

**Characterization of proteomic signature of pancreatic  
and testicular carcinomas by new sensitive methods  
for scarce FFPE tissue proteomics**

Inaugural dissertation

For the attainment of the doctor of chemistry  
in the faculty of medicine at the  
Heinrich Heine University Düsseldorf

Presented by

**Stella Maris Pauls**

From Düsseldorf

Düsseldorf, August 2024

Carried out in the Molecular Proteomics Laboratory (MPL)  
of the Uniklinikum Düsseldorf and  
Heinrich-Heine University Düsseldorf

In collaboration with the  
Institute of pathology and the  
Institute of UrOncology  
of the Uniklinikum Düsseldorf

Printed with permission of the  
Faculty of Mathematics and Natural Science  
of the Heinrich-Heine University Düsseldorf

Correspondence

1. Prof Kai Stühler
2. Prof Andreas Reichert

Date of oral examination:

20. January 2025

*"I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale."*

**Marie Skłodowska Curie**



# Contents

1	Preamble.....	13
2	Summary.....	17
3	Zusammenfassung.....	19
4	Introduction .....	21
4.1	FFPE tissue preservation technique .....	21
4.2	Pancreatic ductal adenocarcinoma (PDAC).....	21
4.2.1	Morphology of pancreatic precursor lesions .....	21
4.2.2	Genetic alteration involved in precursor lesion progression for PDAC .....	22
4.2.3	Mucin expression and glyco-modification in pancreatic precursor lesions .....	23
4.2.4	Similarities to other adenocarcinoma types of pancreatobiliary system .....	25
4.3	Testicular carcinoma – Germ-cell tumors (GCT).....	26
4.3.1	Characterization of GCT .....	26
4.3.2	Yolk-sac tumors (YST) and therapy resistance .....	27
4.3.3	Germ cell tumor-related somatic-type malignancies (STM) .....	27
4.3.4	Growing teratoma syndrome (GTS) .....	28
5	Manuscript 1:.....	29
	Uncover the archived treasures – Challenges and opportunities in FFPE tissue proteomics .....	29
6	Manuscript 2:.....	51
	Proteome analysis of microdissected pancreas tissue by improved HIAR protocol.....	51
7	Manuscript 3:.....	75
	Proteome analysis with combined DIA and Glyco DIA approach of laser microdissected precursor lesions from pancreatobiliary cancer to improve early cancer diagnostics ...	75
8	Manuscript 4:.....	110
	Targeting CLDN6 in germ cell tumors by an antibody-drug-conjugate and studying therapy resistance of yolk-sac tumors to identify and screen specific therapeutic options .....	110
9	Manuscript 5:.....	129
	Characterizing the mutational burden, DNA methylation landscape, and proteome of germ cell tumor-related somatic-type malignancies to identify the tissue-of-origin, mechanisms of therapy resistance, and druggable targets .....	129
10	Manuscript 6:.....	143
	Assessing the risk to develop a growing teratoma syndrome based on molecular and epigenetic subtyping as well as novel secreted biomarkers .....	143
11	Concluding Remarks .....	159
12	Danksagung .....	161
13	References .....	163



## Affidavit

I declare under oath that I produced all data and wrote my thesis independently and without any assistance by a third party under consideration of the “Principles for the Safeguarding of Good Scientific Practice” at Heinrich Heine University Düsseldorf. I declare that my thesis has not been submitted to another faculty and I have not attempted to earn a doctoral degree neither successfully nor unsuccessfully.

Düsseldorf, August 2024

A handwritten signature in black ink, consisting of a large, stylized 'S' and 'P' intertwined.

---

Stella Maris Pauls



## Abbreviations

FFPE	Formalin-fixed paraffin embedded
LC-MS/MS	Liquid-chromatography tandem mass spectrometry
HIAR	Heat-induced antigen retrieval
DIA	Data independent acquisition
DDA	Data dependent acquisition
MS	Mass spectrometry
CLDN6	Claudin-6 Gene
GCT	Germ-cell tumors
YST	Yolk-sac tumors
ADC	Antibody-drug-conjugate
FGF	Fibroblast growth factor
VGf	VGf nerve growth factor inducible
PDGF	Platelet-derived growth factor
mTOR	Mammalian target of rapamycin
CHEK1	Checkpoint kinase 1
AURKA	Aurora kinase A
PARP	Poly (ADP-ribose) polymerase
TSO	TruSight Oncology
DNA	Deoxyribonucleic acid
STM	Somatic-type malignancies
MAPK	Mitogen-activated protein kinase
WNT	Wingless-related integration site
GTS	Growing teratoma syndrome
RNA	Ribonucleic acid
PDAC	Pancreatic ductal adenocarcinoma
CCA	Cholangiocarcinoma
ER	Endoplasmic reticulum
T antigen	Thomsen–Friedenreich-Antigen
Tn antigen	Thomsen-nouvelle antigen
LCM	Laser capture microdissection
GPI	Glycosylphosphatidylinositol
PanIN	Pancreatic intraepithelial neoplasms
IPMN	Intraductal papillary mucinous neoplasms
MCN	Mucinous cystic neoplasms
ADM	Acinar-ductal metaplasia
MUC	Mucins
K-ras	Kirsten rat sarcoma virus
GNAS	Guanine Nucleotide binding protein

RTK	Receptor tyrosine kinases
EGFR	Epidermal growth factor receptor
FGFR	Fibroblast growth factor receptor
GalNAc	<i>N</i> -Acetylgalactosamine
BIIN	Biliary intraepithelial neoplasms
IPNB	Intraductal papillary neoplasm of the bile duct
GCNIS	Germ cell neoplasia in situ
MMAE	Monomethyl Auristatin E
IHC	Immunohistochemistry
PTM	Post-translational modification
NCI	National cancer institute
EU	European Union
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
IMS	Ion Mobility Spectrometry
HCl	Hydrochloric acid
TFA	Trifluoroacetic acid
TFE	Trifluoroethanol
SDS	Sodium dodecyl-sulfate
GRAVY	Grand average of hydropathicity
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
pI	Isoelectric point
IEF	Isoelectric focusing
FASP	Filter-aided sample preparation
TCEP	Tris(2-carboxyethyl)phosphine
DTT	Dithiothreitol
PEG	Polyethylene glycol
PPG	Polypropylene glycol
SP3	Single-pot, solid-phase-enhanced, sample preparation
TMT	Tandem mass tag
Q-TOF	Quadrupole Time-of-Flight
SPEED	Sample preparation by easy extraction and digestion
BCA	Bicinchoninic acid
DIA-NN	Data independent acquisition neural network
ITRAQ	Isobaric tags for relative and absolute quantitation
TMA	Tissue microarray
PB	Pancreatobiliary
TD	Tumor developing
TND	Tumor non-developing
H&E stain	Hematoxylin and eosin stain
EtOH	Ethanol

