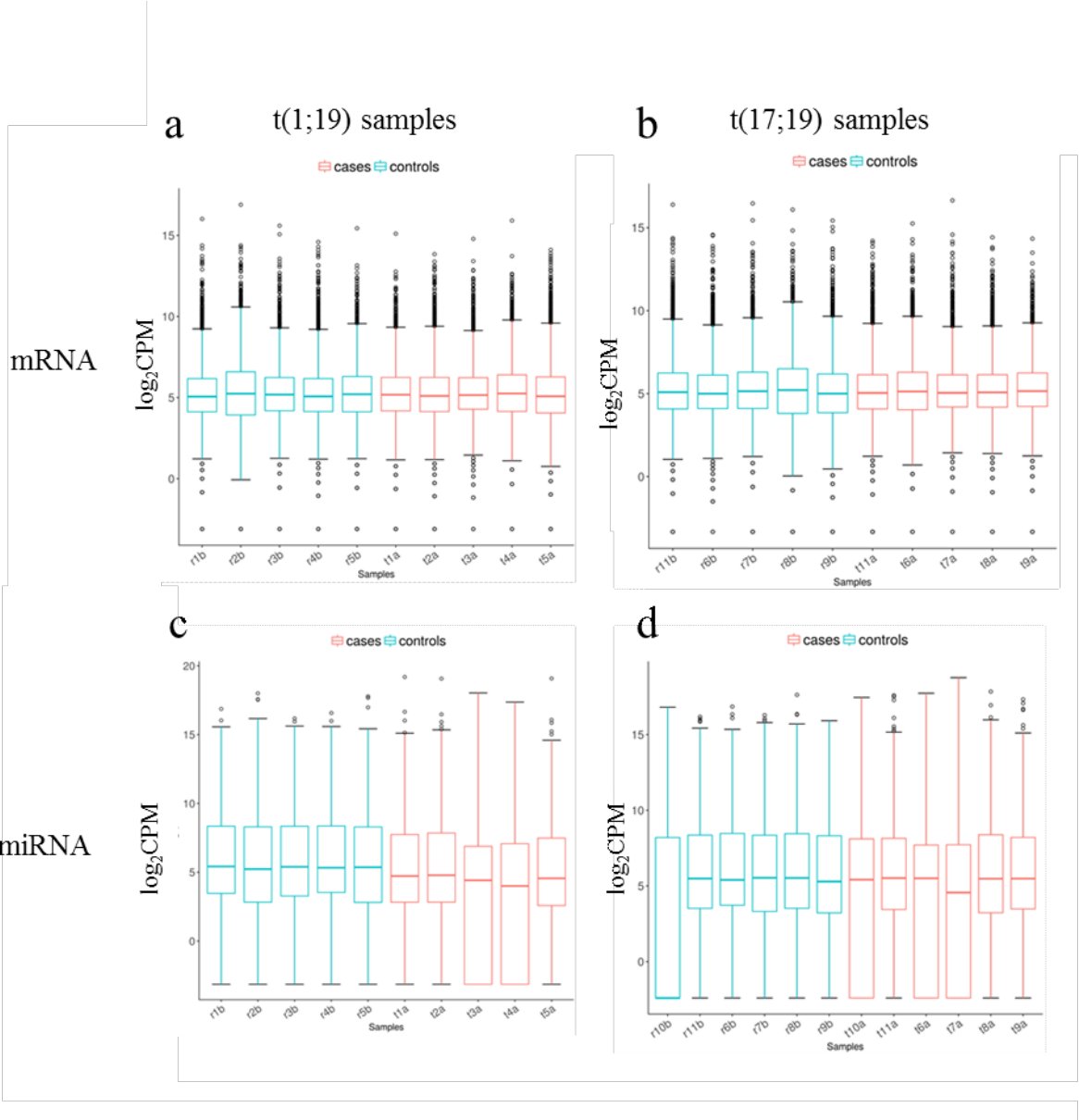
LogFC = logarithmic fold change; p-val = p-value; FDR = false discovery rate; FPKM = Fragments per kilobase of exon per million fragments mapped.

The miR-888 cluster is differentially expressed between t(1;19) and t(17;19)-ALL. One of the four described targets is relevant for cancer progression: Musashi RNA Binding Protein 2 (*MSI2*). To validate the predicted interaction of *MSI2* and hsa-miR-892a a luciferase assay was performed. *MSI2*-3'UTR was cloned into psiCHECK2. HEK293T cells were transfected/co-transfected in four different ways: 1) psiCHECK2-*MSI2*-3'UTR, 2) psiCHECK2-*MSI2*-3'UTR + LUC-siRNA (positive control), 3) psiCHECK2-*MSI2*-3'UTR + Negative-CTRL-mimic and 4) psiCHECK2-*MSI2*-3'UTR + miRNA-892a-mimic. Renilla and firefly luminescence were measured and relative normalized renilla luminescence was calculated. No effect of mimic-892a on *MSI2* expression could be observed after 24 h (n=3), 48 h (n=3) and 72 h (n=2). To validate these results HEK293T cells were transfected with miRNA-199a-mimic. RNA and proteins were extracted for qPCR and Western blot analysis. Both analyses showed no reduction of *MSI2* expression after treatment with miRNA-892a-mimic. Luciferase-assay results, qPCR results as well as Western blot results showed no difference in *MSI2* expression after treatment with miRNA-892a-mimic after 24 h, 48 h and 72 h. Therefore, the predicted interaction between miR-892a and *MSI2*-3'UTR could also not be verified.

#### 4.1.2.2 Re-Analysis of Patient and Germline Material

mRNA-seq and miRNA-seq rawdata were inserted into a new pipeline for RNA-seq data analysis. Boxplots of Figure 13 show the read count distribution of mRNA-seq and miRNA-seq data. Figure 13a shows the read count distribution of t(1;19)-tumor and germline samples. All samples have a symmetrical distribution around the median, which lies around a logarithmic counts per million value ( $\text{Log}_2\text{CPM}$ ) of 5. mRNA-seq of tumor and germline samples of t(17;19)-ALL subtype show a similar distribution (Figure 13b). The distribution of read counts of miRNAs of t(1;19)-tumor and germline samples also shows a symmetrical distribution around 5  $\text{Log}_2\text{CPM}$  with exception of tumor samples 3a and 4a. In both samples the distribution is postponed to a low number of read counts while the median lies still around 5  $\text{Log}_2\text{CPM}$  (

Figure 13c). The distribution of read counts of t(17;19)-tumor and germline samples shows a symmetrical distribution around the median of 5 Log<sub>2</sub>CPM in the most samples. Tumor samples 6a, 7a and 10a show an asymmetrical distribution which is again postponed to lower Log<sub>2</sub>CPM values with a median around 5 Log<sub>2</sub>CPM. In sample 10b Log<sub>2</sub>CPM values reach from < 0 to around 8 with a median of < 0 (Figure 13d). An asymmetrical read count distribution indicates low qualitative samples, which likely do not represent the correct miRNA expression status. Including the samples into further analysis would thus lead to erroneous results.



**Figure 13 Read count distribution** of a/b: mRNA-seq data, c/d: miRNA-seq data a: Distribution of Log<sub>2</sub>CPM values of RNA-seq is symmetrical with a median around 5 Log<sub>2</sub>CPM in all t(1;19)-tumor and -germline samples . b) Distribution of Log<sub>2</sub>CPM values of RNA-seq is symmetrical with a median around 5 Log<sub>2</sub>CPM in all t(17;19)-tumor and -germline samples. c) Distribution of Log<sub>2</sub>CPM values of miRNA-seq data is symmetrical with a median around 5 Log<sub>2</sub>CPM in nearly all t(1;19)-tumor and -germline samples. Exceptions are tumor samples 3a

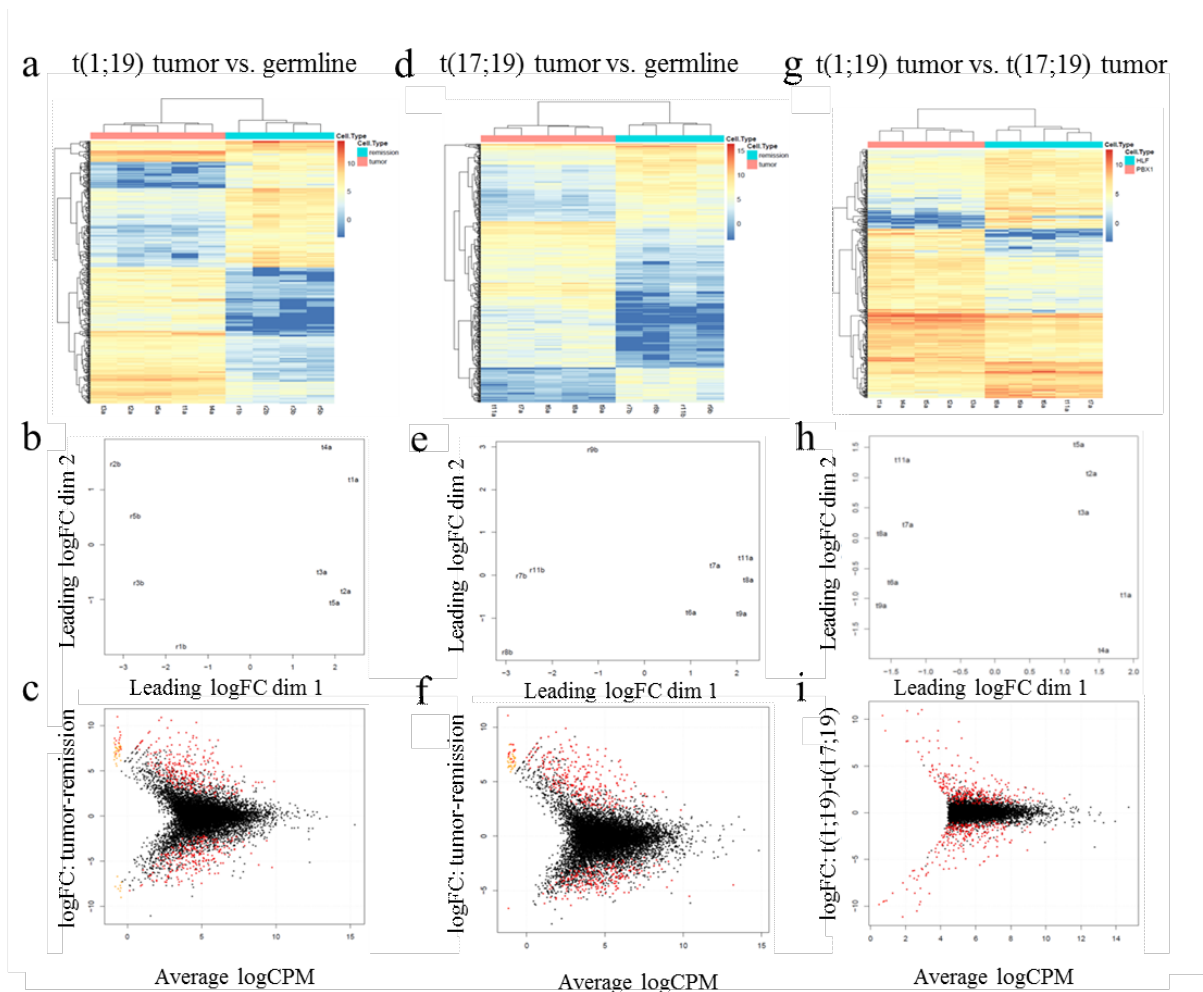


and 4a with an asymmetrical distribution, which is postponed to lower  $\text{Log}_2\text{CPM}$ -values. d) Distribution of  $\text{Log}_2\text{CPM}$  values of miRNA-seq data is symmetrical with a median around 5  $\text{Log}_2\text{CPM}$  in nearly some t(17;19)-tumor and -germline samples. Exceptions are tumor samples 6a, 7a and 10a with an asymmetrical distribution, which is postponed to lower  $\text{Log}_2\text{CPM}$ -values. Germline sample 10b has an asymmetrical distribution around the media of  $< 0$  from  $\text{Log}_2\text{CPM}$  values of  $< 0$  till approximately 8.

An asymmetrical distribution of  $\text{Log}_2\text{CPM}$ -values, which are shifted to lower  $\text{Log}_2\text{CPM}$ -values in the RNA- or miRNA-seq data, would lead to falsified results in the analysis of differential expressed genes or miRNAs. For this reason miRNA-seq data analysis was again performed with exclusion of tumor samples 4a, 6a, 7a and 10a and germline sample 10b. Furthermore, germline samples 4b and 6b were excluded from mRNA-seq and miRNA-seq analysis because patients were not in complete remission at the time of sampling. Furthermore two different filter settings were used to analyze mRNA- and miRNA-seq data. The first analysis approach included all genes/miRNAs with a  $\text{CPM} \geq 1$  in 10% of the samples in one cohort. The second and more strict analysis approach included all genes/miRNAs with a  $\text{CPM} \geq 1$  in 90% of the samples to exclude transcripts from further analysis, which are potentially not sequenced right. This method tries to exclude transcripts with low quality. In this second approach tumor samples of one ALL-subtype were compared to all germline samples.

### **Differentially expressed genes with analysis method 1 (10% of samples $\text{CPM} \geq 1$ ) and 2 (90% of samples $\text{CPM} \geq 1$ )**

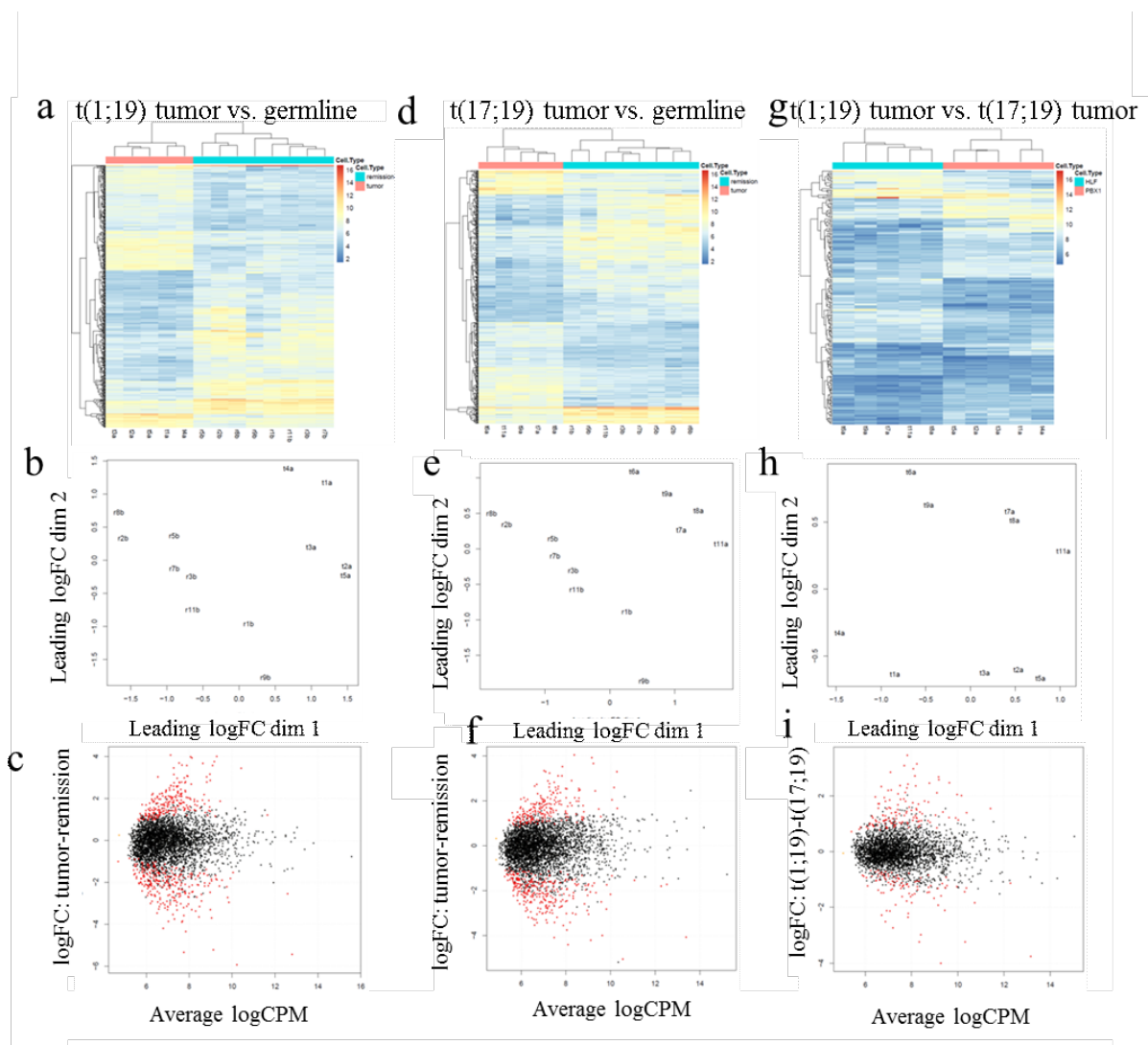
Analysis of differential gene expression between t(1;19)-tumor and their accompanied germline samples with analysis approach 1 revealed 2371 genes differentially expressed between both cohorts ( $|\log\text{FC}| \geq 2$ ,  $\text{FDR} < 0.1$ ). 2170 genes were differentially expressed between t(17;19)-tumor and their accompanied germline samples ( $|\log\text{FC}| \geq 2$ ,  $\text{FDR} < 0.1$ ) and 452 genes were differentially expressed between t(1;19)-tumor and t(17;19)-tumor samples ( $|\log\text{FC}| \geq 2$ ,  $\text{FDR} < 0.1$ ). By hierarchical clustering with the top 500 differentially expressed genes both entities were separated from each other in all three cases (Figure 14a-c). Multidimensional scaling plots of distances showed the same results (Figure 14d-f). The higher the difference between tumor and germline or t(1;19)- tumor and t(17;19)-tumor in gene expression of single genes the lower the average  $\text{CPM}$  values (Figure 14g-i).



**Figure 14: Differential gene expression between t(1;19)-tumor and germline samples (a, d, g) between t(17;19)-tumor and germline samples (b, e, h) and between t(1;19) and t(17;19) tumor samples (c, f, i) with exclusion of patient sample 10a and germline samples 4a, 6a and 10a. a-c:** Heatmaps of hierarchical clustering of the top 500 differentially expressed genes separate two cohorts from each other. Pink: tumor samples, light blue: remission samples. d-e: multidimensional scaling plots of distances of the top 500 differentially expressed genes separate two cohorts from each other. g-i: Mean difference plots for the top 500 differentially expressed genes.

With usage of the more stringent filter setting of analysis approach 2, 232 genes were differentially expressed between t(1;19)-tumor and all germline samples ( $|\logFC| \geq 2$ ,  $FDR < 0.1$ ). 214 genes were differentially expressed between t(17;19)-tumor and all germline samples ( $|\logFC| \geq 2$ ,  $FDR < 0.1$ ) and 33 genes were differentially expressed between t(1;19)-tumor and t(17;19)-tumor samples ( $|\logFC| \geq 2$ ,  $FDR < 0.1$ ). By hierarchical clustering with the top 500 differentially expressed genes both entities were separated from each other in every case (Figure 15a-c). Multidimensional scaling plots of distances showed the same results (Figure 15d-f). The higher the difference between tumor and germline or t(1;19)-tumor and t(17;19)-tumor in gene expression of single genes the lower the average CPM

values (Figure 15g-i).