ACN	Acetonitrile
CV	Compensation Voltage
AGC	Automatic gain control
HCD	High-energy collisional dissociation
ANOVA	Analysis of Variance
FPKM	Fragments Per Kilobase of exon per Million mapped reads
XIC	Extracted Ion chromatogram



# 1 Preamble

This thesis is divided to six manuscripts (three in preparation and three already published) and an introduction in English and German about the development of methods for proteomic analysis of scarce FFPE tissue specimens. The methods developed in this thesis are used for the analysis of 2.5 mm<sup>2</sup> FFPE tissue sections to unlock the biological pathways underlying lesion progression of pancreatobiliary cancer. Furthermore, the more sensitive methods combining an optimized sample preparation with data-independent acquisition method for LC-MS/MS analysis is used for the analysis of FFPE tissue section (macroscopic) with highly variable amounts of tissue for the analysis of testicular carcinomas.

## 1. Uncover the archived treasures – Challenges and opportunities in FFPE tissue proteomics

**Stella Pauls**, Anja Stefanski, Kai Stühler *-in preparation*

The first Manuscript is a review that introduces into the tissue preservation technique of formalin fixation and paraffin embedding (FFPE). It demonstrates the importance of preserving tissue samples taken from patient biopsies and stored into tissue banks with the opportunity to analyze large patient cohorts after a long collection period regarding their Genome, Transcriptome and Proteome. It is used as an introduction into the topic of FFPE tissue preservation and summarizes the current state-of-the-art sample preparation methods for bottom-up proteomics, including tissue lysis, protein clean-up and digestion methods. For the development of more sensitive sample preparation methods for the use of scarce FFPE tissue samples, the current state of the art analytical methods are used as starting point.

## 2. Proteome analysis of microdissected pancreas tissue by improved HIAR protocol

**Stella Pauls**, Anja Stefanski, Friederike Opitz, Sandra Biskup, Irene Esposito, Kai Stühler *-in preparation*

The second Manuscript describes the method optimization of the heat-induced antigen retrieval (HIAR) protocol and the subsequent protein clean-up and digestion method via SP3 protocol to produce high quality peptide mixtures coming from scarce FFPE tissues for proteomic analysis via LC-MS/MS. The development of the analysis platform is extensively described in this manuscript. It also points out the importance of using data-independent acquisition (DIA) methods over data-dependent acquisition (DDA) for LC-MS/MS analysis to increase the depth of proteome analysis, especially for scarce FFPE tissue samples. These optimized methods are used for the analysis of several different FFPE tissue samples in the field of pancreatic and testicular cancer research. In the following manuscripts (manuscript 1 – 3) this optimized sample preparation platform is used and enables the efficient proteomic analysis of laser-microdissected FFPE tissue

samples and FFPE samples in strongly differing sizes and numbers, like tissue microarrays and different numbers of pooled FFPE tissue samples.

### **3. Proteome analysis with combined DIA and Glyco DIA approach of laser microdissected precursor lesions from pancreatobiliary cancer to improve early cancer diagnostics**

**Stella Pauls**, Anja Stefanski, Friederike Opitz, Sandra Biskup, Irene Esposito, Kai Stühler  
*-in preparation*

The sixth manuscript is a study with over 170 samples of precisely laser capture microdissected PDAC/CCA precursor lesions in sizes of 2.5 mm<sup>2</sup>. These samples are analyzed using the optimized HIAR and sample preparation protocol and DIA as LC-MS/MS method. Due to the high sensitivity of the optimized methods a combined analysis regarding the Proteome and the Glyco-Proteome is feasible. In the proteome dataset over 5100 proteins are identified regarding the whole human proteome and around 400 O-glycosylated proteins are identified using a spectral library from Ye, Mao, Clausen and Vakhrushev <sup>10</sup> This gives insights into biological processes involved in lesion progression, including protein translation, protein (de)neddylation, N-glycosylation and vesicle-mediated protein targeting from/to ER and/or Golgi. Glyco-proteome data revealed several different findings, such as an overall accumulation of T/Tn antigens and their sialylated subforms on the surface of several different proteins in the tumor developing group of precursor lesions, with a special focus on the modification candidate 14-3-3 theta.

### **4. Targeting CLDN6 in germ cell tumors by an antibody-drug-conjugate and studying therapy resistance of yolk-sac tumors to identify and screen specific therapeutic options**

Skowron, M.A., Kotthoff, M., Bremmer, F. *et al.* Targeting CLDN6 in germ cell tumors by an antibody-drug-conjugate and studying therapy resistance of yolk-sac tumors to identify and screen specific therapeutic options. *Mol Med* 29, 40 (2023).

The third Manuscript is a published work in collaboration with the Nettersheim group of translational UroOncology. The publication is targeting CLDN6 as a drug target in germ cell tumors (GCT) and studies therapy resistance of recurring yolk-sac tumors (YST) to identify new specific therapeutic options for therapy resistant GCT sub-forms. The first optimized conditions concerning the HIAR protocol in combination with DDA as MS acquisition method are used to analyze FFPE tissue samples in differing sizes of 1 – 10 cm<sup>2</sup> from different types of GCT to provide proteomic analysis results, supporting the understanding of therapy resistance in YST. Furthermore, a new CLDN6 antibody-drug-conjugate (ADC) is screened. It is demonstrated that CLDN6-ADC treatment enhances apoptosis induction in GCT cells, compared to non-cancerous control cells. Proteomic profiling reveals new therapeutic targets for YST, including several markers involved in different signaling pathways. Factors involved in MAPK signaling, RNA translation initiation and binding, oxidative stress, extracellular matrix-related processes and

immune response are involved in modulating therapy resistance in YST. Finally, the study presents novel pharmacological inhibitors blocking the FGF, VGF, PDGF, mTOR, CHEK1, AURKA, or PARP signaling for the treatment of therapy-resistant YST.

## **5. Characterizing the mutational burden, DNA methylation landscape, and proteome of germ cell tumor-related somatic-type malignancies to identify the tissue-of-origin, mechanisms of therapy resistance, and druggable targets**

Bremmer, F., Pongratanakul, P., Skowron, M. *et al.* Characterizing the mutational burden, DNA methylation landscape, and proteome of germ cell tumor-related somatic-type malignancies to identify the tissue-of-origin, mechanisms of therapy resistance, and druggable targets. *Br J Cancer* **129**, 1580–1589 (2023).

The fourth Manuscript is another published work in collaboration with the Nettersheim group of translational UroOncology, using the first optimized conditions of the HIAR protocol to analyze pooled FFPE tissue microarrays in differing sizes and numbers from 2.5 – 10 mm<sup>2</sup>. These FFPE tissue microarrays are analyzed with different analytical methods, such as TSO assays, 850k DNA methylation arrays and mass spectrometry. The results show that carcinoma-related somatic-type malignancies (STM) resemble yolk-sac tumors (YST), while sarcoma-related STM are associated with teratoma. Mutations in FGF signaling factors, such as FGF6/23 and FGFR1/4 are highlighted as potential therapeutic targets. Other signaling pathways in STM, including AKT, FGF, MAPK, and WNT, are indicated as possible strategies to regulate oxidative stress, toxin transport, DNA helicase activity, apoptosis and cell cycle, which can explain the high therapy resistance of STM. Ultimately, putative novel biomarkers are EFEMP1, MIF and DNA methylation at specific CpG dinucleotides.

## **6. Assessing the risk to develop a growing teratoma syndrome based on molecular and epigenetic subtyping as well as novel secreted biomarkers**

Pongratanakul, P., Bremmer, F., Pauls, S., *et al.* Assessing the risk to develop a growing teratoma syndrome based on molecular and epigenetic subtyping as well as novel secreted biomarkers. *Cancer Letters*, Vol. 585, 216673, ISSN 0304-3835 (2024)

The fifth manuscript is another published work in collaboration with the Nettersheim group of translational UroOncology. In this study the final optimized HIAR protocol and sample preparation method is used for sample preparation, combined with DIA as LC-MS/MS acquisition method. The growing teratoma syndrome (GTS) is a phenomenon in which a reduced amount of tumor markers is observed during chemotherapy treatment. Due to a worldwide small number of available cases of GTS our sensitive proteomic analysis platform ensures a deep proteomic analysis with great reproducibility, even in FFPE samples with differing sizes and numbers. A clinical characterization of 50 patients with GTS are selected and 12 samples are molecular

analyzed. Based on different growth rates, GTS is subdivided into three groups, one slow (<0.5 cm/month), medium (0.5–1.5) and rapid (>1.5) growing group. These groups are analyzed concerning DNA methylation, microRNA expression and the proteome/secretome. Proteins enriched in GTS compared to teratoma samples unravel biological pathways concerning proliferation, DNA replication and the cell cycle, and proteins interacting with the immune system are depleted.

## 2 Summary

FFPE tissues are an important source for gaining new insights into various diseases mechanisms. Many helpful methods, such as laser microdissection, can aid in examining biological processes at the cellular level from the fixed tissue material. With new, sensitive sample preparation and data acquisition methods, it is possible to enable deep proteome analyses to advance the study of various proteins involved in disease mechanisms. This work primarily focuses on cancer research regarding pancreatobiliary and testicular cancers. The optimized sample preparation for LC-MS/MS based proteomics allows for the discovery of new drug targets, biomarker candidates, and give biological insights into disease development. In the field of testicular carcinomas, new insights into drug resistance of recurring yolk-sac tumors have been gained. These findings help develop new antibody-drug-conjugates to treat these resistant tumors and enable broad drug screenings. Furthermore, the biological processes involved in the development of therapy resistance were explored using multiomic applications (genomics, methylomics, proteomics), revealing the tissue of origin (teratomas) from which germ cell tumors develop. The robustness of the sample preparation methods also makes it possible to study FFPE tissues of very rare diseases using LC-MS/MS-based proteomics. The growing teratoma syndrome (GTS) is a rare subform of germ cell tumors, characterized by a reduced occurrence of tumor markers during chemotherapy treatment and continuing uncontrolled growth. Here, insights into the biological processes involved in the development of GTS are provided, achieved through multiomic approaches (genomics, methylomics, transcriptomics, proteomics, secretomics). Finally, the optimized FFPE sample preparation methods are applied to laser microdissected FFPE tissues of precursor lesions from pancreatic cancer, enabling a new multiomics approach. Using a data-independent acquisition (DIA) method, it was possible to uncover parts of both the entire proteome and the O-glycoproteome in the recorded mass spectrometric dataset. This analysis provided insights into the diverse biological processes underlying the progression of precursor lesions into carcinomas. Additionally, it is shown that the T/Tn antigen glyco-modification and its sialylated subforms are increasingly found on many different proteins in the tumor-developing tissue lesions, supporting the thesis that these modifications accumulate more in tumor tissues. The protein 14-3-3 theta was identified as a potential biomarker candidate in a differential modification analysis, with the Tn antigen and the sialylated form being more prevalent in tumor-developing tissues. The optimized sample preparation methods presented here are suitable for very small amounts of FFPE tissue, as well as for sample sets with highly variable tissue quantities, as they ensure high sensitivity and reproducibility.