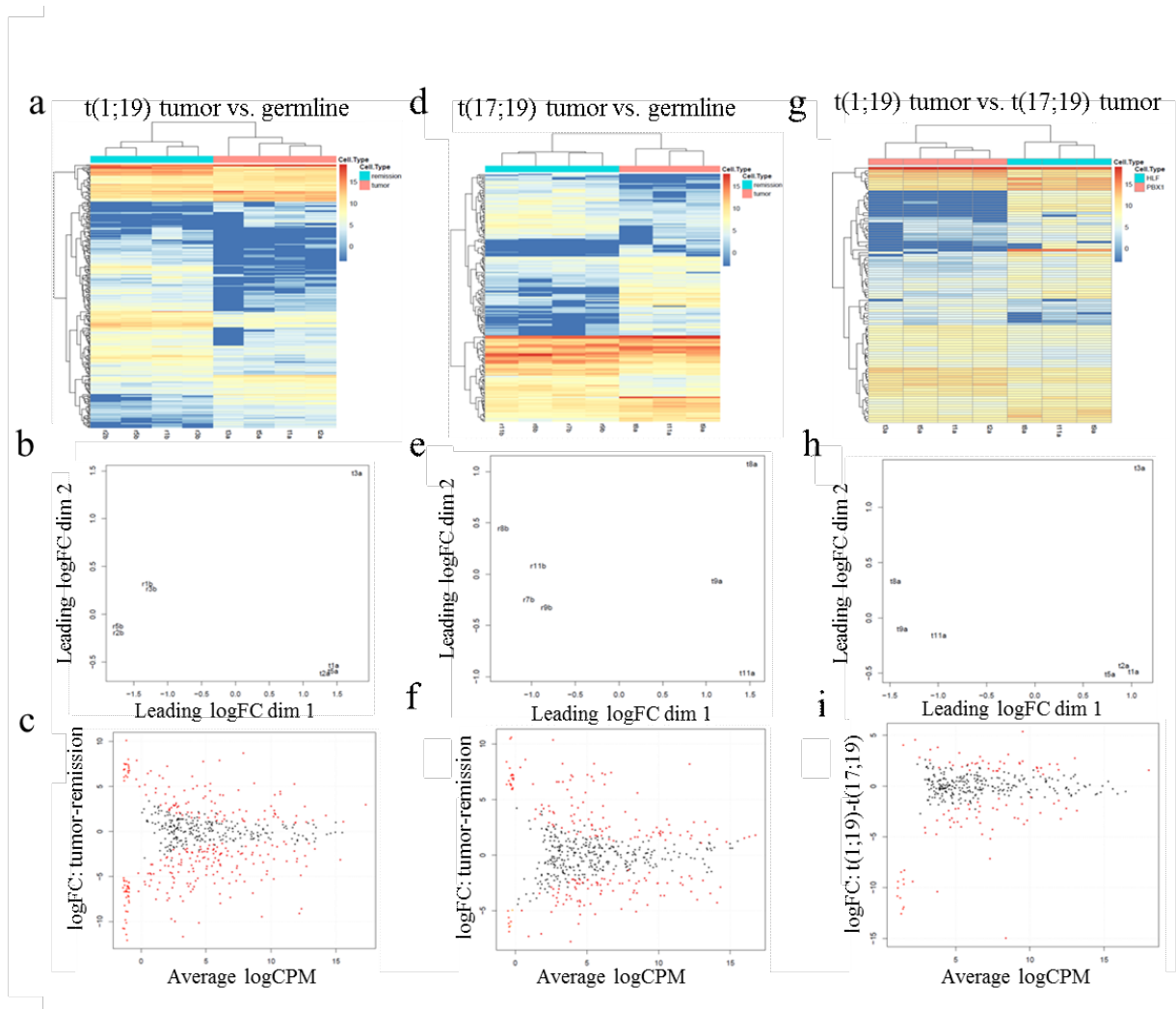


**Figure 15: Differential gene expression between t(1;19)-tumor and all germline samples (a, d, g) between t(17;19)-tumor and all germline samples (b, e, h) and between t(1;19) and t(17;19) tumor samples (c, f, i) with the use of a more stringent filter setting. a-c: Heatmaps of hierarchical clustering of the top 500 differentially expressed genes separate two cohorts from each other. d-e: multidimensional scaling plots of distances of the top 500 differentially expressed genes separate two cohorts from each other. g-i: Mean difference plots for the top 500 differentially expressed genes.**

## Differentially expressed miRNAs with analysis method 1 and 2

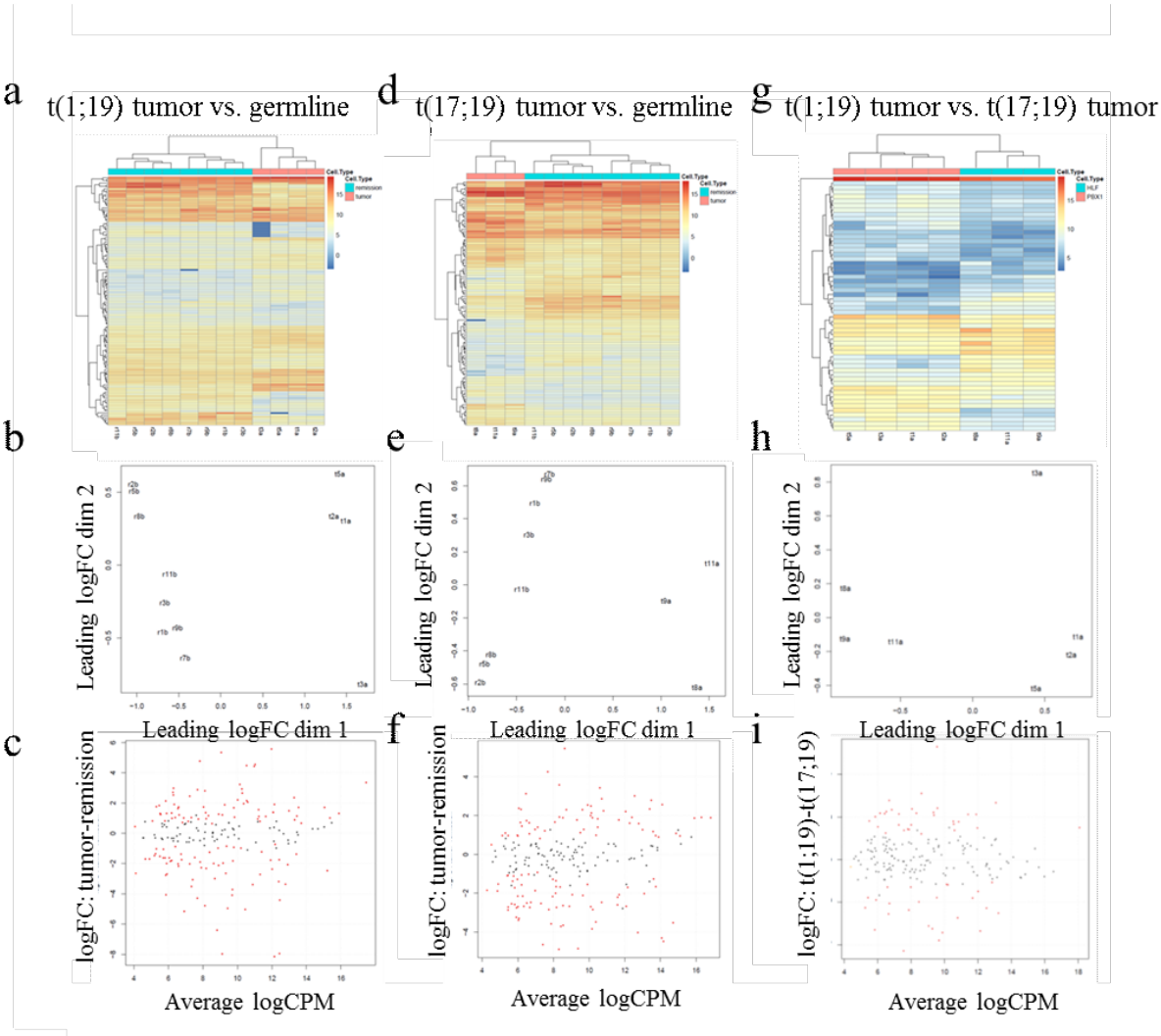
Analysis of differential miRNA expression between t(1;19)-tumor and their accompanied germline samples with analysis approach 1 and exclusion of the before described samples revealed 230 miRNAs differentially expressed between both cohorts ( $|\logFC| \geq 2$ ,  $FDR < 0.05$ ). 152 miRNAs were differentially expressed between t(17;19)-tumor and their accompanied germline samples ( $|\logFC| \geq 2$ ,  $FDR < 0.05$ ) and 68 miRNAs were

differentially expressed between t(1;19)-tumor and t(17;19)-tumor samples ( $|\logFC| \geq 2$ , FDR < 0.05). By hierarchical clustering with the top 500 differentially expressed genes both entities were separated from each other in all three cases (Figure 16a-c). Multidimensional scaling plots of distances showed the same results (Figure 16d-f). The higher the difference between tumor and germline or t(1;19)-tumor and t(17;19)-tumor in gene expression of single genes the lower the average CPM values (Figure 16g-i).



**Figure 16: Differential miRNA expression between t(1;19)-tumor and germline samples (a, d, g) between t(17;19)-tumor and germline samples (b, e, h) and between t(1;19) and t(17;19) tumor samples (c, f, i) with exclusion of tumor samples 4a, 6a, 7a and 10a and germline samples 4b, 6b and 10b and the usage of analysis approach 1. a-c: Heatmaps of hierarchical clustering of the top 500 differentially expressed genes separate two cohorts from each other. d-e: multidimensional scaling plots of distances of the top 500 differentially expressed miRNAs separate two cohorts from each other. g-i: Mean difference plots for the top 500 differentially expressed genes.**

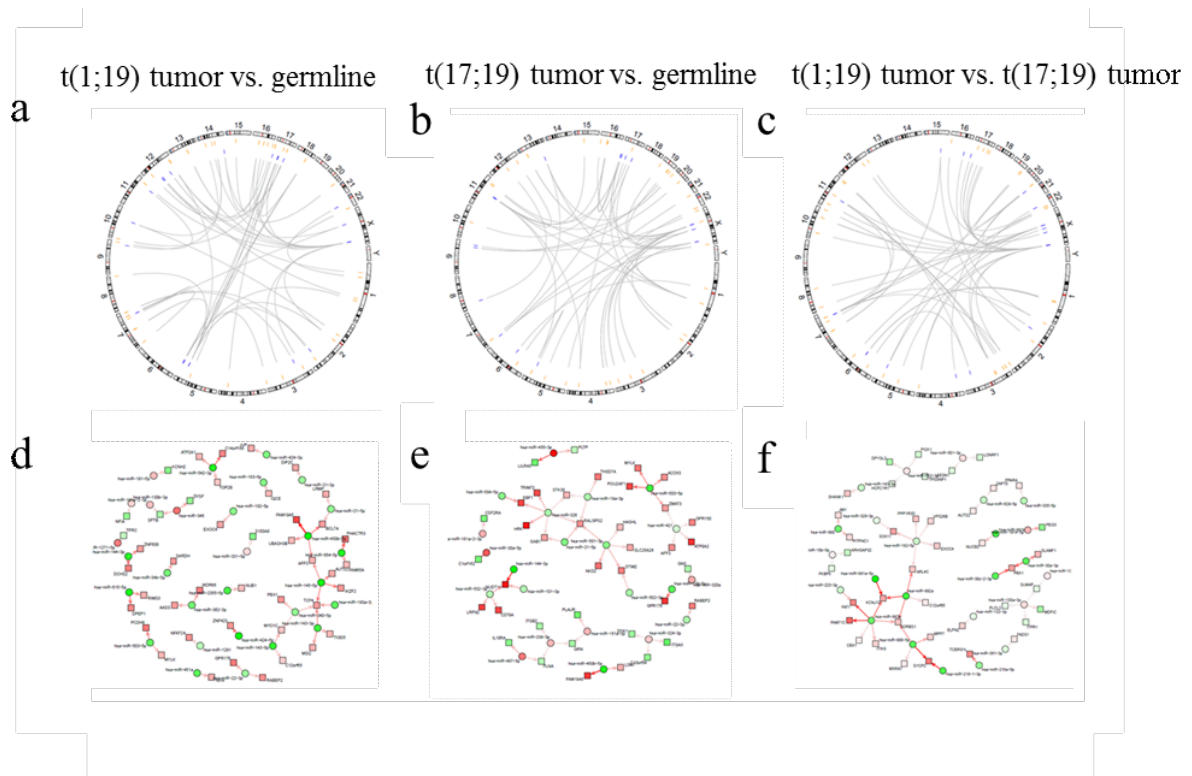
Analysis of differential miRNA expression between t(1;19)-tumor and all germline samples performed with analysis approach 2 revealed 54 genes differentially expressed between both cohorts ( $|\logFC| \geq 2$ ,  $FDR < 0.05$ ). 62 miRNAs were differentially expressed between t(17;19)-tumor and all germline samples ( $|\logFC| \geq 2$ ,  $FDR < 0.05$ ) and 31 miRNAs were differentially expressed between t(1;19)-tumor and t(17;19)-tumor samples ( $|\logFC| \geq 2$ ,  $FDR < 0.05$ ). By hierarchical clustering with the top 500 differentially expressed miRNAs both entities were separated from each other in all three cases (Figure 17a-c). Multidimensional scaling plots of distances showed the same results (Figure 17d-f). Difference between tumor and germline or t(1;19)-tumor and t(17;19)-tumor in gene expression of single genes and the average CPM values equally distributed. miRNAs with low expression difference between two cohorts have either a big average CPM or a low one. The same is true for miRNAs with a high expression difference between two cohorts (Figure 17g-i).



**Figure 17: Differential miRNA expression between t(1;19)-tumor and germline samples (a, d, g) between t(17;19)-tumor and germline samples (b, e, h) and between t(1;19) and t(17;19) tumor samples (c, f, i) with exclusion of tumor samples 4a, 6a, 7a and 10a and germline samples 4b, 6b and 10b and the usage of analysis approach 2.** a-c: Heatmaps of hierarchical clustering of the top 500 differentially expressed genes separate two cohorts from each other. d-e: multidimensional scaling plots of distances of the top 500 differentially expressed miRNAs separate two cohorts from each other. g-i: Mean difference plots for the top 500 differentially expressed genes.

### **Target prediction and correlation analysis of differentially expressed genes of method 1**

Target prediction with the above shown differentially expressed mRNAs and the above shown differentially expressed miRNAs was performed as before with TargetScan v6.2-20 and Microcosm\_5\_20 to find potentially mRNA-miRNA target pairs. Circos plots, which show the chromosomal distribution of potential target pairs, were generated. For t(1;19)-tumor and germline samples (Figure 18a), for t(17;19)-tumor and germline samples (Figure 18b) and for t(1;19) and t(17;19) (Figure 18c) samples the top 50 potential interaction were represented in the circos plots. Both parts of potential interactions of t(1;19)-tumor and germline samples and t(17;19)-tumor and germline samples are distributed around the whole set of chromosomes with exception of the Y-chromosome, from which no potential interactions originate. An evenly distributed gender distribution is given. Chromosome 17 has a higher density of located genes potentially interacting with differentially expressed miRNA. Potential target pairs of both tumor subtypes show an accumulation of miRNAs located on chromosome X. Potential mRNA-miRNA targeting networks of the top 50 pairs of comparison between t(1;19)-tumor and germline samples (Figure 18d), the comparison between t(17;19)-tumor and germline samples (Figure 18e) and t(1;19) and t(17;19) samples (Figure 18f) were generated.



**Figure 18: mRNA-miRNA correlation data.** a – c: Circos plots for the top 50 mRNA-miRNA pairs, sorted by adjusted p-value, with a corrected p-value lower 0.1 and an appearance in microCosm\_5\_20 or targetScan\_v6.2-20. Blue: miRNA, orange: mRNA. All three circos plots show a miRNA-mRNA target distribution around the whole genome. d – f: Potential mRNA-miRNA target network for a maximum of 50 mRNA-miRNA pairs that have a corrected p-value of lower then 0.1 and appear at least one time in the following databases: microcosm\_v5\_20 or TargetScan\_v6.2\_20. Circles: miRNAs, squares: mRNAs. Red: upregulated, green: downregulated. Arrows represent a potential mRNA-miRNA target pair.

Correlation analysis of the results of analysis method 1 revealed a list of 31 potential mRNA-miRNA target pairs (correlation value  $\leq -0.90$ , p-value  $< 0.001$ , quality score  $> 25$ ). Out of them, two were of interest in tumor development (**Fehler! Ungültiger Eigenverweis auf Textmarke.**, marked in yellow). Hsa-miR-223 and its potential target *FAT1* were chosen for validation with qPCR.

**Table 5: Potential mRNA-miRNA target pairs with a correlation value  $\leq -0.90$  a p-value  $< 0.001$  and a quality score  $> 25$ .**

miRNA	mRNA	cor	pval	meanExp. miRNA	meanExp. mRNA	dat.micro Cosm_v5_20	dat.target Scan_v6.2_20	score
hsa-miR-892a	<i>KCNJ12</i>	-0.96	0.0004	427	31	1	0	154
hsa-miR-892a	<i>SORBS1</i>	-0.95	0.0006	427	31	1	0	98
hsa-miR-892a	<i>ARL4C</i>	-0.99	0.0000	427	165	0	1	99
hsa-miR-891a-5p	<i>KCNJ12</i>	-0.94	0.0008	24678	31	1	0	205
hsa-miR-891b	<i>MYT1L</i>	-0.94	0.0009	302	17	0	1	124
hsa-miR-891b	<i>ATRNL1</i>	-0.95	0.0006	302	19	0	1	93



hsa-miR-891b	<i>NUCB2</i>	-0.94	0.0009	302	29	1	0	39
hsa-miR-30c-2-3p	<i>PBX1</i>	-0.99	0.0000	126	130	1	0	139
hsa-miR-888-5p	<i>SYCP2</i>	-0.94	0.0007	148	457	1	0	198
hsa-miR-888-5p	<i>MPP7</i>	-0.99	0.0000	148	28	1	0	69
hsa-miR-888-5p	<i>SORBS1</i>	-0.95	0.0004	148	31	1	0	85
hsa-miR-30a-3p	<i>PBX1</i>	-0.96	0.0003	215	130	1	0	135
hsa-miR-30a-3p	<i>SLAMF1</i>	-0.95	0.0004	215	25	1	0	96
hsa-miR-892b	<i>FAT1</i>	-0.98	0.0000	47	280	0	1	156
hsa-miR-892b	<i>ITIH3</i>	-0.98	0.0001	47	158	1	0	57
hsa-miR-892b	<i>FAM71C</i>	-0.97	0.0001	47	165	1	0	197
hsa-miR-892b	<i>IGSF3</i>	-0.96	0.0003	47	12	0	1	116
hsa-miR-892b	<i>KCNJ12</i>	-0.96	0.0003	47	31	1	0	116
hsa-miR-892b	<i>SORBS1</i>	-0.95	0.0006	47	31	1	0	74
hsa-miR-892b	<i>CECR2</i>	-0.97	0.0002	47	11	1	0	62
hsa-miR-890	<i>PITPNC1</i>	-0.94	0.0007	40	77	0	1	45
hsa-miR-890	<i>JMY</i>	-0.96	0.0003	40	22	0	1	42
hsa-miR-182-5p	<i>SOX11</i>	-0.97	0.0001	861	30	0	1	41
hsa-miR-182-5p	<i>PPP1R3D</i>	-0.94	0.0008	861	42	0	1	25
hsa-miR-182-5p	<i>ARL4C</i>	-0.94	0.0008	861	165	0	1	25
hsa-miR-143-3p	<i>SHANK1</i>	-0.94	0.0007	1314	24	1	0	28
hsa-miR-29a-3p	<i>MYBPH</i>	-0.97	0.0002	6317	12	1	0	31
hsa-miR-223-3p	<i>FAT1</i>	-0.95	0.0006	1399	280	0	1	51
hsa-miR-216a-5p	<i>TCERGIL</i>	-0.95	0.0005	23	28	0	1	105
hsa-miR-216a-5p	<i>VWCE</i>	-0.98	0.0001	23	12	1	0	27
hsa-miR-130a-3p	<i>MDF1C</i>	-0.97	0.0002	451	31	0	1	28

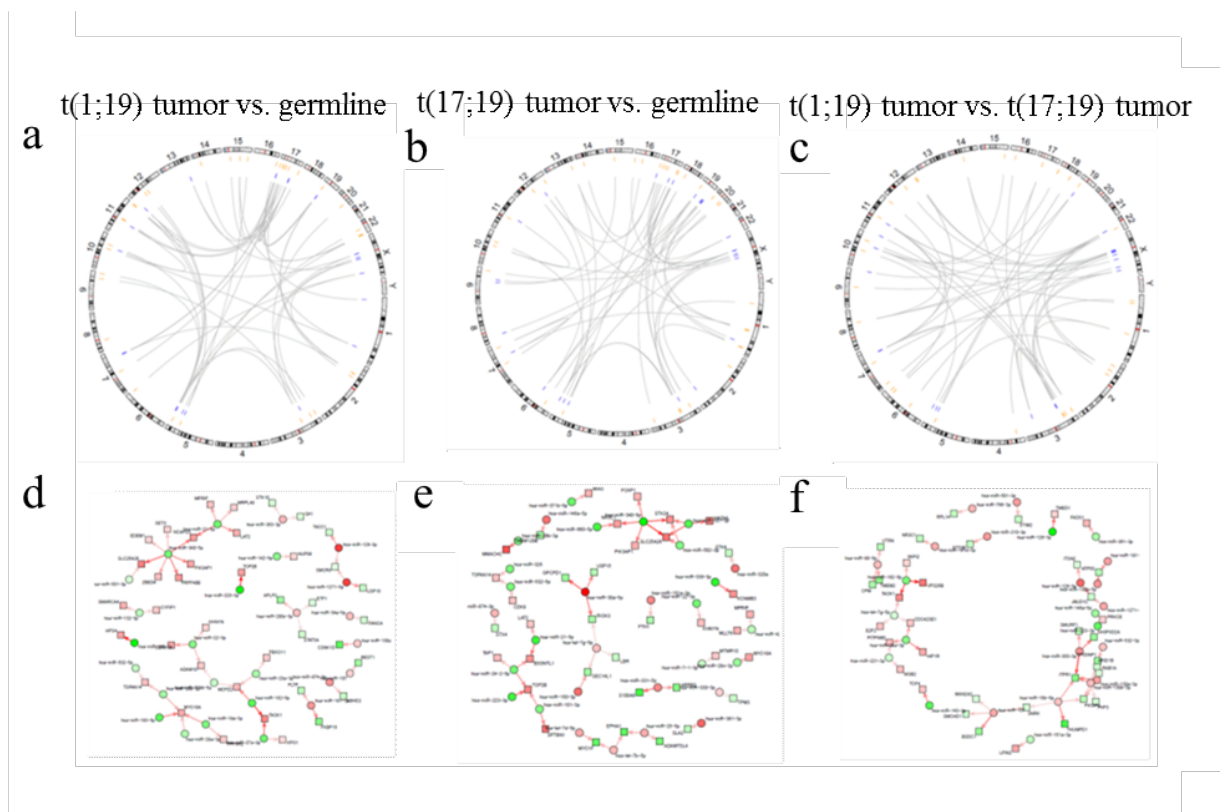
Target pairs marked in yellow are interesting mRNA/miRNA pairs with respect to tumor development. Cor = correlation value; meanExp. = mean expression.

## Target prediction and correlation analysis of differentially expressed genes of method 2

Target prediction of the differentially expressed mRNA and miRNAs resulting of method 2 was performed like in method 1 with TargetScan v6.2-20 and Microcosm\_5\_20. Circos plots, which show the chromosomal distribution of potential target pairs, were generated. For t(1;19)-tumor and germline samples (Figure 19a), for t(17;19)-tumor and germline samples (Figure 19b) and for t(1;19) and t(17;19) (Figure 19c) samples the top 50 potential interaction were represented in the circos plots. Both parts of potential interactions of t(1;19)-tumor and germline samples and t(17;19)-tumor and germline samples are distributed around the whole



set of chromosomes with exception of the Y-chromosome, from which no potential interactions originate. An evenly distributed gender distribution is given. Chromosome 17 has a higher density of located genes potentially interacting with differentially expressed miRNA comparing both tumor subtypes with the germline samples. Potential target pairs of both tumor subtypes show a certain accumulation of miRNAs located on chromosome X. Potential mRNA-miRNA targeting networks of the top 50 pairs of comparison between t(1;19)-tumor and germline samples (Figure 19d), the comparison between t(17;19)-tumor and germline samples (Figure 19e) and t(1;19) and t(17;19) samples (Figure 19f) were designed.



**Figure 19: mRNA-miRNA correlation data.** a – c: Circos plots for the first 50 mRNA-miRNA pairs, sorted by adjusted p-value, with a corrected p-value lower 0.1 and an appearance in microcosm\_5\_20 or targetScan\_v6.2-20. Blue: miRNA, orange: mRNA. All three circos plots show a miRNA-mRNA target distribution around the whole genome with a higher impact of chromosome 17. d – f: Potential mRNA-miRNA target network for a maximum of 50 mRNA-miRNA pairs that have a corrected p-value of lower then 0.1 and appear at least one time in the following databases: microcosm\_v5\_20 or TargetScan\_v6.2\_20. Circles: miRNAs, squares: mRNAs. Red: upregulated, green: downregulated. Arrows represent a potential mRNA-miRNA target pair.