### 3 Zusammenfassung

FFPE Gewebe stellen eine wichtige Quelle zur Gewinnung neuer Erkenntnisse über verschiedenen Erkrankungen dar. Viele hilfreiche Methoden, wie die Laser-Mikrodissektion können dabei helfen biologische Prozesse aus dem fixierten Gewebematerial auf Zellebene zu untersuchen. Mit neuen, sensitiven Probenvorbereitungs- und Datenerfassungs-Methoden ist es möglich tiefe Proteom-Analysen zu ermöglichen, um die Erforschung verschiedener Proteine, die am Krankheitsgeschehen beteiligt sind voran zu treiben. Im Fokus der Forschung steht im allgemeinen die Krebsforschung von pankreatobiliären und testikulären Karzinomen. Die optimierte Probenvorbereitung für die LC-MS/MS-basierte Proteom-Analyse ermöglicht es neue Wirkstofftargets, Biomarker-Kandidaten und biologische Erkenntnisse zur Krankheitsentwicklung zu erlangen. Im Bereich der testikulären Karzinome konnten dahingehend neue Erkenntnisse zur Wirkstoffresistenz einiger wiederkehrender Dottersack-tumore erlangt werden. Diese Erkenntnisse helfen neue Wirkstoff-Antikörper-Konjugate zur Behandlung dieser resistenten Tumore zu entwickeln und ein breites Wirkstoffscreening zu ermöglichen. Weiterhin konnten die biologischen Prozesse, die an der Entwicklung der Therapieresistenz beteiligt sind durch Multiomic-Anwendungen (Genomics, Methyloomics, Proteomics) erforscht werden und entschlüsseln das Ursprungsgewebe (Teratome), aus dem sich die Keimzelltumore entwickeln. Durch die Robustheit der Probenvorbereitungsmethoden ist es ebenso möglich FFPE Gewebe von Erkrankungen, die sehr selten auftreten, mittels LC-MS/MS-basierter Proteomik zu untersuchen. Das Growing Teratoma Syndrome (GTS) stellt eine seltene Subform der Keimzelltumore dar, charakterisiert durch ein verringertes Auftreten von Tumormarkern während der Behandlung mittels Chemotherapeutika und weiterhin unkontrolliertem Wachstum. Hier konnten Einblicke in die biologischen Prozesse gegeben werden, die an der Entstehung des GTS beteiligt sind, welche wiederum durch Multiomic-Ansätze (Genomics, Methyloomics, Transcriptomics, Proteomics, Secretomics) erreicht wurden. Final wurden die optimierten FFPE-Probenvorbereitungsmethoden an Laser-mikrodissektierten FFPE-Geweben der Vorläuferläsionen des Pankreas-Karzinoms eingesetzt und ermöglichen einen neuen Multiomics-Ansatz, in dem es durch eine data-independent acquisition (DIA) Methode möglich war, in den aufgenommenen Daten sowohl Teile des gesamten Proteoms, als auch des O-Glyco-Proteoms zu erschließen. Mit Hilfe dieser Analyse konnten Einblicke in die vielfältigen biologischen Prozesse gegeben werden, die zu Grunde liegen, wenn sich Vorläuferläsionen zu Karzinomen entwickeln. Außerdem konnte gezeigt werden, dass die T/Tn-antigen Glyco-Modifikation und ihre sialylierten Subformen vermehrt auf vielen verschiedenen Proteinen in den Tumor-entwickelnden Gewebeläsionen zu finden sind, welches die These unterstützt, dass diese Modifikationen vermehrt in Tumorgeweben akkumulieren. Das Protein 14-3-3 theta wurde mit einem signifikanten Unterschied in einer differentiellen Modifikationsanalyse als möglicher Biomarker-Kandidat gefunden, bei dem das Tn-antigen und die sialylierte Form im Tumor-entwickelnden Gewebe vermehrt vorliegt. Die hier präsentierten, optimierten Probenvorbereitungsmethoden eignen sich für sehr geringe FFPE-Gewebemengen, aber auch

für Probensets in denen die Gewebemenge stark variiert, da sie eine hohe Sensitivität und Reproduzierbarkeit gewährleistet.

## 4 Introduction

### 4.1 FFPE tissue preservation technique

Formalin fixation and paraffin embedding of tissue samples taken from biopsies is a long-known preservation technique introduced by Ferdinand Blum in 1893.<sup>11</sup> Until today it is the most common preservation technique to create tissue banks over huge time-scales to build access point to large patient cohorts for statistical valid study sizes. Over the years analytical techniques have emerged and can be used to analyze these large patient cohorts, such as with more sensitive isolation method like laser capture microdissection (LCM), combined with more sensitive analytical methods, such as Genomics, Transcriptomics and Proteomics and several new immunohistochemical staining methods. This provides the potential to analyze tissue material on scales down to cellular levels and enables a deep understanding of several different diseases, mostly with the focus on cancer research. The current state of the art methods for proteomic FFPE tissue analysis are described in more detail in the review manuscript 1: **Uncover the archived treasures – Challenges and opportunities in FFPE tissue proteomics.**

### 4.2 Pancreatic ductal adenocarcinoma (PDAC)

The 12<sup>th</sup> most common cancer worldwide is pancreatic cancer, for which in 2020 over 495,000 new cases were reported and it was reported as the 4<sup>th</sup> leading cause of cancer deaths at all ages and genders.<sup>12, 13</sup> Reasons for the high mortality of pancreatic cancer are the lack of sufficient treatment, which is resection of the tumor and the fact that over 50% of the patients are diagnosed, when the tumor already expanded to other organs (metastasis). At this stage the 5-year overall survival rate is under 5%.<sup>13</sup> Pancreatic cancer primarily develops in the exocrine cells located in the head of the pancreas, near the pancreatobiliary ducts, with about 95% of all cases being pancreatic ductal adenocarcinoma (PDAC).<sup>14</sup> Pancreatobiliary precursor lesions are possible starting points for the development of PDAC and cholangiocarcinoma (CCA), for which increasing morphological changes are reported at lesion progression.<sup>15-17</sup>

#### 4.2.1 Morphology of pancreatic precursor lesions

Histologically PDAC develops from three main lesion types, called pancreatic intraepithelial neoplasms (PanIN), intraductal papillary mucinous neoplasms (IPMN) and mucinous cystic neoplasms (MCN).<sup>18</sup> The most prominent PDAC precursor lesions and widely studied and characterized in mouse models are PanIN lesions, which can be subclassified into PanIN1A, PanIN1B and PanIN2-3 with distinct morphological and genetical characteristics.<sup>19-22</sup> PanIN arise from small pancreatic ducts and appear as papillary or flat non-invasive epithelial neoplasm, characterized by columnar to cuboidal cells on different states of cytologic and architectural atypia and varying amounts of mucin.<sup>23</sup> The process of acinar-ductal metaplasia (ADM) in the centroacinar-acinar compartment is the suggested mechanism for the development of the most mucinous precursor lesions for PDAC.<sup>22, 24, 25</sup> One morphological characteristic of PDAC precursor lesions is the size, while PanIN lesions are <0.5 mm in diameter, IPMN and MCN

lesions are more enlarged with sizes >1.0 mm in diameter. Due to various histological findings (intraluminal nodular growth, tissue atrophy, mucous lake formation, pyloric glandlike structure) IPMN can be divided into two subtypes, the intestinal-type IPMN, characterized by intraluminal nodular growth, tissue atrophy and/or mucous lake formation and the gastric-type IPMN, which is mainly characterized to form pyloric glandlike structure and low-grade PanIN-like complexes.<sup>26</sup> The morphological similarity of gastric-type IPMN to low-grade PanIN and the similar mucine immunophenotype (MUC5AC+/MUC2-) suggests similarities in the biological pathways involved in lesion progression.<sup>27, 28</sup>

#### 4.2.2 Genetic alteration involved in precursor lesion progression for PDAC

The main early genetic driver mutation for the formation of PDAC precursor lesion is the activation of the *K-ras* oncogene, initiated by point mutations in over 99% of low-grade PanIN lesions.<sup>21, 29-31</sup> In the progression model for pancreatic cancer, *K-ras* mutation is an early event but not necessarily the gatekeeper for pancreatic ductal neoplasia.<sup>32</sup> *K-ras* mutation occurs in ~50% of nonpapillary lesions and is mainly present in more advanced stages of PanIN but are also found in early stages for gastric type IPMN more frequently than for intestinal type IPMN.<sup>33, 34</sup> Mutation in the *GNAS* gene is a common genetic alteration for intestinal-type IPMN and is associated with suppression of Notch signaling, antagonizing the *K-ras* pathway and attenuates tumor aggressiveness.<sup>35, 36</sup> *GNAS* mutation is more frequently found in intestinal-type IPMN, but also appears in gastric-type IPMN and PanIN precursor lesion and is an early genetic variation.<sup>31</sup> A frequent genetic alteration for many adenocarcinomas is the overexpression of *HER-2/neu*, an epidermal growth factor found in about 70% of infiltrating ductal adenocarcinomas and associated with mucin hypersecretion.<sup>37, 38</sup> Cell-cycle regulation is initiated by *p16* tumor suppressor gene (also known as *CDKN2A* or *CDK41*), which is found to be mutated in 71% of papillary duct lesions with significant atypia.<sup>39, 40</sup> In the progression model *p16* inactivation occurs slightly later than *K-ras* activation and drives atypical tissue progression and promotes invasiveness.<sup>32</sup> The transition of low-grade to high-grade precursor lesions is associated with inactivation of *DPC4* Gene (also known as *SMAD4*), which contributes malignant progression and is a late stage genetic alteration.<sup>40, 41</sup> The mutation of the tumor suppressor gene *p53* (*TP53*) is one of the most frequent genetic variations across all cancer types and is found to be mutated in 70% of pancreatic cancers.<sup>40, 42</sup> One germline mutation, which causes predisposition of women to breast and ovarian cancer is located on the breast cancer associated gene 1 and 2 (*BRCA1*, *BRCA2*), which is also found to be mutated in other cancer types, such as pancreatobiliary cancers.<sup>40, 43-45</sup> Mutations in *p53*, *SMAD4* and *BRCA1/2* are late stage genetic alteration involved in cancer progression. A schematic overview of the relevant genetic alteration involved in pancreatobiliary cancer and its occurrence in early or late-stage dysplasia is illustrated in Figure 1.

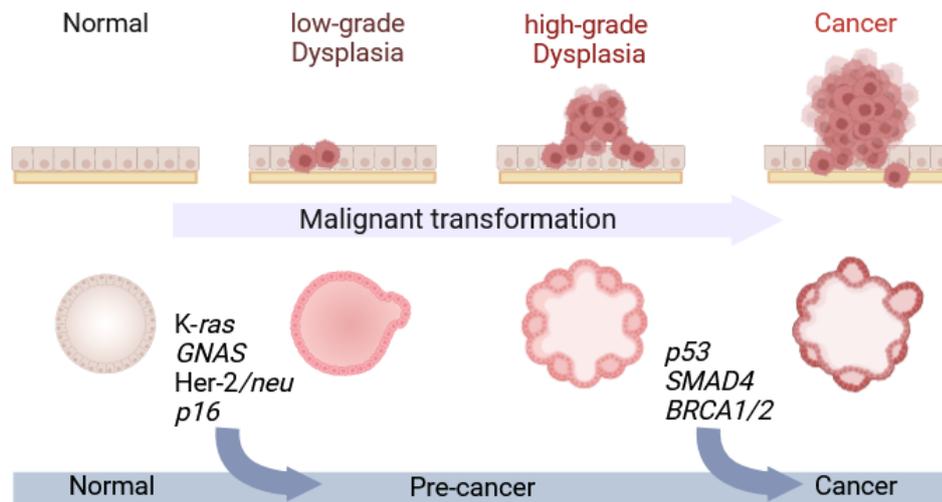


Figure 1: Schematic Overview of lesion progression with early and late driver mutations involved in pancreaticobiliary cancer formation. (Created with BioRender.com)