Correlation analysis of the results of analysis method 2 revealed a list of 43 potential mRNA-miRNA target pairs with a correlation value  $< -0.75$  and a p-value  $< 0.005$ . Out of them, four are of potential interest in tumor development (Table 6, marked in yellow). Hsa-let-7g and its potential target *E2F2* was chosen for further validation.

**Table 6: Potential mRNA-miRNA target pairs with a correlation value < -0,75 and p-value < 0.005.**

miRNA	mRNA	cor	pval	meanExp. miRNA	meanExp. mRNA	dat.micro Cosm_v5_20	dat.target Scan_v6.2_20	score
hsa-miR-182-5p	<i>VPS26B</i>	-0.90	0.0002	837	424	0	1	15
hsa-miR-182-5p	<i>TAOK1</i>	-0.85	0.0010	837	416	0	1	13
hsa-miR-143-3p	<i>TCF4</i>	-0.78	0.0039	1625	368	1	0	8
hsa-miR-126-3p	<i>TMED1</i>	-0.81	0.0024	325	97	1	0	10
hsa-miR-223-3p	<i>PRKCE</i>	-0.82	0.0020	1539	76	0	1	6
hsa-miR-29a-3p	<i>HIP1R</i>	-0.84	0.0012	6066	315	1	0	8
hsa-miR-29a-3p	<i>PITPNM2</i>	-0.81	0.0021	6066	120	0	1	6
hsa-miR-29a-3p	<i>CDC42SE1</i>	-0.90	0.0002	6066	157	0	1	6
hsa-miR-501-3p	<i>STIM2</i>	-0.89	0.0003	974	154	0	1	4
hsa-miR-130a-3p	<i>ITPRI</i>	-0.87	0.0006	333	164	0	1	9
hsa-miR-130a-3p	<i>TP53INP1</i>	-0.85	0.0008	333	194	0	1	8
hsa-miR-130a-3p	<i>PIK3IP1</i>	-0.89	0.0003	333	125	0	1	5
hsa-miR-130a-3p	<i>ARID1B</i>	-0.83	0.0016	333	175	1	0	4
hsa-miR-191-3p	<i>ATP5S</i>	-0.77	0.0043	273	223	1	0	6
hsa-miR-361-3p	<i>FADS1</i>	-0.87	0.0006	6893	175	1	0	5
hsa-miR-128-3p	<i>ITSN2</i>	-0.86	0.0007	9905	61	0	1	5
hsa-miR-128-3p	<i>JMJD1C</i>	-0.88	0.0004	9905	175	0	1	6
hsa-miR-130b-5p	<i>ATP5S</i>	-0.78	0.0038	4275	223	1	0	6
hsa-miR-766-3p	<i>RPL14</i>	-0.81	0.0024	273	240	1	0	3
hsa-miR-363-3p	<i>ITPRI</i>	-0.88	0.0004	1105	164	1	1	11
hsa-miR-363-3p	<i>SH3PXD2A</i>	-0.85	0.0010	1105	534	0	1	9
hsa-miR-363-3p	<i>SMURF1</i>	-0.81	0.0023	1105	320	0	1	6
hsa-miR-107	<i>SNRK</i>	-0.83	0.0014	330	108	0	1	3
hsa-miR-107	<i>MAN2A2</i>	-0.80	0.0029	330	145	1	0	3
hsa-miR-107	<i>SMCHD1</i>	-0.91	0.0001	330	105	0	1	4
hsa-miR-130b-3p	<i>ITPRI</i>	-0.79	0.0031	1497	164	0	1	7
hsa-miR-130b-3p	<i>TP53INP1</i>	-0.83	0.0013	1497	194	0	1	5
hsa-miR-130b-3p	<i>PIK3IP1</i>	-0.87	0.0005	1497	125	0	1	4
hsa-miR-130b-3p	<i>PHF3</i>	-0.81	0.0024	1497	98	0	1	2
hsa-miR-130b-3p	<i>RAB1A</i>	-0.85	0.0009	1497	90	0	1	2
hsa-miR-1271-5p	<i>JMJD1C</i>	-0.79	0.0032	1329	175	0	1	3
hsa-miR-98-5p	<i>UTRN</i>	-0.82	0.0020	3558	301	0	1	4

hsa-miR-98-5p	<i>CPM</i>	-0.79	0.0034	3558	716	0	1	5
hsa-miR-15b-5p	<i>ITPRI</i>	-0.80	0.0025	1739	164	1	0	6
hsa-miR-15b-5p	<i>THUMPD1</i>	-0.82	0.0019	1739	1308	0	1	9
hsa-miR-15b-5p	<i>PIK3IP1</i>	-0.89	0.0003	1739	125	1	0	3
hsa-miR-15b-5p	<i>SNRK</i>	-0.87	0.0005	1739	108	0	1	2
hsa-miR-151a-3p	<i>LPIN2</i>	-0.82	0.0019	3648	160	1	0	5
hsa-miR-221-3p	<i>PITPNM2</i>	-0.78	0.0039	7327	120	0	1	3
hsa-miR-221-3p	<i>WSB2</i>	-0.77	0.0048	7327	150	1	1	3
hsa-miR-181d-5p	<i>NR3C1</i>	-0.82	0.0019	175	205	0	1	4
hsa-let-7g-5p	<i>TAOK1</i>	-0.80	0.0027	24710	416	0	1	4
hsa-let-7g-5p	<i>E2F2</i>	-0.79	0.0034	24710	196	0	1	2

Target pairs marked in yellow were chosen as interesting mRNA-miRNA target pairs for tumor development. Cor = correlation value; meanExp. = mean expression.

### Experimental validation of potential mRNA-miRNA target pairs

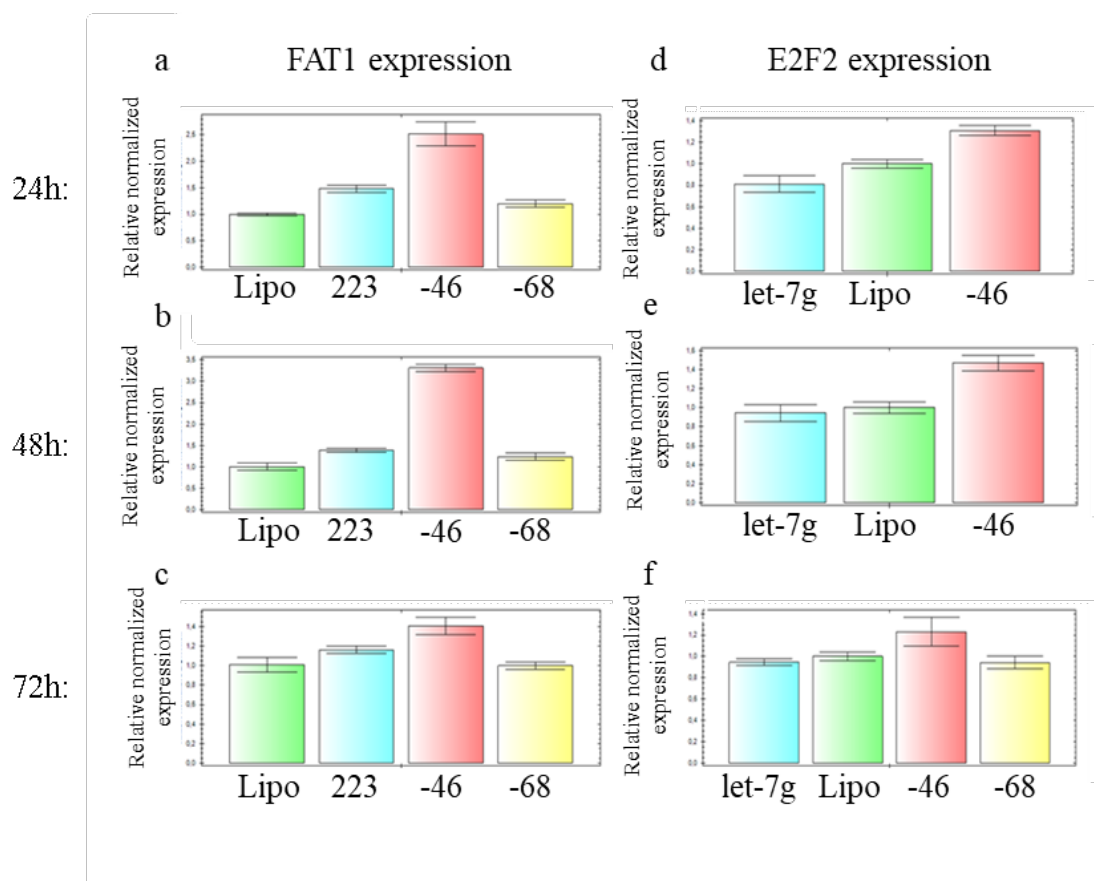
FAT1 with hsa-miR-223 as well as E2F2 and hsa-let-7g were chosen for further validation.

FAT1 is a tumor suppressor important for proliferation, but upregulated in leukemia, which is correlated with a poor prognosis of pre-B-ALL<sup>112</sup> (review by Katoh, 2012). The transcription factor E2F2 is important for cell cycle control and proliferation ADDIN EN.CITE<sup>113</sup>.

Validation was performed by qPCR. siRNAs, which mimic hsa-miR-223 and hsa-let-7g (both fluorescence-labeled) were designed. HEK293T cells were transfected with either hsa-miR-223-siRNA, hsa-let-7g, Neg-CTRL-46-siRNA (a siRNA negative control with a GC content of 46%), Neg CTRL 68-siRNA (a siRNA negative control with a GC content of 68%) or with hsa-miR-130a-Mimic (a non-fluorescence labeled miRNA-Mimic for analysis of transfection efficiency by FACS). Just lipofectamine treated HEK293T cells were also cultivated for usage as a negative control. Cells were harvested and RNA was extracted after 24 h, 48 h and 72 h after transfection. Seventy-two hours after transfection, cells were analysed by flow cytometry to analyse the transfection efficiency. Because hsa-miR-130a-Mimic was not labeled with a fluorescent dye, HEK293T cells transfected with this Mimic were used to define the gates. Three days after transfection around 40% of HEK293T cells were positively transfected with hsa-miR-223-siRNA, around 22% of HEK293T cells were positively transfected with hsa-let-7g, around 29% of HEK293T cells were positively transfected with Neg CTRL 46-siRNA and 64% of HEK293T cells were positively transfected with Neg CTRL 68-siRNA. Optimization of transfection protocol revealed a transfection efficiency of over 90% 24 h after transfection.

Extracted RNA was reversed transcribed into cDNA and used for qPCR. *FAT1* expression was analyzed 24 h (Figure 20a), 48 h (Figure 20b) and 72 h (Figure 20c) after transfection of hsa-miR-223-siRNA. No downregulation of *FAT1* expression could be observed at all three timepoints, if compared to lipofectamine treated cells or Neg-CTRL-68%-siRNA-treated cells. *FAT1* expression in Neg-CTRL 46%-siRNA treated cells was increasing at all three timepoints compared to both other controls, what makes it useless in the usage as a negative control. For this reason Neg-CTRL-68%-siRNA was used as negative control.

*E2F2* expression was analyzed 24 h (Figure 20d), 48 h (Figure 20e) and 72 h (Figure 20f) after transfection of hsa-let-7g transfection. No downregulation of *E2F2* was observed after treatment with hsa-let-7g-siRNA compared to lipofectamine treated cells at all three time points. *E2F2* expression in Neg-CTRL 46%-siRNA treated cells was increasing at all three time points compared to both other controls. For this reason, it was not considered as a negative control.



**Figure 20: Expression of potential target genes after siRNA transfection in HEK293T.** A-C) Expression of *FAT1* after hsa-miR-223-siRNA treatment. No reduction of *FAT1* expression after treatment with hsa-miR-223-siRNA compared to lipofectamine treated cells and Neg CTRL 68-siRNA treated cells A) 24 h, B) 48 h and C) 72 h after siRNA treatment. Treatment with Neg CTRL 46-siRNA shows an increase in *FAT1* expression after 24 h, 48 h and 72 h. The experiment was performed in triplicates with n = 1. D-F) Expression of *E2F2* after hsa-miR-let-7g-siRNA treatment. No reduction of *E2F2* expression after treatment with hsa-let-7g-siRNA compared

to lipofectamine treated cells D) 24 h, E) 48 h and F) 72 h after siRNA treatment. The experiment was performed in triplicates with n = 1.

None of the two predicted mRNA-miRNA target pairs chosen for validation could be verified with qPCR.

### 4.1.3 Novel miRNAs

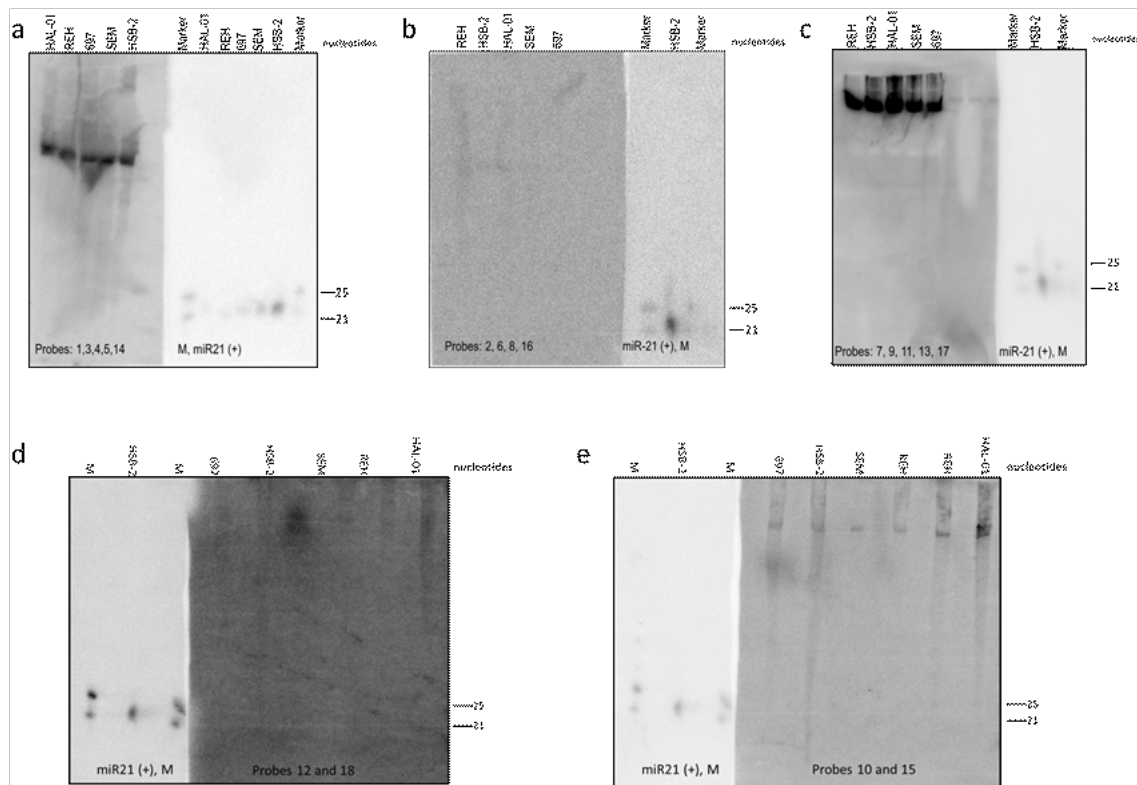
By miRNA seq around 10,000 potential novel miRNAs were found in at least one sample by using miRDeep. A set of 18 potential novel miRNAs was shortlisted for further validation (Table 7). These 18 potential novel miRNAs were exclusively or nearly exclusively found in tumor samples and were also found with more stringent filtering. The validation of these novel miRNAs was performed by Northern blot analysis. Because there was not enough RNA material of patient samples available, RNA of five different ALL cell lines (REH, 697, HAL-01, HSB-2 and SEM) was extracted and used for validation. Hsa-miR-21 was used as a positive control. A pool of two up to five probes for detection of novel miRNAs was used in one Northern blot experiment. All probes had a delta value of cross-interaction of more than -8 to exclude binding of probes to each other. All 18 potential novel miRNAs could not be detected in the employed cell lines (Figure 21a-e).

**Table 7: List of potential novel miRNAs.**

Name	sum of all scores	mean of all scores $\geq 3$	1a	2a	3a	4a	5a	6a	7a	8a	9a	10a	11a	13a	11c	13c
Novel-miR-1	30,6	5,1		5,1	5,1	5,1	5,1	5,1		5,1						
Novel-miR-2	244,4	48,9	33	26	9,5		5,9				0,6				0,6	170
Novel-miR-3	482	120,5	37				35						210			200
Novel-miR-4	12,9	4,3		4,3						4,3			4,3			
Novel-miR-5	16,8	5,6	5,7								5,7		5,4			
Novel-miR-6	15,8	5,3	5,4	5,4												5
Novel-miR-7	17,1	5,7	5,7		1,9		5,7								5,7	
Novel-miR-8	19,9	6,6					12			4,7			3,2			
Novel-miR-9	17,1	5,7		5,7						5,7	5,7					1,9
Novel-miR-10	137	45,7	69		16		52									
Novel-miR-11	15,6	5,2	3,9				5,4				6,3				1	
Novel-miR-12	34,2	5,7	5,7								5,7		5,7		5,7	5,7
Novel-miR-13	49,1	8,2	5,5	5,5	3,6								12		17	

Novel-miR-14	25,7	5,1							4,5	5				5,4	5,4
Novel-miR-15	235	47						12	26	53			34		
Novel-miR-16	57,5	11,5	3,5		0,6				16				12	13	
Novel-miR-17	18,9	4,7								5,1		5,1	5,1		
Novel-miR-18	341	85,3	90							65					92

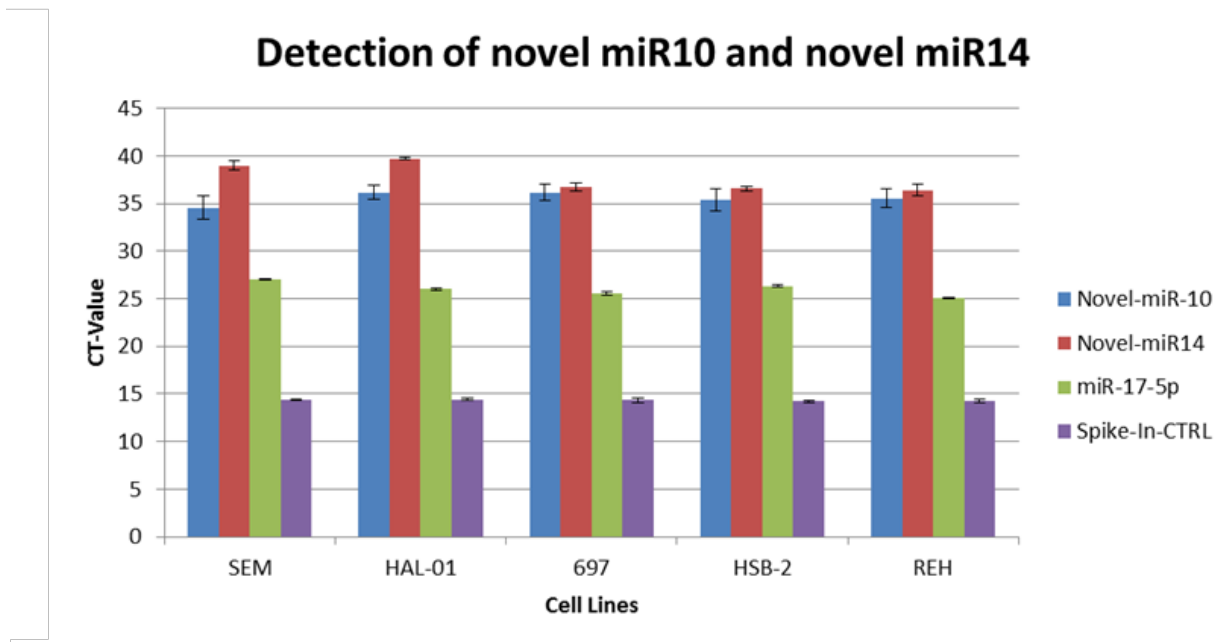
18 potential novel miRs with a miRDeep score  $\geq 3$  were found just or mostly in tumor samples and were shortlisted for validation experiments.



**Figure 21: Validation of novel miRs by Northern Blot analysis.** 30 $\mu$ g of total RNA of ALL cell lines (HAL-01, REH, 697, SEM, HSB-2) and 2  $\mu$ g of Marker were loaded on a polyacrylamide gel, separated and transferred to a membrane, to validate the expression of the 18 potential novel miRs. Membranes were treated with a pool of complementary nucleotides of a) novel miR1, 3, 4 and 5, b) novel miR2, 6, 8 and 16, c) novel miR7, 9, 11, 13, 17, d) novel miR12 and 18 e) novel miR10 and 15. miR21 was used as a positive control. None of the 18 potential novel miRs were detectable by Northern Blot analyses.

Because qPCR is a more sensitive method than Northern Blot, the expression of the two most interesting novel miRs were analyzed by qPCR. Novel miR-10 was exclusively found in patients with t(1;19)-translocation and novel miR-14 was exclusively found in patients with t(17;19)-translocation. Therefore, these two novel miRs were also checked by qPCR. For normalization, miR-17-5p was used as an endogenous control. A spike-in control was also used. However, again no expression of these two miRNAs could be detected in the five ALL

cell lines (Figure 22).



**Figure 22: Validation of novel miRs by qPCR:** Expression of the potential novel miRs10 and 14 in 5 ALL cell lines (SEM, HAL-01, 697, HSB-2 REH) was analyzed by qPCR. miR-17-5p as well as Spike-In-CTRL were used as controls. Both potential novel miRs were not detectable in the used cell lines by qPCR analysis.

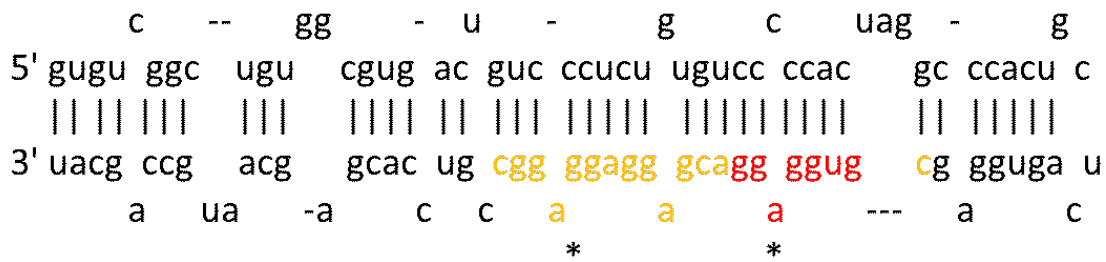
#### 4.1.4 Mutated miRNAs

To analyze miRNAs, which were mutated in patient samples, exome sequencing of tumor and germline samples was performed. The coverage of miRNA sequence obtained by exome sequencing is around 90%. Occurring mutations were analyzed with SNUPy (an in-house analysis tool for detection of mutations in exomic sequencing data). By using following filters: quality score < 50, read depth < 10, mutation frequency in germline samples = 0, mutation frequency in tumor samples = 5 v 6, eight listings were received. Five of these listings describe the mutation G/C on chromosome 7 at position 2297157 (GRCh37) in patient samples 2a, 3a, 7a, 8a and 9a. This mutation lies in the promoter region of hsa-miR-6836.

By using the filter setting: quality score < 50, read depth < 10, mutation frequency in germline samples = 0, mutation frequency in tumor samples = 4, 3 listings were received. They describe two mutations on chromosome 11 in the sequence region of hsa-miR-7847. Samples 1a and 2a harbor an A/C mutation on position 1901340, which lies in the miRNA seed region, and patient 3a harbors a G/C mutation at position 1901315 (Figure 23).



## hsa-miR-7847



**Figure 23: Mutations of hsa-miR-7847 in t(1;19)-positive cells.** Orange: sequence of hsa-miR-7847, red: seed sequence of hsa-miR-7847. Mutations are marked with an \*. Patient samples 1a and 2a show an A to C conversion inside the miRNA seed region. Patient sample 3a show a G to C conversion outside the miRNA seed region.

These mutations found by exome sequencing were re-examined by miRNA-sequencing data. Occurrence and sequence context of G/C mutation on chromosome 7 at position 2297157 point to a sequencing artefact. The A/C mutation at chromosome 11 at position 1901340 was also found in a small frequency of reads in patient sample 4a, 5a, 6a, 8a and 11b. The G/C mutation at position 1901315 was just found in patient sample 3a in the miRNA sequencing files and points to a heterozygous mutation.

The latter mutation seems to be a real mutation with a high probability, because exome and miRNA sequencing data confirm this mutation. A validation by Sanger sequencing could not be performed because there was no patient material left for further analyses. Nevertheless this mutation should be considered for the analyses of further patient material.

### 4.1.4 Treatment of *TCF3/HLF*-positive cell line with HDAC6

#### Inhibitor KSK

The outcome of *TCF3/HLF*-positive leukemia is poor. Until now, no successful medical therapy strategies are available for treatment of this rare leukemia subtype. Therefore the development of potential drugs is an important research topic. Apoptosis experiments with the novel HDAC inhibitor KSK showed an increase of apoptotic cells compared to healthy B-cells, after treatment with KSK. To analyze the effect of the compound KSK on the expression pattern in t(17;19)-positive cells, HAL-01-cells were treated for 24 h with KSK, Ricolinostat and DMSO. RNA was harvested and cDNA-libraries out of mRNA were cloned for NGS. Seventeen genes were differentially expressed between KSK- and DMSO-treated cells with a generalized fold change (gfold) = 2 (Table 8). Out of them, seven genes were also

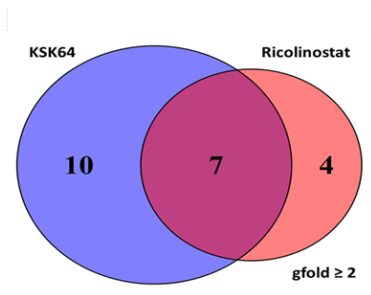


differentially expressed between Ricolinostat- and DMSO-treated cells (Figure 24). Some of these genes like *RASSF4*, *MACC1* or *PRAME* have an influence on tumor development and progression and should be validated in the future.

**Table 8: Differentially expressed genes after KSK treatment.**

GeneSymbol	GeneName	GFOLD(0.01)	E-FDR	log2fdc	1stRPKM	2ndRPKM
ENSG00000065717	<i>TLE2</i>	-2.24	0.16	-3.61	1.48	0.10
ENSG00000072694	<i>FCGR2B</i>	-3.04	0.16	-4.46	1.11	0.04
ENSG00000100968	<i>NFATC4</i>	-3.51	0.16	-4.88	1.21	0.03
ENSG00000105048	<i>TNNT1</i>	-3.22	0.16	-4.77	1.91	0.03
ENSG00000105419	<i>MEIS3</i>	-2.11	0.16	-3.87	0.28	0.00
ENSG00000107551	<i>RASSF4</i>	-2.18	0.16	-3.32	2.14	0.21
ENSG00000154310	<i>TNIK</i>	-2.20	0.16	-3.74	0.19	0.00
ENSG00000155367	<i>PPMIJ</i>	-2.36	0.16	-3.94	0.78	0.01
ENSG00000163520	<i>FBLN2</i>	-2.63	0.16	-3.97	0.79	0.03
ENSG00000166292	<i>TMEM100</i>	-2.03	0.16	-3.82	0.80	0.03
ENSG00000167371	<i>PRRT2</i>	-2.37	0.16	-3.89	0.97	0.04
ENSG00000167680	<i>SEMA6B</i>	-2.36	0.16	-3.92	0.51	0.02
ENSG00000183742	<i>MACC1</i>	-2.30	0.16	-3.63	1.13	0.09
ENSG00000185686	<i>PRAME</i>	-2.06	0.16	-3.96	0.36	0.00
ENSG00000198829	<i>SUCNR1</i>	-2.41	0.16	-4.12	0.71	0.01
ENSG00000229191	<i>RP11-168O16.1</i>	-2.25	0.16	-3.47	2.09	0.18
ENSG00000235142	<i>RP1-60O19.1</i>	-2.33	0.16	-3.47	1.09	0.08

The table shows differentially expressed genes between KSK- and DMSO-treated cells with a gfold  $\leq -2$ . Gfold: generalized Fold Change; E-FDR: Enhanced False Discovery Rate; log<sub>2</sub>fdc: log<sub>2</sub> fold change; RPKM: Reads per Kilobase Million.



**Figure 24: Venn diagram of differentially expressed genes after HDAC6 inhibitor treatment.** The Venn diagram shows the number of differentially expressed genes after HDAC6 inhibitor treatment (KSK or Ricolinostat) compared to DMSO-treated cells after 24 h of exposing with a gfold  $\geq 2$ .

## 4.2 Differences between Infant and Childhood T-ALL

### 4.2.1 Patient Cohort

The development of T-ALL in infancy is very rare and hard to cure. Most of these young patients die. Treatment of T-ALL in the later childhood is accompanied with a much better prognosis. To find the reason for this, samples of infants with T-ALL and samples of children with T-ALL were compared to each other on a molecular level.

Blast material from infant T-ALL (iT-ALL) patients (until 12 months of age) and blast material from childhood T-ALL patients (1 to 16 years of age) were analyzed individually and then compared to each other (Table 9 and Table 10). The median age of the infant patients was nine months. The median age of the childhood patients was eleven years.

**Table 9: Patient information.**

Patient ID	Age at Diagnosis	Sex	Material	Outcome
101	5 months	m	PB	dec
102	9 months	f	BM	dec
103	10 months	f	PB	dec
201	15 years	m	PB	rem
202	14 years	m	BM	rem
203	5 years	m	PB	rem
204	16 months	f	BM	rem
205	12 years	m	PB	rem
206	10 years	f	PB	rem

m: male; f: female; dec: deceased; rem: remission; PB: peripheral blood, BM: bone marrow.

**Table 10: Information on patient genetics.**

Patient ID	<i>BCR/ABL1</i>	<i>KMT2A/AFF1</i>	<i>ETV6/RUNX1</i>	<i>KMT2A/MLLT1</i>	Cytogenetics	Classification

101	neg	neg	neg	n.d.	neg	mature T
102	neg	neg	neg	n.d.	n.d.	pre-T
103	neg	neg	neg	neg	complex karyotype	cortical T
201	neg	neg	neg	neg	neg	pre-T
202	neg	neg	neg	n.d.	neg	cortical T
203	neg	neg	neg	neg	neg	cortical T
204	neg	neg	neg	neg	46,XX,der(11)(?q14)(13)/ 46,XX(10)	cortical T
205	neg	neg	neg	neg	neg	cortical T
206	n.d.	n.d.	n.d.	n.d.	46,XX,t(11;14)(p13;q11) <i>TRD/LMO2</i> , CDKN2A deletion	cortical T

neg: negative/unremarkable; n.d.: not determined.

#### 4.2.2 Mutations in iT-ALL Samples

To analyze the mutational spectrum in iT-ALL, exomes of all three infant patients were sequenced with NGS. In total, 4504 mutations in 1595 genes were detected in any of the three infant samples. One thousand three hundred and five recurrent mutations in 798 genes were detected in at least two of the infant samples. Five hundred fifty-seven recurrent mutations in 426 genes were mutated in all three infant samples. Because there was no germline material available, these mutations include leukemic-specific mutations as well as germline mutations. We analyzed 19 of these mutations in genes, which were recurrently mutated in T-ALL and validated them by Sanger sequencing (Table 11).

**Table 11: Mutation in iTALL patients validated by Sanger sequencing.**

Coordinates	Gene	Type of alteration	Patient	Genotype
1:120478125-120478125	<i>NOTCH2</i>	SNV	102	heterozygous
19:15285382-15285386	<i>NOTCH3</i>	4 bp deletion	103	homozygous
19:15285382-15285386	<i>NOTCH3</i>	4 bp deletion	101	heterozygous

19:15285382-15285386	<i>NOTCH3</i>	4 bp deletion	102	heterozygous
5:35857308-35857309	<i>IL7R</i>	1 bp insertion	103	homozygous
5:35857308-35857309	<i>IL7R</i>	1 bp insertion	101	heterozygous
10:89653620-89653621	<i>PTEN</i>	1 bp deletion	101	heterozygous
10:89690952-89690957	<i>PTEN</i>	5 bp insertion	102	homozygous
10:89690952-89690957	<i>PTEN</i>	5 bp insertion	101	heterozygous
10:89717674-89717676	<i>PTEN</i>	2 bp insertion	103	heterozygous
10:89725886-89725887	<i>PTEN</i>	1 bp insertion	101	heterozygous
10:89728633-89728634	<i>PTEN</i>	1 bp deletion	102	heterozygous
10:89731315-89731317	<i>PTEN</i>	2 bp insertion	101	heterozygous
12:25358662-25358664	<i>KRAS</i>	2 bp deletion	101	heterozygous
12:25358662-25358664	<i>KRAS</i>	2 bp deletion	103	homozygous
4:154626317-154626317	<i>TLR2</i>	SNV	102	heterozygous
9:120466929-120466930	<i>TLR4</i>	1 bp deletion	103	heterozygous
9:120466929-120466930	<i>TLR4</i>	1 bp deletion	101	heterozygous
9:120466929-120466930	<i>TLR4</i>	1 bp deletion	102	heterozygous

Coordinates describing the base position in GRCh38. SNV: single nucleotide variant; bp: base pair.

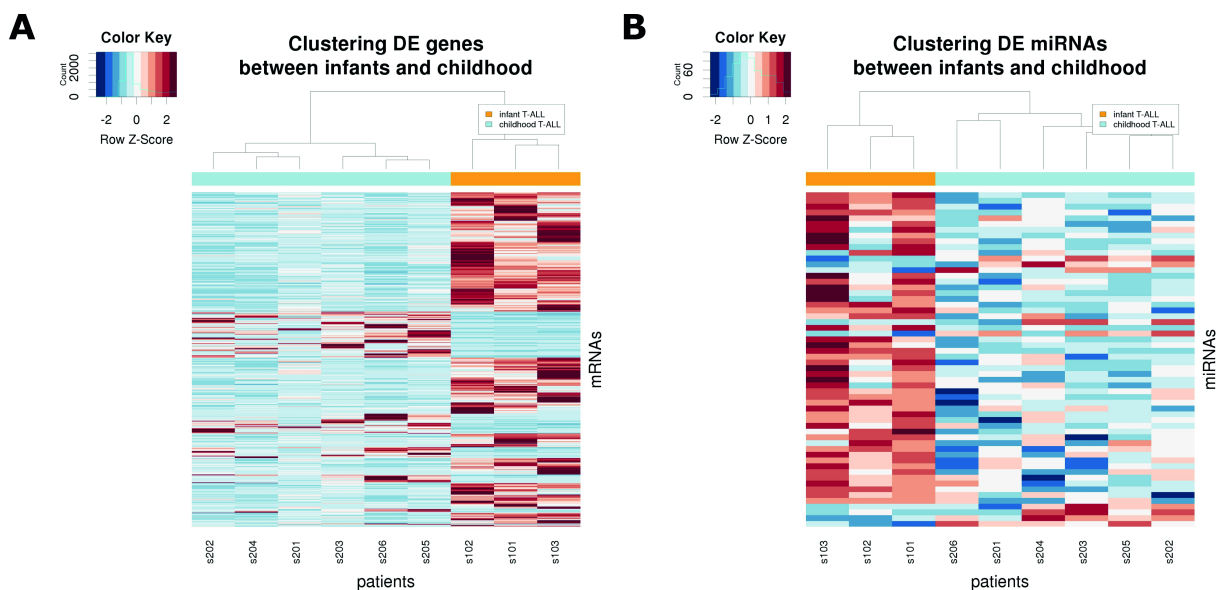
We found a heterozygous SNV in *NOTCH2* in one iTALL patient (#102), which was predicted to be deleterious. This non-synonymous mutation leads to an amino acid exchange from phenylalanine to valine. The mutation occurs in the extracellular EGF-like domain in the NOTCH2 protein, which is needed for Ca<sup>2+</sup> depended ligand binding<sup>114</sup>. In the same patient a homozygous 10 bp deletion in the *NOTCH3* gene was found. Also, a 4 bp deletion in the *NOTCH3* gene was found in all three infant patients (heterozygous in 102 and 103, homozygous in 101), in a repetitive area.

There were also a series of mutations in the tumor suppressor gene *PTEN*. In patient 103, a heterozygous 2 bp insertion was found. Patient 102 had one heterozygous *PTEN* mutation – a one bp deletion. Patient 101 harbored four heterozygous *PTEN* mutations – a one bp deletion and three insertions of one, two and five bases. The five bp insertion also occurred homozygous in patient 102.

Another gene, which was mutated in all three infant patients, was *KRAS*. Patient 103 had two homozygous 1 bp insertions, which also occurred homozygous in patient 101. Patients 102 carried one of these mutations homozygous, the other heterozygous. Patients 103 (homozygous) and patient 101 (heterozygous) carried also a two bp deletion in *KRAS*.

### 4.2.3 Differences in Gene Expression

To analyze the differences between childhood and iT-ALL on a transcriptomic level, we performed RNA sequencing (RNA seq) with tumor samples of three infant and six childhood T-ALL patients. We obtained a total of 135,216,180 reads mapping to coding genes. We found 760 genes differentially expressed between childhood and iT-ALL ( $|\logFC| \geq 1$ ,  $p$ -value  $\leq 0.01$ ). Two hundred and seven genes were downregulated and 553 genes were upregulated in the infant samples. Hierarchical clustering of significantly differential expressed genes clearly separated infant from childhood samples indicating a genetic difference between infant and childhood T-ALL (Figure 25A).



**Figure 25: Hierarchical Clustering of infant and childhood T-ALL.** Infant and childhood samples were separated by significantly differentially expressed A) mRNAs and B) miRNAs.

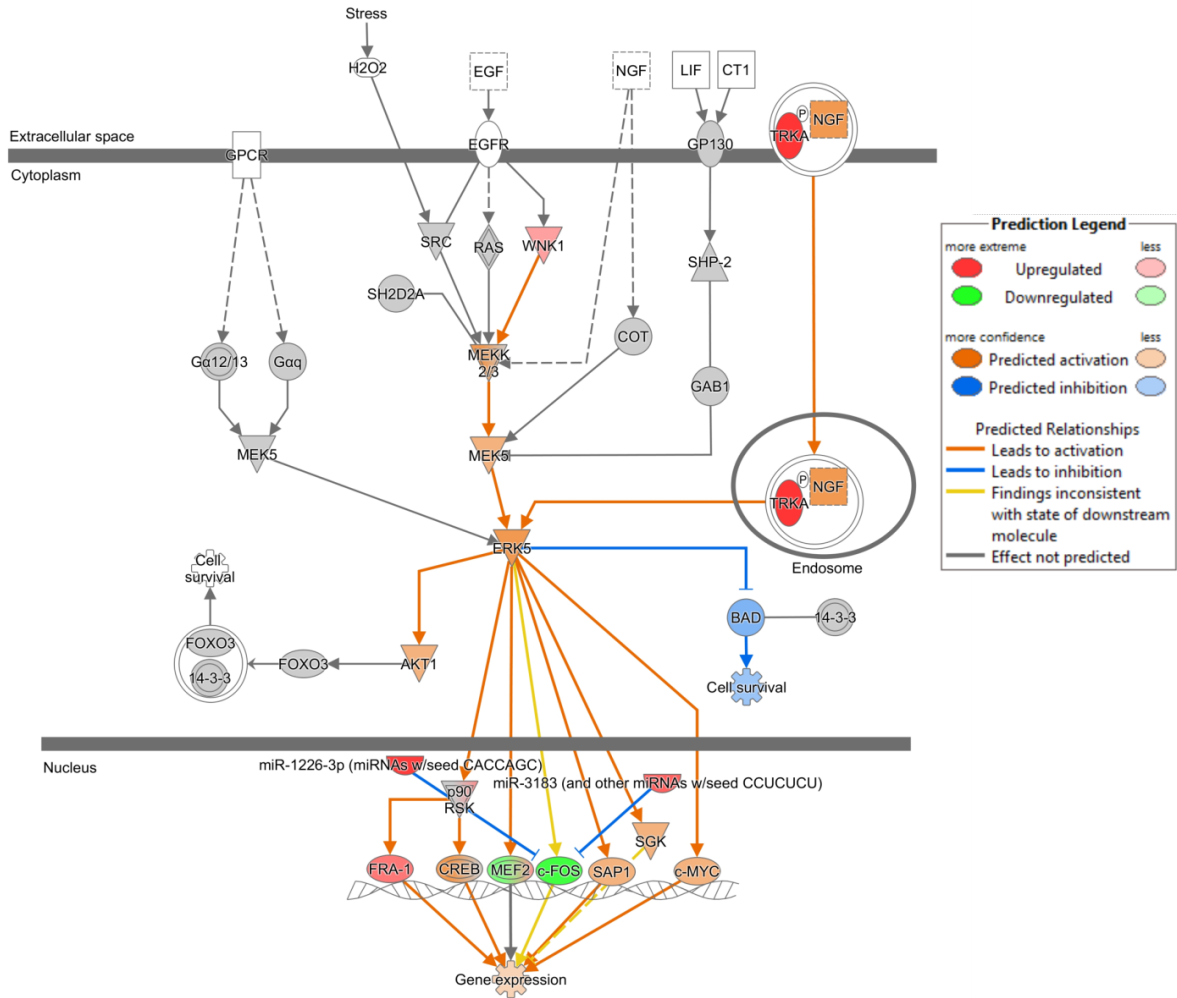
To gain deeper insights into the pathways, in which the differentially expressed genes are involved, a pathway analysis with Ingenuity Pathway analysis from Qiagen was performed. We found nine pathways, which were differentially regulated between infant and childhood cases ( $p$ -value  $< 0.05$ ,  $z$ -score  $\geq \pm 1$ ) (Table 12). The given  $z$ -score represents a statistical calculated value, which describes the match between expected and observed state of the pathway (activation or inhibition). If there is a positive correlation between the expression of a transcription factor and the expression of its downstream target, this interaction gets the value 1. If there is a negative correlation between the expression of a transcription factor and the expression of its downstream target, this interaction get the value -1. The sum of all

interactions is taken and divided by the root of the amount of interactions:

immune system and some were  
differentiation and cell survival. One  
protein-kinase 7 (ERK5)-pathway,  
1.134) (Figure 26). The ERK5  
including growth factors, cellular  
in childhood T-ALL samples an activation  
) leads to activation of ERK5.  
known as proto-oncogenes, which  
include FOS like antigen 1 (  
) and cellular myelocytomatosis

**Childhood T-ALL.** Pathways are filtered by  
z-score < -1) in iT-ALL.

<b>p-value</b>	<b>z-score</b>
0.0005	-1.000
0.0006	-1.387
0.0014	1.134
0.0021	-1.342
0.0026	-1.134
0.0062	1.000
0.0331	1.000
0.0380	-1.134
0.0417	1.134



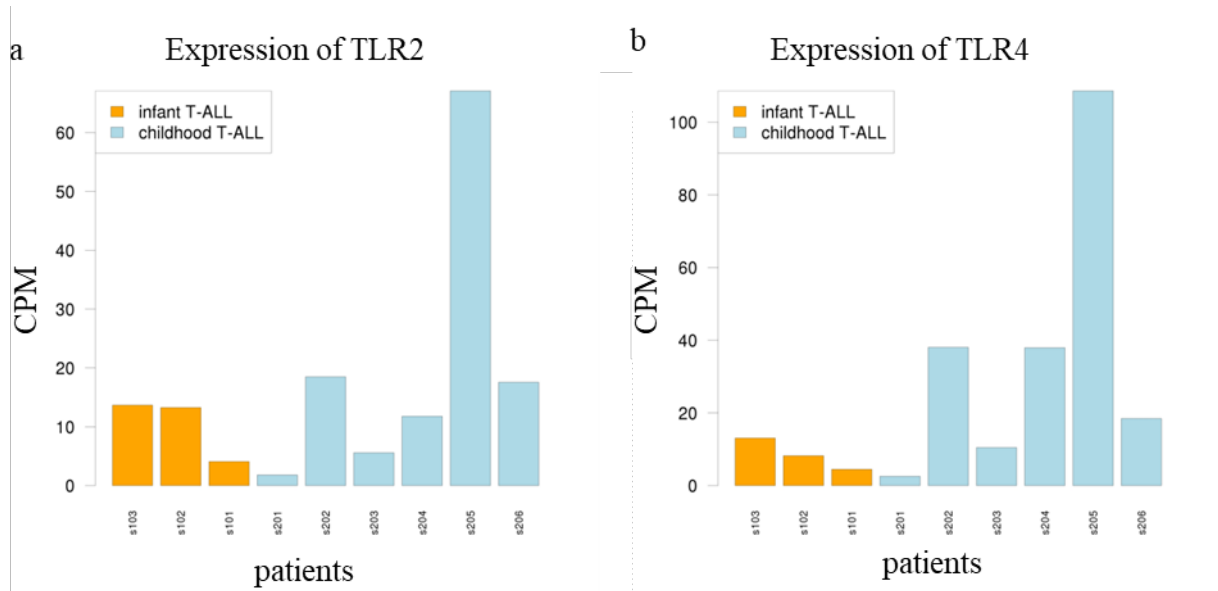
© 2000-2016 QIAGEN. All rights reserved.