#### 4.2.3 Mucin expression and glyco-modification in pancreatic precursor lesions

Mucins are predominantly extracellular or membrane-bound high molecular weight proteins, carrying different O- and N-linked glycan structures, maintaining mucosal barriers against pathogens and an aggressive environment due to acids and/or other chemicals. For pancreatic precursor progression, aberrant mucin expression of several different types of mucins is involved in tissue transformation for all precursor lesions suggesting an influence in morphological changes, i.e., for mucin lake formation as well as in altered reactions of the immune system to the damaged tissue. Expression of mucins is a specificity of gastrointestinal tissue, in which several mucins are either secreted as gel-forming mucins (MUC2, MUC5B, MUC5AC, MUC6) or membrane-bound mucins (i.e., MUC1, MUC3, MUC4, MUC16) and altered localization and/or glycosylation are involved in several biological processes associated with mediation of immune evasion, oncogenic signaling, angiogenesis and metastasis.<sup>28, 46-52</sup> Thompson, et al. indicate an influence of alternative splice variants of mucins involved in survival outcome for PDAC, including unique expression patterns of these isoforms.<sup>53</sup> In several studies regarding mucin expression, MUC1 is associated with high aggressiveness of the tumor, resulting in invasive growth and with low patient survival rates.<sup>48, 54</sup> Besides the expression level of mucins, the time-dependent expression of different mucins at different PanIN progression states can be a useful characteristic for subtyping. It has been shown, that expression of MUC6 and MUC5A is an early event (mainly occurs in PanIN-1A, PanIN1B, PanIN2) and MUC1 expression appeared to be a late event (mainly occurs in PanIN2, PanIN3) during lesion progression.<sup>55</sup> A further dimension of variety is that mucins carry high molecular weight glycan structures of high complexity and not all glycan structures are yet known and characterized, why analysis of mucins are challenging, especially for proteomic analysis. Membrane-bound mucins are linked to the membrane via glycosylphosphatidylinositol (GPI)-anchor, which supports the variety of modifications on mucins in the extracellular space. Further suggestions are, that O-glycosylation, which is a modification of O-linked  $\beta$ -N-acetylglucosamine (O-GlcNAc) on hydroxy groups of serine or threonine, can

block these sites for phosphorylation events involved in cell signaling.<sup>56, 57</sup> The reciprocity between O-GlcNAc and O-phosphorylation expands the binary nature of phospho-signaling (switching on or off via phosphorylation or not), which was proofed by several working groups.<sup>58-60</sup> This is achieved through different protein states arise from these modifications (phosphorylated, glycosylated, or non-modified), that can respond to various conditions as well as to influence intra- and extracellular protein interactions. Extracellular mucins are strongly involved in the formation of a tumor microenvironment and modulate cell-cell communication in oncogenic signaling. It is shown, that mucins interact with various receptors, e.g., receptor tyrosine kinases (RTKs), epidermal growth factor receptor (EGFR) and fibroblast growth factor receptor (FGFR), which influences signaling cascades involved in cell survival, proliferation and metastasis.<sup>51, 52</sup> The diversity of glyco-modifications leads to a variety of possible epitopes for antibodies. Several clinicopathological relevant antigens are located on the surface of mucins, e.g. the blood group precursor Thomsen-Friedenreich-Antigen (TF-Antigen, or T-Antigen) and its structural related sub form, the Tn (T-antigen nouvelle) with isoforms carrying sialic acid modification to form the sialyl-T/Tn antigens.<sup>61, 62</sup> Under healthy conditions, these antigens are highly modified, carrying long carbohydrate side-chains, while in over 90% cases of adenocarcinoma, the unmasked T/Tn antigens are present, only carrying sialic acid residues (sialyl-T/Tn-antigen), the core 1 glycan (Gal $\beta$ -3 GalNAc-R, T-antigen) or *N*-acetylgalactosamine (GalNAc, Tn-antigen).<sup>63-65</sup> These cancer-associated changes caused by inappropriate glycosylation of mucins suggests downregulation of glycosyl-transferases involved in carbohydrate chain extension and the upregulation of sialyl-transferases, which results in the accumulation of these antigen precursors (T, Tn, sialyl-Tn).<sup>66</sup> Hypersialylation is a known deregulation in protein glycosylation of many cancer types and is known to be involved in modulation of immune response and inflammatory processes. Besides sialylation, fucosylation play a crucial role in forming the so-called Lewis-antigens (Le<sup>a/x/y</sup>, sialyl-Le<sup>a/x/y</sup>), for which the most prominent antigens are CA19-9 and CA-50. In several adenocarcinomas Lewis antigens are linked to mucins, glycolipids and other non-mucin glycoproteins and are released into blood circulation or secreted to body fluids, which makes them appropriate biomarker for disease progression. A high specificity of Lewis antigen detection in pancreatobiliary cancers has been widely published.<sup>67-71</sup> The monoclonal antibody SPan-1 MAb can be used for the detection of both CA19-9 and CA-50, while DU-Pan2 Mab is specific for CA-50 epitopes.<sup>69, 72, 73</sup> Even though other mucin-bound, carbohydrate antigens for pancreatic cancer are known, such as CA242, CA195, CA125, DUPAN 2 and SPan-1, but they all lack in sensitivity and/or specificity. CA19-9 with a median sensitivity of 79 (70 – 90%) and specificity of 82 (68 – 91%) is recommended to use as prognostic marker for pancreatic cancer in the clinical context.<sup>74-77</sup> Differential diagnosis of pancreatic cancer with CA19-9 is still challenging due to the high serum levels in benign conditions, such as pancreatitis, cholestasis and jaundice, which makes it insufficient for diagnostics.<sup>78, 79</sup>

#### 4.2.4 Similarities to other adenocarcinoma types of pancreatobiliary system

Pancreatic cancer shares commonalities with various types of adenocarcinomas, reflecting similarities in precursor lesions and pathological processes. For instance, precursor lesions of the pancreatobiliary system exhibit parallels with those found in other anatomical locations. These precursor lesions exhibit various grades of dysplasia and share similarities in their morphological and genetic characteristics. Aberrant expression of Pdx1 in BillIN became frequent along with progression of its atypia. In this context, it is suggested that transcription factors related to the development of the pancreas and biliary tract, such as Pdx1 and Hes1, are involved in the malignant transformation of the pancreas and also biliary tract.<sup>15</sup> These processes may be also involved in the similarities of hilar CCA and PDAC. Aberrant expression of S100P was frequent in PanIN and BillIN in addition to PDAC and hilar CCA, also supporting this suggestion.<sup>15</sup> Understanding these similarities aids in elucidating common pathways of tumorigenesis and may inform diagnostic and therapeutic strategies across different adenocarcinoma types. (Table 1 summarizes the precursor lesions of the pancreatobiliary system.)