**Figure 26: Activation of ERK5 signaling cascade in iT-ALL.** The signaling map shows differences between childhood and infant T-ALL in ERK5 signaling including differentially expressed genes and miRNAs as well as predictions for activation and inhibition. Predictions are based on NGS data. Small horizontal ovals: chemicals, dashed squares: growth factors, squares: cytokines, rectangles: G-protein-coupled-receptors, vertical ovals: other transmembrane receptors, triangles: kinases, diamonds: G-protein, big horizontal ovals: transcription factors, circles: other proteins; double edged proteins: protein complexes, semicircles: miRNAs.

#### 4.2.4 Upstream Regulator Analysis

By an upstream regulator analysis toll like receptor 2 (TLR2) and TLR4 seem to be inhibited in iT-ALL based on the expression of their downstream targets. Transcriptomic NGS data showed no significant downregulation of *TLR* in iT-ALL compared to childhood T-ALL (logFC of -0.98; p-value = 0.29) (Figure 27A). *TLR4* was significantly downregulated in iT-ALL (logFC of -2.07 p-value = 0.03) (Figure 27B). Exomic NGS data showed mutations in TLRs in all three infant patients. Patient 102 harbored two heterozygous *TLR2* mutations. All three infant patients harbored a heterozygous 1 bp deletion in *TLR4* (Table 11). Downstream

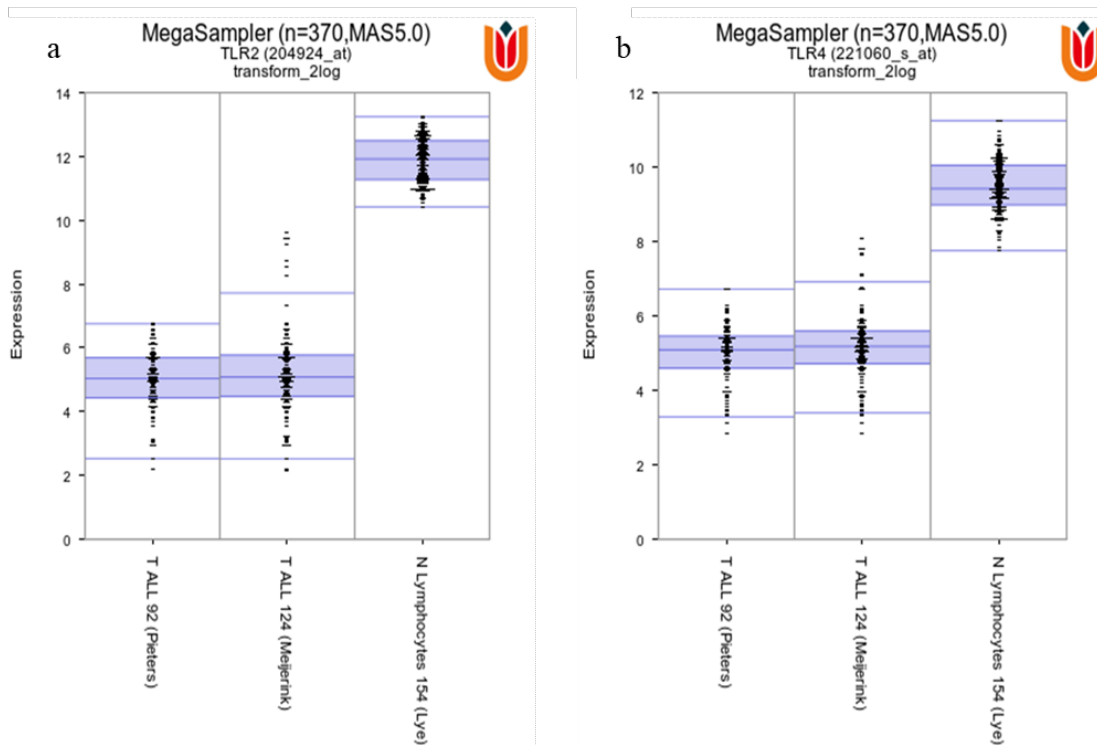
targets of *TLR2* are selectin P (*SELP*), integrin subunit alpha 2b (*ITGA2B*), interleukin 1beta (*IL1B*), cluster of differentiation 86 (*CD86*) and interleukin 6 (*IL6*), which were all significantly downregulated in iT-ALL samples. Downstream targets of *TLR4* are *CD86*, *IL1B*, *IL6*, C-C motif chemokine receptor 7 (*CCR7*) and C-C motif chemokine ligand 5 (*CCL5*).



**Figure 27: Expression of TLRs in T-ALL samples.** a) Expression values in CPM of *TLR2* and b) *TLR4* in infant (orange) and childhood (light blue) T-ALL samples.

By analysis of expression values between T-ALL blasts and normal lymphocytes (including T-cells, B-cells and natural killer cells) with R2 MegaSampler, a downregulation of both TLRs can be observed in T-ALL blasts. *TLR2* is downregulated more than 100-fold compared to normal lymphocytes (Figure 28A). *TLR4* is downregulated more than 15-fold compared to normal lymphocytes (Figure 28B).





**Figure 28: Expression of TLRs in T-cell ALL and normal lymphocytes.** a) Expression of *TLR2* and b) Expression of *TLR4* in T-ALL blasts and normal lymphocytes. *TLR2/4* expression is downregulated in T-ALL blasts compared to healthy lymphocytes.

#### 4.2.5 Druggable Targets

Furthermore, transcriptomic NGS data were used to search for overexpressed genes, which were druggable by approved drugs or targeted by drugs which are already in clinical trials. We found six overexpressed genes, which were druggable and so potentially of interest for optimizing treatment strategies (Table 13). *KIT* (logFC = 2.03) is one of the upregulated genes, a receptor tyrosine kinase, which is frequently associated with cancer. Several already approved tyrosine kinase inhibitors (TKI) are available to inhibit *KIT* such as Imatinib, Dasatinib, Sorafenib or Pazopanib.

**Table 13: List of clinically approved drugs or drugs in clinical trials and their significantly upregulated target genes in iTALL.**

Gene	logFC infants	Drugs
------	---------------	-------

<i>BRD3</i>	1.44	OTX015, CPI-0610, I-BET-762 (GSK525762), TEN-010
<i>KIT</i>	2.03	Imatinib, Dasatinib, Sorafenib, Axitinib, Pazopanib, Cabozantinib, Sunitinib, Ponatinib, Regorafenib, Nilotinib, Cediranib (AZD2171), Dovitinib (TKI-258, CHIR-258), Motesanib (AMG 706), Tandutinib (MLN518), Masitinib (AB1010), Telatinib, Tivozanib, OSI-930, Amuvatinib (MP-470), Midostaurin (PKC412), Quizartinib (AC220), MGCD516, Famitinib
<i>BLK</i>	2.48	Dasatinib, Saracatinib (AZD0530)
<i>FLT1</i>	2.62	Axitinib, Cabozantinib, Nintedanib (BIBF 1120), Regorafenib, Pazopanib, Motesanib (AMG706), MGCD-265, Cediranib (AZD2171), Foretinib (GSK1363089), Lenvatinib (E7080), Tivozanib (AV-951), Lucitanib (E-3810), Famitinib, Linifanib (ABT-869), OSI-930, Cabozantinib
<i>NTRK1</i>	4.07	Ponatinib, Cabozantinib, Entrectinib (RXDX-101), TSR-011, PLX7486, LOXO-101, Dovitinib (TKI-258, CHIR-258), MGCD516, Lestaurtinib (CEP-701), DS-6051b, BMS-754807, Milciclib (PHA-848125), Danusertib (PHA-739358), Lestaurtinib (CEP-701), Cabozantinib
<i>ERBB4</i>	5.96	Lapatinib, Dacomitinib (PF299804, PF299), AC480 (BMS-599626)

#### 4.2.6 Target Prediction of DE miRNAs

To analyze differences in the epigenome between infant and childhood T-ALL, miRNA-seq was performed. A total of 31,761,705 sequencing reads was obtained mapping to miRNAs in miRBase V21. Fifty-eight miRNAs were differentially expressed between infant and childhood T-ALL ( $|\log_{2}FC| \geq 1$ ,  $p\text{-value} \leq 0.01$ ) (Table 14). Nine miRNAs were downregulated and 49 were upregulated in the infant samples. Infant and childhood samples were clearly separated by using these differentially expressed miRNAs in hierarchical clustering (Figure 25B) indicating a difference between both groups not only on the transcriptomic, but also on the epigenomic level.

To find genes whose aberrant expression pattern could be explained by differentially expressed miRNAs, a target prediction and correlation analysis based on five public miRNA target databases was performed. The correlation analysis revealed 47 potential mRNA-miRNA target pairs (Spearman's  $Rho \leq -0.6$ ,  $p\text{-value} \leq 0.05$ ) ([\\* MERGEFORMAT](#)

**Table 15).**

Hsa-miR-let7-5p was downregulated in infant T-ALL samples. A downregulation of this miRNA was already described in infant B-ALL with *KMT2A* rearrangements (*KMT2Ar*) by hypermethylation of the promoter region ADDIN EN.CITE <sup>115,116</sup>. *KMT2Ar* in patients 101, 102 and 103 was not detected, but the downregulation of hsa-let7b-5p could explain the high expression of their potential target genes (Dipeptidyl Peptidase Like 6 (*DPP6*), growth regulation by estrogen in breast cancer 1 (*GREB1*) (Figure 29B), Collagen Type I Alpha 2 Chain (*COL1A2*), Fraser Extracellular Matrix Complex Subunit 1 (*FRASI*), Collagen Type IV Alpha 6 Chain (*COL4A6*) and Insulin Like Growth Factor 2 mRNA Binding Protein 1 (*IGF2BP1*) (Figure 29A)).

Tumor suppressor miR hsa-miR-31 was also downregulated in iT-ALL (logFC = -3.86, P-value = 0.004), which could explain the upregulation of its potential target replication protein A1 (RPA1) (Spearman's Rho: -0.62, P-value = 0.04) (Figure 29C).

**Table 14: Differentially expressed miRNAs between infant and childhood T-ALL samples.**

Symbol	logFC	p-value
hsa-let-7b-5p	-2.92	0.0028
hsa-miR-18a-5p	2.38	0.0041
hsa-miR-31-5p	-3.86	0.0036
hsa-miR-30c-5p	2.50	0.0079
hsa-miR-183-5p	2.93	0.0004
hsa-miR-205-5p	-3.71	0.0040
hsa-miR-210-3p	1.62	0.0099
hsa-miR-185-5p	2.15	0.0008
hsa-miR-190a-5p	2.02	0.0019
hsa-miR-200c-3p	1.90	0.0064
hsa-miR-148b-3p	2.49	0.0001
hsa-miR-331-3p	2.65	0.0000
hsa-miR-324-5p	1.96	0.0015
hsa-miR-196b-5p	-5.23	0.0016
hsa-miR-502-5p	2.26	0.0059
hsa-miR-652-3p	2.12	0.0014
hsa-miR-421	1.92	0.0037
hsa-miR-671-5p	1.78	0.0079
hsa-miR-766-3p	3.29	0.0001
hsa-let-7b-3p	-3.61	0.0004
hsa-let-7f-1-3p	-2.52	0.0079
hsa-miR-223-5p	2.35	0.0026

hsa-miR-125b-2-3p	3.53	0.0004
hsa-miR-29c-5p	1.95	0.0092
hsa-miR-874-3p	2.25	0.0015
hsa-miR-1226-3p	3.13	0.0001
hsa-miR-1301-3p	1.83	0.0093
hsa-miR-1180-3p	1.87	0.0033
hsa-miR-1249-3p	1.90	0.0050
hsa-miR-1276	2.47	0.0078
hsa-miR-196b-3p	-5.66	0.0033
hsa-miR-3143	2.04	0.0096
hsa-miR-3186-3p	3.57	0.0001
hsa-miR-3620-3p	2.77	0.0033
hsa-miR-3661	2.86	0.0014
hsa-miR-3681-5p	3.36	0.0001
hsa-miR-3909	2.16	0.0014
hsa-miR-3922-3p	2.27	0.0024
hsa-miR-4485-3p	-5.60	0.0008
hsa-miR-4687-5p	2.43	0.0023
hsa-miR-5010-3p	2.34	0.0092
hsa-miR-664b-3p	4.32	0.0000
hsa-miR-5581-3p	2.36	0.0033
hsa-miR-5683	4.02	0.0000
hsa-miR-561-5p	2.90	0.0079
hsa-miR-652-5p	2.18	0.0030
hsa-miR-1306-5p	2.16	0.0038
hsa-miR-6503-5p	-5.17	0.0033
hsa-miR-210-5p	2.61	0.0015
hsa-miR-128-1-5p	2.36	0.0017
hsa-miR-6802-3p	2.51	0.0061
hsa-miR-6803-3p	2.68	0.0006
hsa-miR-6806-3p	2.81	0.0037
hsa-miR-6855-3p	2.26	0.0071
hsa-miR-6769b-3p	2.55	0.0019
hsa-miR-6894-3p	2.57	0.0054
hsa-miR-7706	2.72	0.0001
hsa-miR-128-2-5p	4.42	0.0001

LogFC: log fold change.

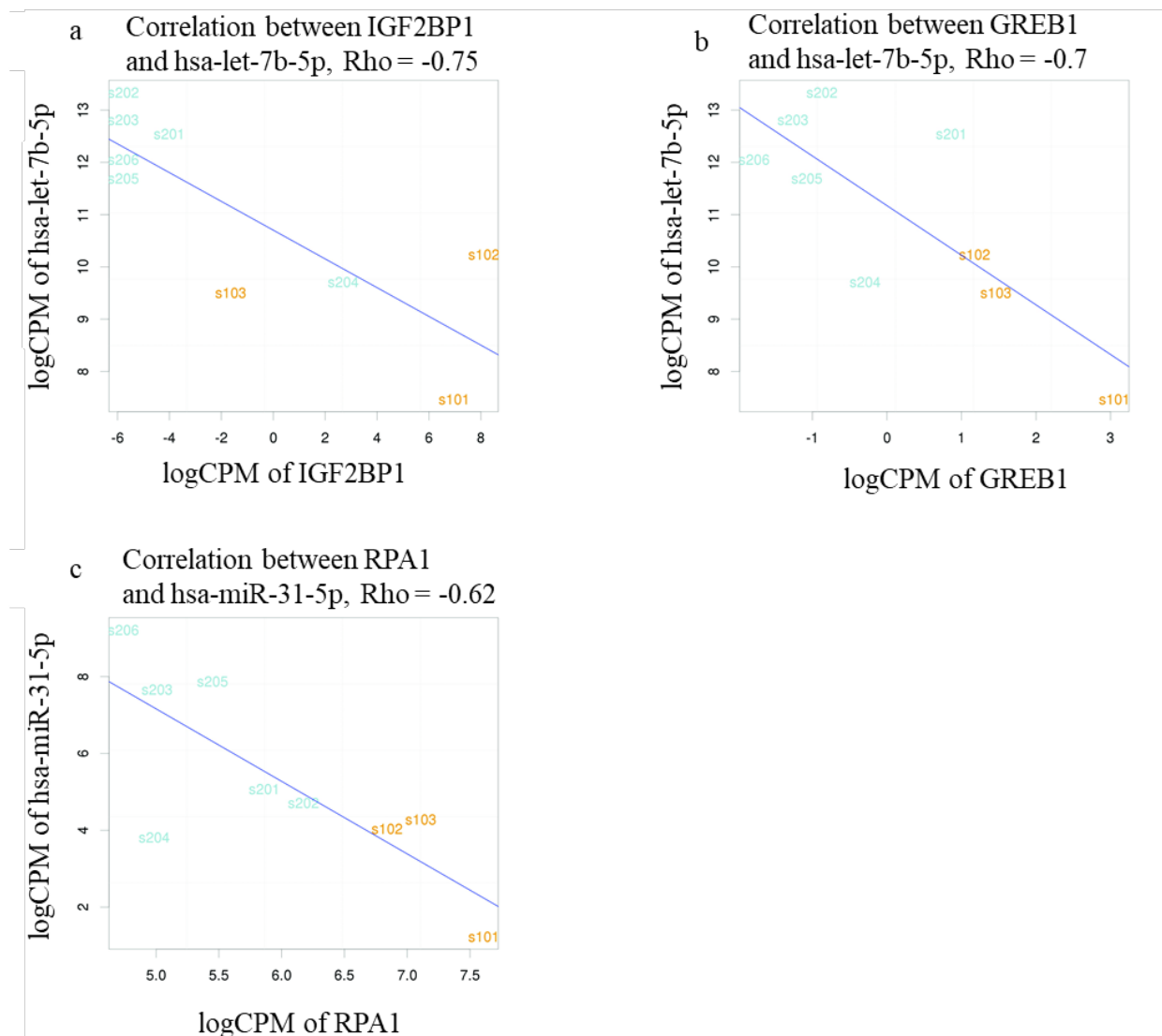
**Table 15: Potential mRNA-miRNA target pairs.**

miRNA	miRNA logFC	targeted mRNA	mRNA logFC	Spearman Rho	Correlation p-value
-------	-------------	---------------	------------	--------------	---------------------

<b>hsa-miR-5683</b>	4.0	<b>HLF</b>	-3.0	-0.73	0.02
<b>hsa-miR-205-5p</b>	-3.7	<b>NFAT5</b>	1.4	-0.75	0.01
<b>hsa-miR-421</b>	1.9	<b>HOXA9</b>	-5.0	-0.6	0.05
<b>hsa-miR-3909</b>	2.2	<b>NOG</b>	-2.6	-0.62	0.04
<b>hsa-let-7b-5p</b>	-2.9	<b>SCAF4</b>	1.4	-0.68	0.03
<b>hsa-let-7b-5p</b>	-2.9	<b>GREB1</b>	2.6	-0.7	0.02
<b>hsa-miR-766-3p</b>	3.3	<b>NR3C2</b>	-4.5	-0.78	0.01
<b>hsa-miR-5683</b>	4.0	<b>CDK14</b>	-2.3	-0.93	0.00
<b>hsa-miR-183-5p</b>	2.9	<b>CRI</b>	-2.9	-0.63	0.04
<b>hsa-miR-3909</b>	2.2	<b>CD300E</b>	-2.9	-0.6	0.05
<b>hsa-miR-3143</b>	2.0	<b>ABCD2</b>	-4.0	-0.72	0.02
<b>hsa-miR-18a-5p</b>	2.4	<b>MEF2C</b>	-2.0	-0.63	0.04
<b>hsa-let-7f-1-3p</b>	-2.5	<b>ZFAND3</b>	1.8	-0.78	0.01
<b>hsa-let-7b-5p</b>	-2.9	<b>PDPR</b>	1.4	-0.6	0.05
<b>hsa-miR-148b-3p</b>	2.5	<b>NR3C2</b>	-4.5	-0.67	0.03
<b>hsa-let-7b-5p</b>	-2.9	<b>COL1A2</b>	2.9	-0.64	0.03
<b>hsa-miR-421</b>	1.9	<b>MEF2C</b>	-2.0	-0.67	0.03
<b>hsa-let-7f-1-3p</b>	-2.5	<b>PRPF8</b>	1.5	-0.6	0.05
<b>hsa-let-7b-3p</b>	-3.6	<b>WNK1</b>	1.5	-0.72	0.02
<b>hsa-miR-5683</b>	4.0	<b>ZBTB38</b>	-2.9	-0.62	0.04
<b>hsa-miR-183-5p</b>	2.9	<b>KDELC2</b>	-2.3	-0.83	0.00
<b>hsa-let-7b-5p</b>	-2.9	<b>PRDM2</b>	1.4	-0.7	0.02
<b>hsa-miR-5683</b>	4.0	<b>MEF2C</b>	-2.0	-0.87	0.00
<b>hsa-let-7b-5p</b>	-2.9	<b>FRAS1</b>	2.4	-0.82	0.01
<b>hsa-miR-1276</b>	2.5	<b>ADCY9</b>	-2.6	-0.69	0.02
<b>hsa-miR-3143</b>	2.0	<b>NOG</b>	-2.6	-0.67	0.03
<b>hsa-miR-5581-3p</b>	2.4	<b>MEIS1</b>	-3.5	-0.8	0.01
<b>hsa-let-7f-1-3p</b>	-2.5	<b>NBEA</b>	1.5	-0.73	0.02
<b>hsa-let-7b-5p</b>	-2.9	<b>IGF2BP1</b>	6.9	-0.75	0.01
<b>hsa-miR-18a-5p</b>	2.4	<b>ERRFI1</b>	-3.3	-0.75	0.01
<b>hsa-miR-671-5p</b>	1.8	<b>ALDH3A2</b>	-2.1	-0.62	0.04
<b>hsa-miR-421</b>	1.9	<b>PTGER2</b>	-1.9	-0.72	0.02
<b>hsa-let-7b-5p</b>	-2.9	<b>WNK1</b>	1.5	-0.73	0.02
<b>hsa-miR-331-3p</b>	2.7	<b>MEIS1</b>	-3.5	-0.87	0.00
<b>hsa-miR-148b-3p</b>	2.5	<b>PTGER2</b>	-1.9	-0.67	0.03
<b>hsa-miR-205-5p</b>	-3.7	<b>RUNX2</b>	3.6	-0.6	0.05
<b>hsa-miR-148b-3p</b>	2.5	<b>MYBL1</b>	-3.4	-0.68	0.03
<b>hsa-let-7b-5p</b>	-2.9	<b>COL4A6</b>	4.6	-0.83	0.00
<b>hsa-miR-5581-3p</b>	2.4	<b>FOS</b>	-3.0	-0.85	0.00
<b>hsa-let-7b-3p</b>	-3.6	<b>JUP</b>	2.9	-0.7	0.02
<b>hsa-miR-31-5p</b>	-3.9	<b>RPA1</b>	1.8	-0.62	0.04
<b>hsa-miR-31-5p</b>	-3.9	<b>WNK1</b>	1.5	-0.72	0.02
<b>hsa-let-7b-3p</b>	-3.6	<b>SLC18A2</b>	2.7	-0.63	0.04
<b>hsa-miR-5581-3p</b>	2.4	<b>MYBL1</b>	-3.4	-0.87	0.00

<b>hsa-let-7b-3p</b>	-3.6	<b>ZFAND3</b>	1.,8	-0.75	0.01
<b>hsa-miR-148b-3p</b>	2.5	<b>RAB34</b>	-3.8	-0.63	0.04
<b>hsa-miR-421</b>	1.9	<b>NR3C2</b>	-4.5	-0.73	0.02

Target pairs with a Spearman's Rho of  $\leq -0.6$  and a p-value of  $< 0.05$ . Target prediction was performed using five public available data bases (miRanda, miRDB, TarBase, TargetScan, miRTarBase).



**Figure 29: Negative correlation between *hsa-let-7b* and *IGF2BP1* (A)/*GREB1* (B) as well as *hsa-miR-31* and *RPA1* (C).** Light blue: childhood T-All samples; orange: infant T-ALL samples. s101: sample 101; s102: sample 102 etc. Childhood T-ALL samples show a high expression of A/B: *hsa-let7b* and C: *hsa-miR-31* and a small expression of the corresponding potential target genes (A: *IGF2BP1*, B: *GREB1*, C: *RPA1*). In contrast to this, infant T-ALL samples show little expression of A/B: *hsa-miR-let-7b* and of C: *hsa-miR-31* and a high expression of the corresponding potential target genes.

## 5. Discussion

### 5.1 Differences between mRNA and miRNA target pairs t(1;19) and t(17;19) positive ALL

The translocation t(17;19) is a chromosomal aberration mostly found in pro-B-ALL. It appears very rarely and is accompanied with a poor prognosis in pediatric B-ALL patients. As the more common and well treatable t(1;19)-B-ALL subtype, t(17;19) leads to a protein fusion construct with TCF3 on the molecular level. While t(1;19) leads to a fusion of *TCF3* and *PBX1*, a *TCF3/HLF* fusion construct arises in t(17;19). To find molecular effects of t(17;19), paired leukemic and remission samples of pediatric patients with either t(1,19) or t(17;19) were sequenced and compared to each other. The focus lay on differences in miRNA expression and function. Therefore, first differentially expressed miRNAs and mRNAs between both tumor subtypes and healthy samples were analyzed. Indeed, mRNA and miRNA expression profiles clearly separated the two leukemia entities from one another.

Correlation analyses and target predictions were performed with different methods to identify miRNA-mRNA regulation pairs. Method 1 revealed 14 potential mRNA-miRNA target pairs, of which 2 (*SOX4* and hsa-miR-199a/b) were tested. Method 2 revealed 4 potential mRNA-miRNA target pairs, of which 1 (*MSI2* and has-miR-892a) was tested. Method 3 revealed 31 different potential mRNA-miRNA target pairs, of which 2 (*FAT1* and hsa-miR-223, *MDFIC* and hsa-miR-130) were tested. Method 4 revealed 43 potential mRNA-miRNA target pairs, of which 4 (*PRKCE* and hsa-miR-223, *E2F2* and hsa-let-7g, *TAOK* and hsa-miR-182, *ITPR1* and hsa-miR-130) were tested. The interaction of all tested pairs could not be validated on a mRNA level. Thus we gain a rate of false positives of 14%, 25%, 6% and 9%, respectively. These numbers just consider the tested potential pairs and could therefore be higher.

Because all predicted mRNA-miRNA target pairs could not be validated *in vitro*, the question arises, how useful the prediction tools really are. A target prediction between mRNA and miRNA represents a hard task. Around 20,000 genes are predicted to be regulated by miRNAs in humans by using alignments to conserved miRNA binding sites. For this reason, it is believed that miRNAs have a big impact on functional biological processes and could also have an impact on development of different diseases. But the rate of false positive predictions is high (up to around 50%)<sup>117</sup>.

Specific motifs of 3'UTRs flanked by adenosines are conserved in vertebrates<sup>117</sup>. These sites bind to the miRNA seed region considered to be the first eight nucleotides of the miRNA<sup>118</sup>. Most prediction tools use the complementarity between mRNA-3'UTR and the miRNA seed region by sequence alignments. Different to miRNA targeting in plants, where there is a high complementarity between miRNA and target gene sequence ADDIN EN.CITE <sup>119</sup>, the complementarity between miRNA and target gene in animals is just partially given. The small sequence of a miRNA makes such a prediction even harder. The exact rules and energetics of the pairing between a miRNA, its target gene and also the involved protein complex is not fully understood, so that it cannot easily be predicted by a computational program. This is the reason why most prediction tools deliver a large set of potential mRNA-miRNA target pairs including false positive pairs ADDIN EN.CITE <sup>120</sup>.

MiRanda ADDIN EN.CITE <sup>121</sup> is one prediction tool we used in the infant T-cell leukemia project. It considers the complementarity between miRNA and 3'UTR-sequence of a potential target by aligning the whole miRNA against mRNA 3'UTR and calculates the free energy of the resulting RNA-RNA duplex to gain the thermodynamic stability. Since G:U wobbles are common in RNA structure but detrimental to translational repression, alignments with G:U wobbles are down-weighted. Taken further more in consideration the conservation of miRNA and target gene, this method reaches an outcome of 35% of false positive results regarding genes with one target site and 9% of false positives of genes with more than one target site. Filtering out non-conserved pairs minimizes the error rate of this method ADDIN EN.CITE <sup>120</sup>.

Another prediction tool used in this project is TargetScan<sup>122</sup>. It uses alignments of miRNA seed region to highly conserved mRNA sites over several species. The scoring depends on the level of target conservation, the binding to miRNA on the 8<sup>th</sup> sequence position, if there is an adenosine at the first binding position, the distance from target site to 3'UTR end and the AU composition of flanking area. This method still has an outcome of 31% false positive results in human ADDIN EN.CITE <sup>123</sup>.

A comparison of different target prediction tools in 2009 showed a precision (correctly predicted/totally predicted) of TargetScan results of around 50% and a sensitivity (correctly predicted/total correct) of around 10%. MiRanda showed a precision around 30% and a sensitivity of around 20%. TargetScan mainly predicts by evolutionary conservation of seed region while miRanda mainly predicts by alignment and thermodynamic stability of



complexes. To improve predictivity we combined different prediction tools to increase precision rate. Alexiou et al. tested the effectiveness of combined prediction tools in 2009. They found that using one single prediction tool offers better sensitivity but lower specificity compared to a combination of more<sup>124</sup>. Thus a combination of prediction tools could lead to loss of targets, but would be more powerful for identification of true targets. As the high rate of false positive predictions observed in this and other studies indicates a combination and/or improvement of prediction tools is clearly necessary to identify true miRNA target genes.

A major problem using conserved binding sites for target prediction was described by Pinzón et al. in 2016 ADDIN EN.CITE <sup>125</sup>. They observed that many mRNAs containing conserved miRNA binding sites function as miRNA titrators or competing endogenous RNAs (ceRNAs). Repression of these mRNA targets has no functional effect. They indirectly regulate the expression of other mRNAs by competing for shared miRNAs ADDIN EN.CITE <sup>125,126</sup>.

To increase the probability of receiving true miRNA predicted targets, an Argonaut2-Par-CLIP experiment could be performed. This experiment offers mRNA sequences binding to the miRNA seed region in the RISC complex. However, due to scarcity of patient material the high numbers of patient cells, which are needed for Par-CLIP, represent a problem. To avoid this problem appropriate cell lines could be used like the TCF3-HLF positive cell line HAL-01 for patients carrying t(17;19) and the TCF3-PBX1 positive 697 cell line for patients carrying t(1;19). Sequence information gained from a Par-CLIP experiment using cell lines could be compared to predicted targets of rare patient material. A sequence of a real target gene should be gained by Ago2-ParClip, because this target gene binds to its miRNA for RISC-mediated gene silencing.

For validation of a potential miRNA-mRNA target pair a biotinylated miRNA-pull down assay could be performed, in which cells are transfected with the biotinylated miRNA of interest. After harvesting transfected cells, the miRNA of interest and its bound mRNA targets are separated from remaining RNA by pull down and isolated. Obtained RNA is transcribed into cDNA and can be analysed by qPCR to get information about the target genes<sup>127</sup>.

Furthermore a higher number of patient samples would be helpful to fully understand the differences of t(1;19) and t(17;19) acute lymphoblastic leukemia. Interpretation of NGS results need statistical analysis and the higher number of patients included into this analysis, the more significant results can be obtained. Since t(17;19) is extremely rare in ALL patients, it is hard to obtain such samples. Samples of patient derived xenografts already exist and

would be an alternative to primary patient material. It is shown that the mutational profile as well as the gene expression pattern of patient tumor samples is maintained in xenograft samples ADDIN EN.CITE <sup>110</sup>.

Another point, which should be considered in future work, is the translational inhibition of mRNA accomplished by miRNAs. This miRNA effect was first described in 1993 by Lee and Wightman, who observed a reduction of LIN-14 protein after *LIN-4* miRNA expression and without reduction of *LIN-14* mRNA ADDIN EN.CITE <sup>67,128,129</sup>. Translational expression does not affect the level of target mRNA, but the level of respective protein after sequence-specific miRNA-mRNA binding. These mRNA targets can also be predicted by computational programs, but not verified by qPCR or luciferase assay. In this work, just miRNA mediated mRNA decrease was analyzed. An analysis of miRNA mediated translational inhibition by qualitative and quantitative proteomic analysis for example by mass spectrometry is another challenge for future studies.

## **5.2 Novel miRNAs in t(1;19) and t(17;19) ALL**

The detection of novel miRNAs was computationally performed with miRDeep. Validation tests were performed with Northern blot and qPCR but none of the 18 novel miRNAs found in the sequencing data was detectable in Northern blot. Two of the most interesting potential novel miRNAs were also tried to be validated by the more sensitive qPCR method. Novel miR10 was just detected in t(1;19)-positive tumor samples, while miR14 was just detected in t(17;19)-positive tumor samples. The existence of both of them could not be confirmed by using qRT-PCR. Because patient material was not available, validation experiments had to be performed with leukemia cell lines. For this reason the possibility exists, that the potential novel miRNAs are real, but not expressed in the used cell lines. An alternative explanation would be that the potential novel miRNAs are sequencing artefacts and therefore not detectable. However, as the expression of these novel miRNAs was not very high in patient material (<<1000 reads, not normalized), the chosen validation methods may not have been sensitive enough to detect the novel miRNAs in the used cell lines, although the highly sensitive qPCR-method was used for the most interesting novel miRNAs. If the expression of these microRNAs is so low, that it cannot be detected by qPCR, the question arises, whether these miRNAs can have an important impact on tumor development or other functional processes. To analyze a possible functional effect, the potential novel miRNAs could be

overexpressed in the mentioned ALL cell lines with subsequent functional analysis.

Novel bioinformatic tools, yet to be developed, can hopefully soon be used to reanalyze the data to find novel miRNAs in the leukemic patient samples. Furthermore, for future validations of novel miRNAs primary patient material should be used, because appropriate cell lines are not identical to patient material.

### **5.3 Mutated miRNAs in t(1;19) and t(17;19)**

The analysis of mutations in miRNAs revealed two mutations in hsa-miR-7847 occurring in t(1;19) tumor material. One of them lies in the miRNA-seed region, which is complementary to the mRNA target site and important for mRNA binding and degradation. The miRNA has been reported to be overexpressed in late-stage gastric cancer cells compared to healthy cells<sup>130</sup> and in the chemoresistant breast cancer cell line MDA-MB-231 compared to the chemosensitive MDA-MB-21 cell line. hsa-miR-7847 has a putative effect on the WNT-signaling pathway ADDIN EN.CITE<sup>131</sup>. In future studies novel patient material should especially be analyzed regarding these two mutations. An overexpression in an ALL-cell line of wildtype hsa-miR-7847 and of hsa-miR-7847 containing one of the two described mutations could be helpful for the analysis of mutational effects. Genomic modification by gene editing is another possibility to study mutational effects by inserting point mutations into specific sites of the genome. CRISPR/CAS or zink finger nucleases are used to create double strand breaks at specific genomic sites. By using a required template, the cell inserts a specific mutation by homologous recombination. Also, functional analysis regarding potential gene targets and chemoresistance should be performed.

### **5.4 Treatment of t(17;19) ALL with HDAC6 inhibitor KSK**

*TCF3/HLF*-positive ALL is a rare and lethal subtype of ALL, for which no drugs are available at the moment. Previous analyses revealed an increased apoptosis rate of the *TCF3/HLF*-positive cell line HAL-01 after treatment with the novel HDAC6 inhibitor KSK, compared to healthy B-cells. This makes KSK a potential therapeutic drug for treatment of t(17;19)-ALL. To analyze the impact of KSA on gene expression of t(17;19)-positive cells, HAL-01 cell line was treated with KSK for 24h and sequenced on an Illumina platform. Seventeen genes were differentially expressed by a gfold  $\geq 2$  after 24 h of inhibitor treatment

in HAL-01. However, this number represents a low number of changes considering the general mechanism of HDACs. Tests of other HDAC inhibitors on different cancer cell lines showed similar results ADDIN EN.CITE <sup>132,133</sup>, leading to the assumption that HDAC inhibitor treatment does not cause global changes in gene expression, but with strong impact, as it was shown in the apoptosis assay.

Since we only observed a downregulation of gene expression, we can assume that after 24 h of HDAC6 inhibitor treatment, secondary effects can already be observed. Hyperacetylation could improve the expression of transcriptional repressors leading to a downregulation of observed genes. Downregulation of genes after HDAC inhibitor treatment was also observed by other researchers ADDIN EN.CITE <sup>132-135</sup> and should be further analyzed for example by harvesting treated cells in a dense time series.

Since it is known that HDAC6 predominantly translocates to the cytoplasm ADDIN EN.CITE <sup>136</sup> and has a function on tubulin deacetylation, and other non-histone substrates ADDIN EN.CITE <sup>137</sup>, the observed differential gene expression could also be the effect of inhibiting deacetylation of non-histone substrates like HSP90. HSP90 promotes maturation and stabilization of protein structures ADDIN EN.CITE <sup>138</sup>. Reversible acetylation modulates chaperone activity. Since HSP90 is known to stabilize transcription factors, it is possible that it is also involved in stabilization of the fusion constructs of *TCF3* with *PBX* or *HLF*. Loss of chaperone activity by hypermethylation, caused by inhibition of HDAC6, could therefore lead to an increased reduction of the fusion construct and a subsequent change in gene expression.

Uncontrolled function of histone acetyltransferases (HATS), HDACs or their binding partners can trigger leukemic development in a many cases ADDIN EN.CITE <sup>139,140</sup>. The effect of gene silencing caused by HDACs can be abolished by the use of HDAC inhibitors, which interact with the catalytic site of HDACs leading to a hypermethylation of histones. Different HDAC inhibitors have already been tested and show a strong antitumor effect *in vitro* and *in vivo* ADDIN EN.CITE <sup>141,142</sup>. Furthermore, they induce remission in transgenic leukemic mouse models ADDIN EN.CITE <sup>143</sup>. By promoting gene expression of repressed genes, HDAC inhibitors stimulate growth arrest, differentiation and apoptosis in tumor cell lines while having no or little effect on normal cells<sup>144</sup>. Further functional assays regarding differentiation, proliferation and apoptosis are planned in the HAL-01 cell line treated with KSK as well as the effect of the HDAC inhibitor on xenografts derived from primary patient leukemia cells in mouse models.

## 5.5 Outlook

To fully analyze the differences between *TCF3/PBX1* and *TCF3/HLF* leukemic patients, follow-up studies are needed, especially for mRNA/miRNA target pair validation. Patient material is extremely rare, particularly material of *TCF3/HLF* patients. Ideally, this study can be continued with new patient material to confirm and optimize results and to perform functional studies with new patient material. Alternatively or additionally immortalized cell lines and patient-derived xenograft samples can be used for further studies. The aim is, to get deeper insights into the posttranscriptional effects on development of both tumor subtypes which hopefully identifies new therapeutic targets and increases therapeutic success.

## 5.6 Differences between Infant and Childhood T-ALL

The development of a T-ALL in infancy is very rare and accompanied with a poor prognosis. To detect differences between iTALL and childhood T-ALL, which has a much better prognosis, 3 infant and 6 childhood T-ALL samples were analyzed by NGS.

By whole exome sequencing we detected a series of oncogenic mutations in the three infant patients including *NOTCH2* and *NOTCH3* mutations. A *NOTCH3* mutation was found in all three infant patients. We could not detect activating *NOTCH1* mutations, which are the most frequent oncogenic mutation in childhood T-ALL (around 60%) and mostly occur in the heterodimerization (HD) or the proline, glutamic acid, serine, threonine-rich (PEST) domain ADDIN EN.CITE <sup>65</sup>. We could further not detect any F-box and WB repeat domain containing 7 (*FBXW7*) mutations, which occur frequently in childhood T-ALL ADDIN EN.CITE <sup>145</sup> and are described to occur in some iT-ALL patients ADDIN EN.CITE <sup>66</sup>. We did not find *CDKN2A* mutations in the iT-ALL samples, underlining the suggestion of Mansur et al. that *CDKN2A* mutations have a subordinate role in iT-ALL compared to childhood T-ALL or B lineage ALL ADDIN EN.CITE <sup>146</sup>. Recent publications described non-synonymous mutations in C2 domain of *PTEN*, which cause C-terminal truncation of *PTEN* protein in T-ALL ADDIN EN.CITE <sup>147</sup>. Some deletions in the *PTEN* gene are also associated with treatment failure and have thus an impact on patient outcome. Around 13% to 20% of pediatric T-ALL patients show mutations in *PTEN* gene ADDIN EN.CITE <sup>148,149</sup>. We detected short deletions in *PTEN* gene in two out of three infant T-ALL patients and insertions in all three infant patients, which represents a high proportion combined to pediatric

cases with *PTEN* mutations.

We did not detect deletions or differential expression of the myeloid leukemia factor 1 (*MLF1*) gene in the iT-ALL samples as described by Mansur et al. in 2015 ADDIN EN.CITE<sup>66</sup>.

They found a loss of 3q25.32 in around a quarter of iT-ALL samples which completely delete the *MLF1* gene and is unique for infant T-ALL samples. MLF1 can act as a tumor suppressor preventing degradation of p53 by E3 ubiquitin ligase and therefore promoting p53-dependent cell cycle arrest ADDIN EN.CITE<sup>150</sup>. For this reason *MLF1* gene deletion should be analyzed in future available iT-ALL sample material.

By NGS we detected different aberrant mechanisms in infant and childhood T-ALL. Both groups were separated using hierarchical clustering based on differentially expressed genes and miRNAs. We also found several pathways, which were differentially affected by aberrant expression in both groups. These facts indicate the difference between infant and childhood T-ALL on a genetic and epigenetic level. Upstream regulator analysis showed many *TLR2/4* target genes, which were downregulated in iT-ALL, indicating an inhibition of *TLR2/4*. Both toll-like receptors are not significantly downregulated in infant T-ALL on RNA level, but we found mutations in both genes, which might explain the downregulation of target genes without a downregulation of *TLR2/4*. *TLR2/4* are pattern-recognition receptors, which get activated by binding to surface molecules on pathogens leading to an activation of the immune response<sup>151</sup>. The mutation in *TLR2* is already known and predicted to be a deleterious missense mutation (rs5743708). Loss of *TLR2* function is already described to support cancer development, for example a promotion of liver cancer in mice ADDIN EN.CITE<sup>152</sup>. The *TLR4* mutation is also known (rs5900307) and lies within an intronic region. In hepatocarcinoma patients, a downregulation of both TLRs was detected ADDIN EN.CITE<sup>153</sup>. A connection between TLRs and cancer is described by Chen et. al: TLRs initiate signaling pathways leading to cell proliferation and chemoresistance ADDIN EN.CITE<sup>154</sup>. The function of TLRs in T-cells or T-lymphoblasts is still unclear. By analyzing expression values between T-ALL blasts and normal lymphocytes (including T-cells, B-cells and natural killer cells) using the R2 MegaSampler a downregulation of both TLRs can be observed in T-ALL blasts. *TLR2* is downregulated more than 100-fold compared to normal lymphocytes. *TLR4* is downregulated more than 15-fold compared to normal lymphocytes. Because of the observed downregulation, the assumption can be made that downstream target inhibition by downregulation of *TLR2/4* or by mutational inactivation can have an important influence on T-ALL development in childhood, adult as well as infant T-ALL. Therefore, *TLR2/4* function

on leukemic development should be analyzed in further experiments.

Some of the differentially expressed genes in iT-ALL can serve as potentially druggable targets. One of these genes is *KIT*, for which four approved drugs (Imatinib, Dasitinib, Sorafenib, Pazopanib) are already available. *KIT* encodes the proto-oncogene c-Kit, a type 3 transmembrane receptor which binds its ligand stem cell factor (SCF)<sup>155</sup>. Certain mutations in this gene lead to drug resistance of TKIs like Imatinib ADDIN EN.CITE <sup>156-158</sup>. Therefore, we analyzed the iT-ALL samples for any *KIT* mutations, but could not detect any. Because we identified a higher gene expression of *KIT* in the infant cases compared to the childhood T-ALL patients, an approved TKI like Imatinib could be used as add-on to standard therapy.

A target prediction of the differentially expressed mRNA and miRNA showed 47 potential target pairs with a correlation coefficient of -0.6 or less. Most of these target pairs are associated with cancer development or cancer progression.