*Table 1: Precursor lesions of the pancreatobiliary system. (Abbr.: BillIN = biliary intraepithelial neoplasm, IPNB = intraductal papillary neoplasm of bile duct, MCN = mucinous cystic neoplasm, ICPN = intracholecystic papillary neoplasm of gallbladder, IPMN = intraductal papillary mucinous neoplasm of pancreas, IOPN = intraductal oncocytic papillary neoplasm of pancreas, ITPN = intraductal tubular papillary neoplasm of pancreas, IAPN = intraampullary papillo-tubular neoplasm, table extracted from Nakanuma, Kakuda, Sugino, Sato and Fukumura <sup>16</sup>)*

Anatomical Locations	Precursors
Intrahepatic large bile duct	BillIN (low-grade and high-grade)
Perihilar bile duct	IPNB
Distal bile duct	MCN
Gallbladder	BillIN (low-grade and high-grade)
	ICPN
	MCN
	Pyloric gland adenoma
Pancreas	PanIN (low-grade and high-grade)
	IPMN
	MCN
	IOPN
	ITPN
Ampulla	Intestinal adenoma (low-grade and high-grade) IAPN
	Flat intraepithelial neoplasm

### 4.3 Testicular carcinoma – Germ-cell tumors (GCT)

The most frequent cancer among young men aged between 20 – 34 is testicular carcinoma.<sup>80</sup> Over 84% of testicular cancers are germ-cell tumors, divided to seminomas (44%), mixed germ cell tumors (35%), yolk-sac tumors (YST, 9%), teratoma (8%), embryonal carcinoma (3%) and spermatocytic seminoma (1%), for which the distribution is shown in Figure 2:

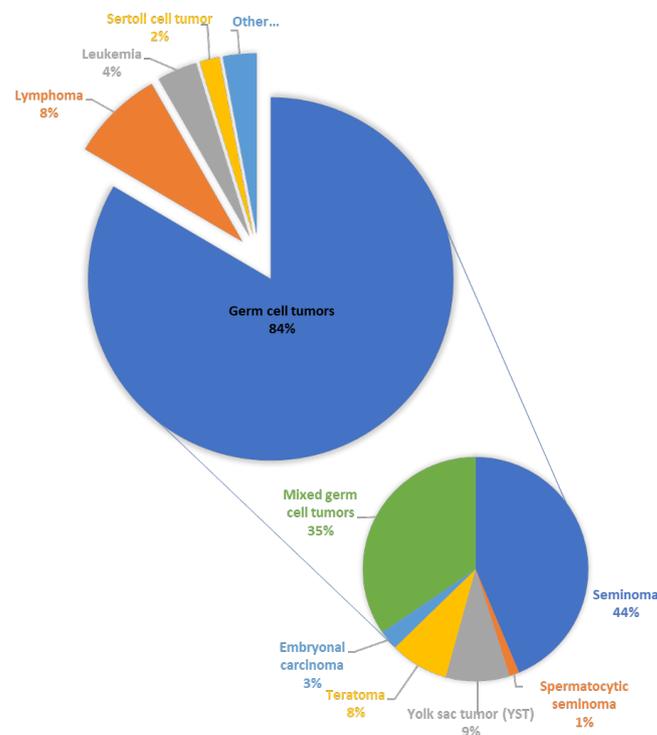


Figure 2: Distribution of testicular carcinoma types, including the sub-distribution of different germ-cell tumor types. Extracted from Gill et al.<sup>81</sup>

Both, seminoma and non-seminoma GCT arise from a defective or misguided primordial germ cell, that develops into a germ cell neoplasia in situ (GCNIS).<sup>82</sup> These tumors can occur in the gonads (testes and ovaries) or extragonadally and exhibit a wide range of histological presentations and clinical behaviors.

#### 4.3.1 Characterization of GCT

Based on the chromosomal complement and developmental potential GCT can be classified into five groups.<sup>83</sup> The first group are (immature) teratoma and yolk-sac tumors further present in neonates or young-aged children. The cells of origin for this group are immature germ-cells and occur extragonadal and gonadal. The type two GCT also occur gonadal and extragonadal but the cells of origin are more mature germ-cells and gonocytes. The type two GCT is more present in older children 13-15 years old and the median age over all cases is 35 and 25.<sup>83</sup> The type three GCT is a spermatocytic seminoma and is only present in the testis for patients older than 50 years. The GCT that is only present in the ovary is classified into the group four GCT, in which an oogonia/oocyte is the cell of origin. This GCT is more present in children/adult women. The

last type of GCT is the group five, that occurs in the fertile period and the cell of origin is an empty ovum or spermatozoa.