A high expression of miRNA hsa-let-7b-5p targets mRNA of *GREB1*, a gene which is described in breast ADDIN EN.CITE <sup>159,160</sup>, ovarian ADDIN EN.CITE <sup>161</sup> and prostate ADDIN EN.CITE <sup>162</sup> cancer. A high expression of *GREB1* was described to increase proliferation of breast cancer cells ADDIN EN.CITE <sup>159</sup>, ADDIN EN.CITE <sup>163</sup> while knockdown of *GREB1* decreased tumor growth in ovarian cancer cells *in vitro* and in mouse xenografts ADDIN EN.CITE <sup>164</sup>. In infant patients, we observed a downregulation of hsa-let7-5p, which potentially leads to observed upregulation of *GREB1* in infant patients. Another potential target of hsa-let7b-5p, which is also upregulated in iT-ALL samples, is *IGF2BP1*. This gene is also overexpressed in *ETV6/RUNX1* ALL subtype ADDIN EN.CITE <sup>165</sup> and aberrantly expressed by fusion with IGH locus ADDIN EN.CITE <sup>166,167</sup>.

Hsa-miR-31 is downregulated in iT-ALL compared to childhood T-ALL. A downregulation of hsa-miR-31 was already observed in adult T-ALL with impact on NF-kB-signaling. Hsa-miR-31 downregulation also leads to promotion of apoptosis resistance in adult T-ALL ADDIN EN.CITE <sup>168</sup>. In iT-ALL hsa-miR31 targets *RPA1*, a gene important for DNA damage respond during DNA replication ADDIN EN.CITE <sup>169</sup>. This gene is also described to be upregulated in CLL ADDIN EN.CITE <sup>170,171</sup>.

By using NGS on T-ALL patient exomes, transcriptomes and miRNomes, the landscape of iTALL was revealed. Differentially expressed genes and miRNAs were detected between iTALL and childhood T-ALL leading to differences in important signaling pathways. Our

study shows a substantial difference on genetic and epigenetic level between iTALL and childhood T-ALL and gives a first idea of optimized treatment strategies for cure of this rare, high-risk disease.

## 6. Literature

- ADDIN EN.REFLIST 1 Orkin, S. H. Diversification of haematopoietic stem cells to specific lineages. *Nature reviews. Genetics* **1**, 57-64, doi:10.1038/35049577 (2000).
- 2 Rieger, M. A. & Schroeder, T. Hematopoiesis. *Cold Spring Harbor perspectives in biology* **4**, doi:10.1101/cshperspect.a008250 (2012).
- 3 Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631-644, doi:10.1016/j.cell.2008.01.025 (2008).
- 4 Passegue, E., Jamieson, C. H., Ailles, L. E. & Weissman, I. L. Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? *Proceedings of the National Academy of Sciences of the United States of America* **100 Suppl 1**, 11842-11849, doi:10.1073/pnas.2034201100 (2003).
- 5 Randolph, T. R. Advances in acute lymphoblastic leukemia. *Clinical laboratory science : journal of the American Society for Medical Technology* **17**, 235-245 (2004).
- 6 Bennett, J. M. *et al.* Proposals for Classification of Acute Leukemias. *Brit J Haematol* **33**, 451-&, doi:DOI 10.1111/j.1365-2141.1976.tb03563.x (1976).
- 7 Vardiman, J. W. *et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937-951, doi:10.1182/blood-2009-03-209262 (2009).
- 8 Smith, M. *et al.* Uniform approach to risk classification and treatment assignment for children with acute lymphoblastic leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **14**, 18-24, doi:10.1200/JCO.1996.14.1.18 (1996).
- 9 Reiter, A. *et al.* Chemotherapy in 998 unselected childhood acute lymphoblastic leukemia patients. Results and conclusions of the multicenter trial ALL-BFM 86. *Blood* **84**, 3122-3133 (1994).
- 10 Trueworthy, R. *et al.* Ploidy of lymphoblasts is the strongest predictor of treatment outcome in B-progenitor cell acute lymphoblastic leukemia of childhood: a Pediatric Oncology Group study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **10**, 606-613, doi:10.1200/JCO.1992.10.4.606 (1992).
- 11 Schultz, K. R. *et al.* Risk- and response-based classification of childhood B-precursor acute lymphoblastic leukemia: a combined analysis of prognostic markers from the Pediatric Oncology Group (POG) and Children's Cancer Group (CCG). *Blood* **109**, 926-935, doi:10.1182/blood-2006-01-024729 (2007).
- 12 Crist, W. *et al.* Pre-B cell leukemia responds poorly to treatment: a pediatric oncology group study. *Blood* **63**, 407-414 (1984).
- 13 Schmiegelow, K. *et al.* Long-term results of NOPHO ALL-92 and ALL-2000 studies of childhood acute lymphoblastic leukemia. *Leukemia* **24**, 345-354, doi:10.1038/leu.2009.251 (2010).
- 14 Pui, C. H., Carroll, W. L., Meshinchi, S. & Arceci, R. J. Biology, risk stratification, and therapy of pediatric acute leukemias: an update. *J Clin Oncol* **29**, 551-565, doi:10.1200/JCO.2010.30.7405 (2011).



- 15 Vogler, L. B. *et al.* Pre-B-cell leukemia. A new phenotype of childhood lymphoblastic leukemia. *The New England journal of medicine* **298**, 872-878, doi:10.1056/NEJM197804202981603 (1978).
- 16 Uckun, F. M. *et al.* Human t(1;19)(q23;p13) pre-B acute lymphoblastic leukemia in mice with severe combined immunodeficiency. *Blood* **81**, 3052-3062 (1993).
- 17 Borowitz, M. J. *et al.* Predictability of the t(1;19)(q23;p13) from surface antigen phenotype: implications for screening cases of childhood acute lymphoblastic leukemia for molecular analysis: a Pediatric Oncology Group study. *Blood* **82**, 1086-1091 (1993).
- 18 Foon, K. A., Schroff, R. W. & Gale, R. P. Surface markers on leukemia and lymphoma cells: recent advances. *Blood* **60**, 1-19 (1982).
- 19 Pui, C. H. & Evans, W. E. Acute lymphoblastic leukemia. *The New England journal of medicine* **339**, 605-615, doi:10.1056/NEJM199808273390907 (1998).
- 20 Hunger, S. P., Ohyashiki, K., Toyama, K. & Cleary, M. L. Hlf, a novel hepatic bZIP protein, shows altered DNA-binding properties following fusion to E2A in t(17;19) acute lymphoblastic leukemia. *Genes & development* **6**, 1608-1620, doi:10.1101/gad.6.9.1608 (1992).
- 21 Pui, C. H. *et al.* Cytogenetic features and serum lactic dehydrogenase level predict a poor treatment outcome for children with pre-B-cell leukemia. *Blood* **67**, 1688-1692 (1986).
- 22 Raimondi, S. C. *et al.* Cytogenetics of pre-B-cell acute lymphoblastic leukemia with emphasis on prognostic implications of the t(1;19). *J Clin Oncol* **8**, 1380-1388, doi:10.1200/JCO.1990.8.8.1380 (1990).
- 23 Crist, W. M. *et al.* Poor prognosis of children with pre-B acute lymphoblastic leukemia is associated with the t(1;19)(q23;p13): a Pediatric Oncology Group study. *Blood* **76**, 117-122 (1990).
- 24 Ohno, H. *et al.* Acute lymphoblastic leukemia associated with a t(1;19)(q23;p13) in an adult. *Internal medicine* **32**, 584-587 (1993).
- 25 Mitani, K. *et al.* Molecular analysis of the t(1;19)(q23;p13) translocation observed in adult leukemias. *International journal of hematology* **60**, 267-271 (1994).
- 26 Mertens, F., Johansson, B. & Mitelman, F. Age- and gender-related heterogeneity of cancer chromosome aberrations. *Cancer genetics and cytogenetics* **70**, 6-11 (1993).
- 27 Hayashi, Y., Hanada, R., Yamamoto, K., Taguchi, N. & Shikano, T. Leukoencephalopathy in children with t(1;19) acute lymphoblastic leukaemia. *Lancet* **340**, 316 (1992).
- 28 Yen, H. J. *et al.* Pediatric acute lymphoblastic leukemia with t(1;19)/TCF3-PBX1 in Taiwan. *Pediatric blood & cancer* **64**, doi:10.1002/pbc.26557 (2017).
- 29 Carroll, A. J. *et al.* Pre-B cell leukemia associated with chromosome translocation 1; 19. *Blood* **63**, 721-724 (1984).
- 30 Pui, C. H., Behm, F. G. & Crist, W. M. Clinical and biologic relevance of immunologic marker studies in childhood acute lymphoblastic leukemia. *Blood* **82**, 343-362 (1993).
- 31 Mellentin, J. D. *et al.* The gene for enhancer binding proteins E12/E47 lies at the t(1; 19) breakpoint in acute leukemias. *Science* **246**, 379-382 (1989).
- 32 Pui, C. H. *et al.* Immunologic, cytogenetic, and clinical characterization of childhood acute lymphoblastic leukemia with the t(1;19) (q23; p13) or its derivative. *J Clin Oncol* **12**, 2601-2606, doi:10.1200/JCO.1994.12.12.2601 (1994).
- 33 Williams, D. L. *et al.* New chromosomal translocations correlate with specific immunophenotypes of childhood acute lymphoblastic leukemia. *Cell* **36**, 101-109 (1984).

- 34 Nourse, J. *et al.* Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. *Cell* **60**, 535-545 (1990).
- 35 Kamps, M. P., Murre, C., Sun, X. H. & Baltimore, D. A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. *Cell* **60**, 547-555 (1990).
- 36 Murre, C., McCaw, P. S. & Baltimore, D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* **56**, 777-783 (1989).
- 37 Bain, G., Gruenwald, S. & Murre, C. E2A and E2-2 are subunits of B-cell-specific E2-box DNA-binding proteins. *Molecular and cellular biology* **13**, 3522-3529 (1993).
- 38 Sawada, S. & Littman, D. R. A heterodimer of HEB and an E12-related protein interacts with the CD4 enhancer and regulates its activity in T-cell lines. *Molecular and cellular biology* **13**, 5620-5628 (1993).
- 39 Monica, K., Galili, N., Nourse, J., Saltman, D. & Cleary, M. L. PBX2 and PBX3, new homeobox genes with extensive homology to the human proto-oncogene PBX1. *Molecular and cellular biology* **11**, 6149-6157 (1991).
- 40 Lu, Q., Wright, D. D. & Kamps, M. P. Fusion with E2A converts the Pbx1 homeodomain protein into a constitutive transcriptional activator in human leukemias carrying the t(1;19) translocation. *Molecular and cellular biology* **14**, 3938-3948 (1994).
- 41 Hunger, S. P. *et al.* The t(1;19)(q23;p13) results in consistent fusion of E2A and PBX1 coding sequences in acute lymphoblastic leukemias. *Blood* **77**, 687-693 (1991).
- 42 Izraeli, S., Kovar, H., Gadner, H. & Lion, T. Unexpected heterogeneity in E2A/PBX1 fusion messenger RNA detected by the polymerase chain reaction in pediatric patients with acute lymphoblastic leukemia. *Blood* **80**, 1413-1417 (1992).
- 43 Kamps, M. P., Look, A. T. & Baltimore, D. The human t(1;19) translocation in pre-B ALL produces multiple nuclear E2A-Pbx1 fusion proteins with differing transforming potentials. *Genes & development* **5**, 358-368 (1991).
- 44 Suci, S., Weh, H. J. & Hossfeld, D. K. T-cell acute lymphoblastic leukemia with translocation (1;18). *Cancer genetics and cytogenetics* **30**, 165-169 (1988).
- 45 Privitera, E. *et al.* Different molecular consequences of the 1;19 chromosomal translocation in childhood B-cell precursor acute lymphoblastic leukemia. *Blood* **79**, 1781-1788 (1992).
- 46 Izraeli, S. *et al.* Expression of identical E2A/PBX1 fusion transcripts occurs in both pre-B and early pre-B immunological subtypes of childhood acute lymphoblastic leukemia. *Leukemia* **7**, 2054-2056 (1993).
- 47 Kamps, M. P. & Baltimore, D. E2A-Pbx1, the t(1;19) translocation protein of human pre-B-cell acute lymphocytic leukemia, causes acute myeloid leukemia in mice. *Molecular and cellular biology* **13**, 351-357 (1993).
- 48 Chiaretti, S., Zini, G. & Bassan, R. Diagnosis and subclassification of acute lymphoblastic leukemia. *Mediterranean journal of hematology and infectious diseases* **6**, e2014073, doi:10.4084/MJHID.2014.073 (2014).
- 49 Nadler, L. M. *et al.* Induction of human B cell antigens in non-T cell acute lymphoblastic leukemia. *The Journal of clinical investigation* **70**, 433-442 (1982).
- 50 Kohlmann, A. *et al.* New insights into MLL gene rearranged acute leukemias using gene expression profiling: shared pathways, lineage commitment, and partner genes. *Leukemia* **19**, 953-964, doi:10.1038/sj.leu.2403746 (2005).
- 51 Look, A. T. Oncogenic transcription factors in the human acute leukemias. *Science* **278**, 1059-1064, doi:DOI 10.1126/science.278.5340.1059 (1997).

- 52 Matsunaga, T. *et al.* Regulation of annexin II by cytokine-initiated signaling pathways and E2A-HLF oncoprotein. *Blood* **103**, 3185-3191, doi:10.1182/blood-2003-09-3022 (2004).
- 53 Yeung, J. *et al.* Characterization of the t(17;19) translocation by gene-specific fluorescent in situ hybridization-based cytogenetics and detection of the E2A-HLF fusion transcript and protein in patients' cells. *Haematologica* **91**, 422-424 (2006).
- 54 Inukai, T. *et al.* Hypercalcemia in childhood acute lymphoblastic leukemia: frequent implication of parathyroid hormone-related peptide and E2A-HLF from translocation 17;19. *Leukemia* **21**, 288-296, doi:10.1038/sj.leu.2404496 (2007).
- 55 Glover, J. M., Loriaux, M., Tyner, J. W., Druker, B. J. & Chang, B. H. In vitro sensitivity to dasatinib in lymphoblasts from a patient with t(17;19)(q22;p13) gene rearrangement pre-B acute lymphoblastic leukemia. *Pediatric blood & cancer* **59**, 576-579, doi:10.1002/pbc.23383 (2012).
- 56 Raimondi, S. C. *et al.* New recurring chromosomal translocations in childhood acute lymphoblastic leukemia. *Blood* **77**, 2016-2022 (1991).
- 57 Thompson, M. A. & Davé, U. P. *Molecular Genetics of Acute Leukemia*, <<https://oncohemakey.com/molecular-genetics-of-acute-leukemia/>> (2016).
- 58 Hunger, S. P., Li, S., Fall, M. Z., Naumovski, L. & Cleary, M. L. The proto-oncogene HLF and the related basic leucine zipper protein TEF display highly similar DNA-binding and transcriptional regulatory properties. *Blood* **87**, 4607-4617 (1996).
- 59 Goldberg, J. M. *et al.* Childhood T-cell acute lymphoblastic leukemia: the Dana-Farber Cancer Institute acute lymphoblastic leukemia consortium experience. *J Clin Oncol* **21**, 3616-3622, doi:10.1200/JCO.2003.10.116 (2003).
- 60 Aifantis, I., Raetz, E. & Buonamici, S. Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nature reviews. Immunology* **8**, 380-390, doi:10.1038/nri2304 (2008).
- 61 Uckun, F. M. *et al.* Clinical features and treatment outcome of childhood T-lineage acute lymphoblastic leukemia according to the apparent maturational stage of T-lineage leukemic blasts: a Children's Cancer Group study. *J Clin Oncol* **15**, 2214-2221, doi:10.1200/JCO.1997.15.6.2214 (1997).
- 62 Winick, N. J., Carroll, W. L. & Hunger, S. P. Childhood leukemia--new advances and challenges. *The New England journal of medicine* **351**, 601-603, doi:10.1056/NEJMe048154 (2004).
- 63 Burger, R., Hansen-Hagge, T. E., Drexler, H. G. & Gramatzki, M. Heterogeneity of T-acute lymphoblastic leukemia (T-ALL) cell lines: suggestion for classification by immunophenotype and T-cell receptor studies. *Leukemia research* **23**, 19-27 (1999).
- 64 Girardi, T., Vicente, C., Cools, J. & De Keersmaecker, K. The genetics and molecular biology of T-ALL. *Blood* **129**, 1113-1123, doi:10.1182/blood-2016-10-706465 (2017).
- 65 Weng, A. P. *et al.* Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science* **306**, 269-271, doi:10.1126/science.1102160 (2004).
- 66 Mansur, M. B. *et al.* Distinctive genotypes in infants with T-cell acute lymphoblastic leukaemia. *British journal of haematology* **171**, 574-584, doi:10.1111/bjh.13613 (2015).
- 67 Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843-854 (1993).
- 68 Peter Nelson, M. K., Anup Sharma, Elsa Maniataki, Zissimos Mourelatos. The microRNA world: small is mighty. *CellPress* **28**, 6 (2003).
- 69 Ballantyne, M. D., McDonald, R. A. & Baker, A. H. lncRNA/MicroRNA interactions in the vasculature. *Clinical pharmacology and therapeutics* **99**, 494-501, doi:10.1002/cpt.355 (2016).

- 70 Calin, G. A. *et al.* Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2999-3004, doi:10.1073/pnas.0307323101 (2004).
- 71 Calin, G. A. *et al.* MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 11755-11760, doi:10.1073/pnas.0404432101 (2004).
- 72 MacDonagh, L. *et al.* The emerging role of microRNAs in resistance to lung cancer treatments. *Cancer treatment reviews* **41**, 160-169, doi:10.1016/j.ctrv.2014.12.009 (2015).
- 73 Sun, X., Charbonneau, C., Wei, L., Chen, Q. & Terek, R. M. miR-181a Targets RGS16 to Promote Chondrosarcoma Growth, Angiogenesis, and Metastasis. *Molecular cancer research : MCR* **13**, 1347-1357, doi:10.1158/1541-7786.MCR-14-0697 (2015).
- 74 Ju, X. *et al.* Differential microRNA expression in childhood B-cell precursor acute lymphoblastic leukemia. *Pediatric hematology and oncology* **26**, 1-10, doi:10.1080/08880010802378338 (2009).
- 75 Schotte, D. *et al.* Identification of new microRNA genes and aberrant microRNA profiles in childhood acute lymphoblastic leukemia. *Leukemia* **23**, 313-322, doi:10.1038/leu.2008.286 (2009).
- 76 Jones, K., Nourse, J. P., Keane, C., Bhatnagar, A. & Gandhi, M. K. Plasma microRNA are disease response biomarkers in classical Hodgkin lymphoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* **20**, 253-264, doi:10.1158/1078-0432.CCR-13-1024 (2014).
- 77 Janssen, H. L. *et al.* Treatment of HCV infection by targeting microRNA. *The New England journal of medicine* **368**, 1685-1694, doi:10.1056/NEJMoa1209026 (2013).
- 78 Beg, M. S. *et al.* Phase I study of MRX34, a liposomal miR-34a mimic, administered twice weekly in patients with advanced solid tumors. *Investigational new drugs* **35**, 180-188, doi:10.1007/s10637-016-0407-y (2017).
- 79 Carvalho de Oliveira, J. *et al.* MiRNA Dysregulation in Childhood Hematological Cancer. *International journal of molecular sciences* **19**, doi:10.3390/ijms19092688 (2018).
- 80 Ohyashiki, K. *et al.* Establishment of a novel heterotransplantable acute lymphoblastic leukemia cell line with a t(17;19) chromosomal translocation the growth of which is inhibited by interleukin-3. *Leukemia* **5**, 322-331 (1991).
- 81 Matsuo, Y. & Drexler, H. G. Establishment and characterization of human B cell precursor-leukemia cell lines. *Leukemia research* **22**, 567-579 (1998).
- 82 Adams, R. A., Flowers, A. & Davis, B. J. Direct implantation and serial transplantation of human acute lymphoblastic leukemia in hamsters, SB-2. *Cancer research* **28**, 1121-1125 (1968).
- 83 Adams, R. A. *et al.* The question of stemlines in human acute leukemia. Comparison of cells isolated in vitro and in vivo from a patient with acute lymphoblastic leukemia. *Experimental cell research* **62**, 5-10 (1970).
- 84 Rosenfeld, C. *et al.* Phenotypic characterisation of a unique non-T, non-B acute lymphoblastic leukaemia cell line. *Nature* **267**, 841-843 (1977).
- 85 Greil, J. *et al.* The acute lymphoblastic leukaemia cell line SEM with t(4;11) chromosomal rearrangement is biphenotypic and responsive to interleukin-7. *Br J Haematol* **86**, 275-283 (1994).
- 86 Reichel, M. *et al.* Fine structure of translocation breakpoints in leukemic blasts with chromosomal translocation t(4;11): the DNA damage-repair model of translocation. *Oncogene* **17**, 3035-3044, doi:10.1038/sj.onc.1202229 (1998).

- 87 Drexler, H. G., Quentmeier, H. & MacLeod, R. A. Malignant hematopoietic cell lines: in vitro models for the study of MLL gene alterations. *Leukemia* **18**, 227-232, doi:10.1038/sj.leu.2403236 (2004).
- 88 Findley, H. W., Jr., Cooper, M. D., Kim, T. H., Alvarado, C. & Ragab, A. H. Two new acute lymphoblastic leukemia cell lines with early B-cell phenotypes. *Blood* **60**, 1305-1309 (1982).
- 89 Rio, D. C., Clark, S. G. & Tjian, R. A mammalian host-vector system that regulates expression and amplification of transfected genes by temperature induction. *Science* **227**, 23-28 (1985).
- 90 DuBridge, R. B. *et al.* Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Molecular and cellular biology* **7**, 379-387 (1987).
- 91 Pear, W. S., Nolan, G. P., Scott, M. L. & Baltimore, D. Production of high-titer helper-free retroviruses by transient transfection. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 8392-8396 (1993).
- 92 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, doi:10.14806/ej.17.1.200
- pp. 10-12 (2011).
- 93 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 94 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 95 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).
- 96 Yates, A. *et al.* Ensembl 2016. *Nucleic acids research* **44**, D710-716, doi:10.1093/nar/gkv1157 (2016).
- 97 Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic acids research* **39**, D152-157, doi:10.1093/nar/gkq1027 (2011).
- 98 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
- 99 Wong, N. & Wang, X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic acids research* **43**, D146-152, doi:10.1093/nar/gku1104 (2015).
- 100 Vergoulis, T. *et al.* TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic acids research* **40**, D222-229, doi:10.1093/nar/gkr1161 (2012).
- 101 Garcia, D. M. *et al.* Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nature structural & molecular biology* **18**, 1139-1146, doi:10.1038/nsmb.2115 (2011).
- 102 Chou, C. H. *et al.* miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic acids research* **44**, D239-247, doi:10.1093/nar/gkv1258 (2016).
- 103 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).
- 104 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 105 Duraku, L. S. *et al.* Re-innervation patterns by peptidergic Substance-P, non-peptidergic P2X3, and myelinated NF-200 nerve fibers in epidermis and dermis of rats

- with neuropathic pain. *Experimental neurology* **241**, 13-24, doi:10.1016/j.expneurol.2012.11.029 (2013).
- 106 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**, 491-498, doi:10.1038/ng.806 (2011).
- 107 McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069-2070, doi:10.1093/bioinformatics/btq330 (2010).
- 108 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).
- 109 Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* **4**, 1073-1081, doi:10.1038/nprot.2009.86 (2009).
- 110 Fischer, U. *et al.* Genomics and drug profiling of fatal TCF3-HLF-positive acute lymphoblastic leukemia identifies recurrent mutation patterns and therapeutic options. *Nature genetics* **47**, 1020-1029, doi:10.1038/ng.3362 (2015).
- 111 Vervoort, S. J., van Boxtel, R. & Coffey, P. J. The role of SRY-related HMG box transcription factor 4 (SOX4) in tumorigenesis and metastasis: friend or foe? *Oncogene* **32**, 3397-3409, doi:10.1038/onc.2012.506 (2013).
- 112 Katoh, M. Function and cancer genomics of FAT family genes (review). *International journal of oncology* **41**, 1913-1918, doi:10.3892/ijo.2012.1669 (2012).
- 113 Zhu, J. W. *et al.* E2F1 and E2F2 determine thresholds for antigen-induced T-cell proliferation and suppress tumorigenesis. *Molecular and cellular biology* **21**, 8547-8564, doi:10.1128/MCB.21.24.8547-8564.2001 (2001).
- 114 Rao, Z. *et al.* The structure of a Ca(2+)-binding epidermal growth factor-like domain: its role in protein-protein interactions. *Cell* **82**, 131-141 (1995).
- 115 Nishi, M. *et al.* Suppression of the let-7b microRNA pathway by DNA hypermethylation in infant acute lymphoblastic leukemia with MLL gene rearrangements. *Leukemia* **27**, 389-397, doi:10.1038/leu.2012.242 (2013).
- 116 Wu, Z. *et al.* HMGA2 as a potential molecular target in KMT2A-AFF1-positive infant acute lymphoblastic leukaemia. *Br J Haematol* **171**, 818-829, doi:10.1111/bjh.13763 (2015).
- 117 Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20, doi:10.1016/j.cell.2004.12.035 (2005).
- 118 Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes & development* **18**, 504-511, doi:10.1101/gad.1184404 (2004).
- 119 Rhoades, M. W. *et al.* Prediction of plant microRNA targets. *Cell* **110**, 513-520 (2002).
- 120 Enright, A. J. *et al.* MicroRNA targets in Drosophila. *Genome biology* **5**, R1, doi:10.1186/gb-2003-5-1-r1 (2003).
- 121 John, B. *et al.* Human MicroRNA targets. *PLoS biology* **2**, e363, doi:10.1371/journal.pbio.0020363 (2004).
- 122 Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome research* **19**, 92-105, doi:10.1101/gr.082701.108 (2009).
- 123 Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787-798 (2003).
- 124 Alexiou, P., Maragkakis, M., Papadopoulos, G. L., Reczko, M. & Hatzigeorgiou, A. G. Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics* **25**, 3049-3055, doi:10.

- 1093/bioinformatics/btp565 (2009).
- 125 Pinzon, N. *et al.* microRNA target prediction programs predict many false positives. *Genome research* **27**, 234-245, doi:10.1101/gr.205146.116 (2017).
- 126 Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146**, 353-358, doi:10.1016/j.cell.2011.07.014 (2011).
- 127 Phatak, P. & Donahue, J. M. Biotinylated Micro-RNA Pull Down Assay for Identifying miRNA Targets. *Bio-protocol* **7**, e2253, doi:10.21769/BioProtoc.2253 (2017).
- 128 Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855-862 (1993).
- 129 Olsen, P. H. & Ambros, V. The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Developmental biology* **216**, 671-680, doi:10.1006/dbio.1999.9523 (1999).
- 130 Bibi, F. *et al.* microRNA analysis of gastric cancer patients from Saudi Arabian population. *BMC genomics* **17**, 751-751, doi:10.1186/s12864-016-3090-7 (2016).
- 131 Zhong, S. *et al.* MicroRNA expression profiles of drug-resistance breast cancer cells and their exosomes. *Oncotarget* **7**, 19601-19609, doi:10.18632/oncotarget.7481 (2016).
- 132 Glaser, K. B. *et al.* Gene expression profiling of multiple histone deacetylase (HDAC) inhibitors: defining a common gene set produced by HDAC inhibition in T24 and MDA carcinoma cell lines. *Molecular cancer therapeutics* **2**, 151-163 (2003).
- 133 Mariadason, J. M., Corner, G. A. & Augenlicht, L. H. Genetic reprogramming in pathways of colonic cell maturation induced by short chain fatty acids: comparison with trichostatin A, sulindac, and curcumin and implications for chemoprevention of colon cancer. *Cancer research* **60**, 4561-4572 (2000).
- 134 Van Lint, C., Emiliani, S. & Verdin, E. The expression of a small fraction of cellular genes is changed in response to histone hyperacetylation. *Gene expression* **5**, 245-253 (1996).
- 135 Lee, H., Lee, S., Bæk, M., Kim, H.-Y. & Jeoung, D.-I. Expression profile analysis of trichostatin A in human gastric cancer cells. *Biotechnology Letters* **24**, 377-381, doi:10.1023/A:1014512819978 (2002).
- 136 de Ruijter, A. J., van Gennip, A. H., Caron, H. N., Kemp, S. & van Kuilenburg, A. B. Histone deacetylases (HDACs): characterization of the classical HDAC family. *The Biochemical journal* **370**, 737-749, doi:10.1042/BJ20021321 (2003).
- 137 Grozinger, C. M., Hassig, C. A. & Schreiber, S. L. Three proteins define a class of human histone deacetylases related to yeast Hda1p. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 4868-4873 (1999).
- 138 Kamal, A. *et al.* A high-affinity conformation of Hsp90 confers tumour selectivity on Hsp90 inhibitors. *Nature* **425**, 407-410, doi:10.1038/nature01913 (2003).
- 139 Rietveld, L. E., Caldenhoven, E. & Stunnenberg, H. G. Avian erythroleukemia: a model for corepressor function in cancer. *Oncogene* **20**, 3100-3109, doi:10.1038/sj.onc.1204335 (2001).
- 140 Minucci, S., Nervi, C., Lo Coco, F. & Pelicci, P. G. Histone deacetylases: a common molecular target for differentiation treatment of acute myeloid leukemias? *Oncogene* **20**, 3110-3115, doi:10.1038/sj.onc.1204336 (2001).
- 141 Saito, A. *et al.* A synthetic inhibitor of histone deacetylase, MS-27-275, with marked in vivo antitumor activity against human tumors. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 4592-4597 (1999).

- 142 Richon, V. M., Zhou, X., Rifkind, R. A. & Marks, P. A. Histone deacetylase inhibitors: development of suberoylanilide hydroxamic acid (SAHA) for the treatment of cancers. *Blood cells, molecules & diseases* **27**, 260-264, doi:10.1006/bcmd.2000.0376 (2001).
- 143 He, L. Z. *et al.* Histone deacetylase inhibitors induce remission in transgenic models of therapy-resistant acute promyelocytic leukemia. *The Journal of clinical investigation* **108**, 1321-1330, doi:10.1172/JCI11537 (2001).
- 144 Marks, P. A., Richon, V. M. & Rifkind, R. A. Histone deacetylase inhibitors: inducers of differentiation or apoptosis of transformed cells. *Journal of the National Cancer Institute* **92**, 1210-1216 (2000).
- 145 Park, M. J. *et al.* FBXW7 and NOTCH1 mutations in childhood T cell acute lymphoblastic leukaemia and T cell non-Hodgkin lymphoma. *Br J Haematol* **145**, 198-206, doi:10.1111/j.1365-2141.2009.07607.x (2009).
- 146 Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758-764, doi:10.1038/nature05690 (2007).
- 147 Gutierrez, A. *et al.* High frequency of PTEN, PI3K, and AKT abnormalities in T-cell acute lymphoblastic leukemia. *Blood* **114**, 647-650, doi:10.1182/blood-2009-02-206722 (2009).
- 148 Zuurbier, L. *et al.* The significance of PTEN and AKT aberrations in pediatric T-cell acute lymphoblastic leukemia. *Haematologica* **97**, 1405-1413, doi:10.3324/haematol.2011.059030 (2012).
- 149 Silva, A. *et al.* PTEN posttranslational inactivation and hyperactivation of the PI3K/Akt pathway sustain primary T cell leukemia viability. *The Journal of clinical investigation* **118**, 3762-3774, doi:10.1172/JCI34616 (2008).
- 150 Yoneda-Kato, N., Tomoda, K., Umehara, M., Arata, Y. & Kato, J. Y. Myeloid leukemia factor 1 regulates p53 by suppressing COP1 via COP9 signalosome subunit 3. *The EMBO journal* **24**, 1739-1749, doi:10.1038/sj.emboj.7600656 (2005).
- 151 Mahla, R. S., Reddy, M. C., Prasad, D. V. & Kumar, H. Sweeten PAMPs: Role of Sugar Complexed PAMPs in Innate Immunity and Vaccine Biology. *Frontiers in immunology* **4**, 248, doi:10.3389/fimmu.2013.00248 (2013).
- 152 Lin, H. *et al.* Loss of immunity-supported senescence enhances susceptibility to hepatocellular carcinogenesis and progression in Toll-like receptor 2-deficient mice. *Hepatology* **57**, 171-182, doi:10.1002/hep.25991 (2013).
- 153 Soares, J. B. *et al.* Increased hepatic expression of TLR2 and TLR4 in the hepatic inflammation-fibrosis-carcinoma sequence. *Innate immunity* **18**, 700-708, doi:10.1177/1753425912436762 (2012).
- 154 Chen, R., Alvero, A. B., Silasi, D. A., Steffensen, K. D. & Mor, G. Cancers take their Toll--the function and regulation of Toll-like receptors in cancer cells. *Oncogene* **27**, 225-233, doi:10.1038/sj.onc.1210907 (2008).
- 155 Ashman, L. K. The biology of stem cell factor and its receptor C-kit. *The international journal of biochemistry & cell biology* **31**, 1037-1051 (1999).
- 156 Heinrich, M. C. *et al.* Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J Clin Oncol* **21**, 4342-4349, doi:10.1200/JCO.2003.04.190 (2003).
- 157 McLean, S. R. *et al.* Imatinib binding and cKIT inhibition is abrogated by the cKIT kinase domain I missense mutation Val654Ala. *Molecular cancer therapeutics* **4**, 2008-2015, doi:10.1158/1535-7163.MCT-05-0070 (2005).
- 158 Roberts, K. G. *et al.* Resistance to c-KIT kinase inhibitors conferred by V654A mutation. *Molecular cancer therapeutics* **6**, 1159-1166, doi:10.1158/1535-7163.MCT-06-0641 (2007).



- 159 Liu, M., Wang, G., Gomez-Fernandez, C. R. & Guo, S. GREB1 functions as a growth promoter and is modulated by IL6/STAT3 in breast cancer. *PLoS one* **7**, e46410, doi:10.1371/journal.pone.0046410 (2012).
- 160 Mohammed, H. *et al.* Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell reports* **3**, 342-349, doi:10.1016/j.celrep.2013.01.010 (2013).
- 161 Bauerschlag, D. O. *et al.* Progression-free survival in ovarian cancer is reflected in epigenetic DNA methylation profiles. *Oncology* **80**, 12-20, doi:10.1159/000327746 (2011).
- 162 Antunes, A. A. *et al.* GREB1 tissue expression is associated with organ-confined prostate cancer. *Urologic oncology* **30**, 16-20, doi:10.1016/j.urolonc.2009.09.014 (2012).
- 163 Rae, J. M. *et al.* GREB 1 is a critical regulator of hormone dependent breast cancer growth. *Breast cancer research and treatment* **92**, 141-149, doi:10.1007/s10549-005-1483-4 (2005).
- 164 Laviolette, L. A., Hodgkinson, K. M., Minhas, N., Perez-Iratxeta, C. & Vanderhyden, B. C. 17beta-estradiol upregulates GREB1 and accelerates ovarian tumor progression in vivo. *International journal of cancer* **135**, 1072-1084, doi:10.1002/ijc.28741 (2014).
- 165 Stoskus, M. *et al.* Identification of characteristic IGF2BP expression patterns in distinct B-ALL entities. *Blood cells, molecules & diseases* **46**, 321-326, doi:10.1016/j.bcmd.2011.02.005 (2011).
- 166 Gu, G., Sederberg, M. C., Drachenberg, M. R. & South, S. T. IGF2BP1: a novel IGH translocation partner in B acute lymphoblastic leukemia. *Cancer genetics* **207**, 332-334, doi:10.1016/j.cancergen.2014.07.002 (2014).
- 167 Jeffries, S. J., Jones, L., Harrison, C. J. & Russell, L. J. IGH@ translocations co-exist with other primary rearrangements in B-cell precursor acute lymphoblastic leukemia. *Haematologica* **99**, 1334-1342, doi:10.3324/haematol.2014.103820 (2014).
- 168 Yamagishi, M. *et al.* Polycomb-mediated loss of miR-31 activates NIK-dependent NF-kappaB pathway in adult T cell leukemia and other cancers. *Cancer cell* **21**, 121-135, doi:10.1016/j.ccr.2011.12.015 (2012).
- 169 Lin, Y. *et al.* Engineering of functional replication protein homologs based on insights into the evolution of oligonucleotide/oligosaccharide-binding folds. *Journal of bacteriology* **190**, 5766-5780, doi:10.1128/JB.01930-07 (2008).
- 170 Poncet, D. *et al.* Changes in the expression of telomere maintenance genes suggest global telomere dysfunction in B-chronic lymphocytic leukemia. *Blood* **111**, 2388-2391, doi:10.1182/blood-2007-09-111245 (2008).
- 171 Hoxha, M. *et al.* Relevance of telomere/telomerase system impairment in early stage chronic lymphocytic leukemia. *Genes, chromosomes & cancer* **53**, 612-621, doi:10.1002/gcc.22171 (2014).

## 7. List of Abbreviations

AD	activation domain
ALL	acute lymphatic leukemia
AML	acute myeloid leukemia
Amp	gene for ampicillin resistance
APS	Ammoniumpersulfat
BFM	Berlin-Frankfurt-Münster study group
BM	bone marrow
BMFZ	Biologisch-Medizinisches Forschungszentrum
bp	base pair
bZIP	basic leucine zipper
CD86	cluster of differentiation 86
CCG	Childrens' Cancer Group
CCL5	C-C motif chemokine ligand 5
CCR	continuous complete remission
CCR7	C-C motif chemokine receptor 7
CDP	common dendritic progenitor
cIgM	cytoplasmic immunoglobulin M
CLL	chronic lymphoblastic leukemia
CLP	common lymphoid progenitor
C $\mu$	$\mu$ heavy chain
CML	chronic myeloid leukemia

CMP	common myeloid progenitor
c-Myc	cellular myelocytomatosis viral oncogene homolog
CNS	central nervous system
COL1A2	Collagen Type I Alpha 2 Chain
CPM	counts per million
CSF	cerebrospinal fluid
CTEP	Cancer Therapy Evaluation Program
dec	deceased
DFCI	Dana-Farber Cancer Institute
DPB	D-box binding PAR BZIP transcription factor
DPP6	Dipeptidyl Peptidase Like 6
DTT	1,4.dithio-DL-threitol
E. coli	Escherichia coli
E-FDR	Enhanced False Discovery Rate
EFS	event free survival
ELP	early lymphoid progenitor
EmGFP	Emerald green fluorescent protein
ERK5	ERK5-mitogen-activated protein-kinase 7
EV	empty vector
FAB	French-American-British
FBS	Fetal Bovine Serum
FDR	false discovery rate
FOSL1, FRA-1	FOS like antigen 1

FPKM	Fragments per kilobase of exon per million fragments mapped
FRAS1	Fraser Extracellular Matrix Complex Subunit 1
Gfold	generalized Fold Change
GH	growth hormone
GMP	granulocyte-macrophage progenitor
GREB1	growth regulation by estrogen in breast cancer 1
HCV	hepatitis C virus
HD	homeodomain
HLF	hepatic leukemic factor
Hluc	firefly luciferase gene
hRluc	Renilla luciferase reporter gene
HSC	hematopoietic stem cell
HSCT	hematopoietic stem cell transplantation
HSV	Herpes Simplex Virus
iBFM	international Berlin-Frankfurt-Münster
IGF2BP1	Insulin Like Growth Factor 2 mRNA Binding Protein 1
IL1B	interleukin 1beta
IL6	interleukin 6
indel	insertion/deletion
IPA	Ingenuity® Pathway Analysis
IT	intermediate-term
iT-ALL	iT-ALL
ITGA2B	integrin subunit alpha 2b

KMT2Ar	KMT2A rearrangements
lacZ $\alpha$	gene for $\beta$ -galactosidase
LCK	Lymphocyte Cell-Specific Protein-Tyrosine Kinase
logFC	logarithmic fold change
LT	long-term
MCS	multiple cloning site
MDP	monocyte-dendritic cell progenitor
MEP	megakaryocyte-erythrocyte progenitor
mRNA	messenger RNA
miRNA, miR	microRNA
MLPP	lymphoid-primed MPP
MPP	multipotent progenitor
MSI2	Musashi RNA Binding Protein 2
NGF	neural growth factor
NGS	Next Generation Sequencing
NK	natural killer cell
ori	origin of replication
PAR	proline and acidic amino-acid-rich region
PB	peripheral blood
PBS	Dulbecco's Phosphate Buffered Saline
PBX1	Pre-B cell leukemic homeobox 1
PLB	Passive Lysis Buffer
POG	Pediatric Oncology Group

Pre-miRNA	precursor miRNA
pri-miRNA	primary-miRNA
prl	pre B-cell leukemia
PRS	PBX1 recognition sequence
p-val	p-value
rem	remission
RIN	RNA integrity number
RISC	RNA induced silencing complex
RPA1	replication protein A1
SCF	stem cell factor
SDS	Sodiumdodecylsulfate
SELP	selectin P
sIg	surface immunoglobulin
SJCRH	St Jude Children's Research Hospital
SNV	single nucleotide variant
SOX4	Sex Determining Region Y-related Box4
ST	short-term
SV40	Simian-Virus 40
TCR-B	T-cell receptor beta
TdT	terminal deoxynucleotidyl transferase
TEF	thyrotrophic embryonic factor
TK	Thymidine Kinase
TKI	tyrosine kinase inhibitors

TRKA	tyrosine kinase A
WBC	white blood cell count
WHO	World Health Organization
WT	wild type

## **8. Acknowledgements**

First of all, I would like to thank Prof. Arndt Borkhardt for giving me the opportunity to perform my PhD thesis in the KMT lab for these interesting projects. Furthermore, I would like to thank Prof. Dr. Holger Gohlke evaluating my thesis as a second referee.

I want to thank Jessica Höll and Ute Fischer for supporting and guidance during the PhD thesis, for all the ideas, which pushing on the projects, for the motivation and for proofreading this thesis.

I also want to thank all present and past members of the KMT lab for the great working atmosphere. Especially I would like to thank Sanil Bhatia for collaboration and Daniel Hein for proofreading and motivation. I also would like to thank all the other colleagues, who are involved in the project of my PhD thesis: Andreas Klötgen, Daniel Picard, Kebria Hezaveh, Jens Warfsmann, Sujitha Duggimpudi and Melf Sönnichsen.

Last but not least, I would like to thank my whole family, especially my wife Katharina, for the great support.



## Curriculum Vitae

### Personal Data

Name: Mareike Dörrenberg  
Adress: Heerdter Landstr. 244  
Date and Place of Birth: February 21<sup>st</sup>, in Recklinghausen, Germany  
Nationality: German

### Research and Education

01/2014 – 03/2018      PhD Thesis in the Department of Pediatric Oncology,  
Hematology and Clinical Immunology, Heinrich-Heine  
University Düsseldorf  
  
Title: "Genetic and Epigenetic Landscapes of Pediatric Acute  
Lymphoblastic Leukemia Subtypes"  
  
(Supervisor: Prof. Dr. Arndt Borkhardt)

09/2012 – 12/2013      Research assistant at Westfälische Hochschule, Recklinghausen  
  
Topic: Expression and Purification of the Recombinant Protein  
VPR of HIV1

09/2011 – 08/2012      Master Thesis at the Institute of Transfusion Medicine,  
University Hospital Essen  
  
Title: Functional Analysis of the Cell Fate Determinant Numb,  
Known from *Drosophila Melanogaster*, in Human Mesenchymal  
Stromal Cells  
  
(Supervisor: Dr. Bernd Giebel)

10/2010 – 08/2012      Westfälische Hochschule, Recklinghausen: Master of Science

04/2010 – 08/2010      Bachelor Thesis at TaconicArtemis, Cologne and Westfälische  
Hochschule, Recklinghausen  
  
Title: Establishment of an *in-vitro* System for the Differentiation  
of Murine Embryonic Stem Cells to Cardiomyocytes and  
Neurons  
  
(Supervisor: Prof. Dr. Schwenk)

10/2007 – 08/2010      Westfälische Hochschule, Recklinghausen, Bachelor of Science

09/2005 – 07/2007      Hans-Böckler-Berufskolleg, Marl: Biological Technical  
Assistant

## Publications

1. Doerrenberg, Mareike, et al. "T-cell acute lymphoblastic leukemia in infants has distinct genetic and epigenetic features compared to childhood cases." *Genes, Chromosomes and Cancer* 56.2 (2017): 159-167.
2. Shinde, Prashant V., et al. "Tumor necrosis factor-mediated survival of CD169+ cells promotes immune activation during vesicular stomatitis virus infection." *Journal of virology* 92.3 (2018): e01637-17.

## Affirmation

Hereby, I declare on oath that I composed this dissertation independently by myself. I used only the references and resources indicated in this thesis. With the exception of such quotations, the work presented in this thesis is my own. I have accredited all the sources of help. This PhD thesis was never submitted or presented in a similar form to any other institution or examination board. I have not undertaken a doctoral examination without success so far.

Düsseldorf, September 11th, 2019

---

Mareike Dörrenberg