#### 4.3.2 Yolk-sac tumors (YST) and therapy resistance

The most frequent GCT in children at an age of 0-4 years are the teratoma and yolk-sac tumors (YST).<sup>81, 82, 84</sup> Yolk-sac tumors and choriocarcinomas develop from so called non-seminomatous stem cell-like embryonal carcinomas, which have the potential to differentiate into all three germ layers, including teratoma or extraembryonic tissues.<sup>82</sup> YST represent a very aggressive form of GCT and is characterized by their rapid growth and high potential to metastasize. For GCT the standard treatment option is cisplatin-based chemotherapy, but many cases of YST receive a poor prognosis due to the development of therapy resistance.<sup>85</sup> Mechanisms underlying the formation of therapy resistance include activation of NRF2-pathway, elevated DNA repair mechanisms or decreased/deregulated apoptosis induction.<sup>86</sup> To overcome cisplatin therapy resistance in YST, new treatment approaches, such as antibody-drug conjugates or multikinase inhibitor-based therapeutic options are in the focus of research. The tetraspanin membrane protein CLDN6 is a cancer associated cell surface biomarker with high selectivity for seminomas, embryonal carcinomas and yolk-sac tumors.<sup>87</sup> Due to the high selectivity, it is suitable to use as a target for antibody-drug conjugates (ADC), in which a specific antibody is coupled to a potent cytotoxin, such as monomethyl auristatin E (MMAE).<sup>85, 88</sup> With the help of molecular profiling the transcription factor FOXA2 is found as key driver for the formation of therapy resistant YST, subsequently inducing AFP, GPC3, APOA1/APOB, ALB and GATA3/4/6 expression, addressing signaling pathways, such as BMP-, WNT- and MAPK signaling.<sup>89</sup> For the efficient treatment of cisplatin therapy resistant YST, the use of multikinase inhibitor-based therapeutic is recommended for testing from *Skowron et al.*<sup>85</sup> These new therapeutic options offer possibilities to treat resistant carcinomas, such as resistant YST with several advantages, including less side effects and simplified intake of these chemotherapeutics. ADCs are indeed becoming an increasingly favored approach for targeting cancer, especially in advanced stages where metastasis is prevalent. In comparison to conventional chemotherapeutics, both ADC and multikinase inhibitors exhibit fewer side effects.

#### 4.3.3 Germ cell tumor-related somatic-type malignancies (STM)

A somatic-type malignancy (STM) is a rare lethal subtype of malignant teratoma, occurring in up to 6% of GCT cases.<sup>90, 91</sup> It is a secondary tumor component of non-seminomas and resembles cancers typically found in other organs and tissues.<sup>92</sup> Due to genetical and morphological characteristics, teratoma and YST are in the same group of GCT subtypes, which also explains the high occurrence of chemotherapy resistance against cisplatin-based chemotherapy for both types. STM refer to a recurrence rate of 81% and an overall disease-specific mortality rate of 24%.<sup>91</sup> Because of the rarity of STM occurrence, the molecular mechanisms have not yet been extensively studied. *Bremmer et al.* firstly unravels the molecular mechanisms underlying STM transformation.<sup>93</sup> With the help of genomic, proteomic, and DNA methylation (epigenomic) data they found similarities of carcinoma-related STM to YST and sarcoma-related STM to teratoma.

Signaling pathways, such as AKT, FGF, MAPK, and WNT are found to be deregulated with special focus on FGF signaling as possible therapeutical target pathway.<sup>93</sup> Several biological pathways are found and indicate the high rate of therapy resistance within STM, such as pathways involved in oxidative stress, DNA helicase activity, toxin transport, apoptosis and cell cycle processes.

#### 4.3.4 Growing teratoma syndrome (GTS)

A very rare phenomenon, firstly described for six case reports in 1982 by *Logothetis et al.* is the so called growing teratoma syndrome (GTS).<sup>94</sup> This condition is characterized by a continued growing tumor mass present in some patients during cisplatin chemotherapy treatment and reduced serum tumor markers, such as alpha-fetoprotein (AFP), beta-human choriogonadotropin (beta-hCG), lactate dehydrogenase (LDH). The only treatment option for GTS is the complete surgical resection of the tumor masses, and due to the rarity of this condition, the biological mechanisms underlying its progression are yet unknown.<sup>95</sup> Based on molecular and epigenetic subtyping *Pongratanakul et al.* proposed a new definition for GTS as follows: "*The GTS describes a continuously growing teratoma that might harbor occult non-seminomatous components considerably reduced during therapy but outgrowing over time again.*"<sup>96</sup>

## 5 Manuscript 1:

Uncover the archived treasures – Challenges and opportunities in FFPE tissue proteomics



# Uncover the archived treasures – Challenges and opportunities in FFPE tissue proteomics

Stella Pauls<sup>2</sup>, Anja Stefanski<sup>1</sup>, Kai Stühler<sup>1,2</sup>

<sup>1</sup>Proteome Research, Institut für Molekulare Medizin, Medizinische Fakultät, Heinrich-Heine-Universität, Universitätsstraße 1, D-40225 Düsseldorf

<sup>2</sup>Molecular proteomics laboratory (MPL), Institut für Molekulare Medizin, Medizinische Fakultät, Heinrich-Heine-Universität, Universitätsstraße 1, D-40225 Düsseldorf

## 1 Introduction

### 1.1 Formalin fixation, paraffin embedding (FFPE)

Formalin fixation is an important standard method in immunohistochemistry (IHC), since Ferdinand Blum introduced it in 1893.<sup>11</sup> Over the past years pathologies all over the world collected several different tissue samples and stored them in huge tissue archives. Nowadays, estimations assume that every year in Europe, 20-30 million cases of tissues are taken for biopsy and in many hospitals multiple tissue samples are stored, resulting in an ongoing growth of these tissue banks.<sup>97</sup> These archives provide access to large patient cohorts for -omic research, to study molecular patterns of a wide range of diseases for the discovery of novel biomarkers.<sup>98</sup> Formalin fixation followed by embedding in paraffin wax (FFPE) is nowadays the gold standard method for the preservation of tissues.<sup>99</sup> Besides the high availability of FFPE tissues compared to fresh frozen tissues, the high quality of morphology reveal a precise microdissection of several tissue areas of interest.<sup>97</sup> Furthermore, FFPE tissues are molecular snap-shots of a given condition/state, why it is suitable to use in biomarker discovery and for the analysis of the developmental progression of tissues in diseases. An important field to provide information about protein abundance, location, modification and protein-protein interactions is the analysis of the proteome by using proteomic analysis.<sup>100</sup> Study proteins from FFPE tissues beyond immunostaining has been technical challenging in the last years due to the crosslinking properties of formalin fixation and low protein recovery for LC-MS analysis.<sup>101</sup> Therefore many working groups developed extraction methods for LC-MS proteomic analysis to achieve comparable results to the matching fresh frozen tissue.<sup>8, 102-106</sup>

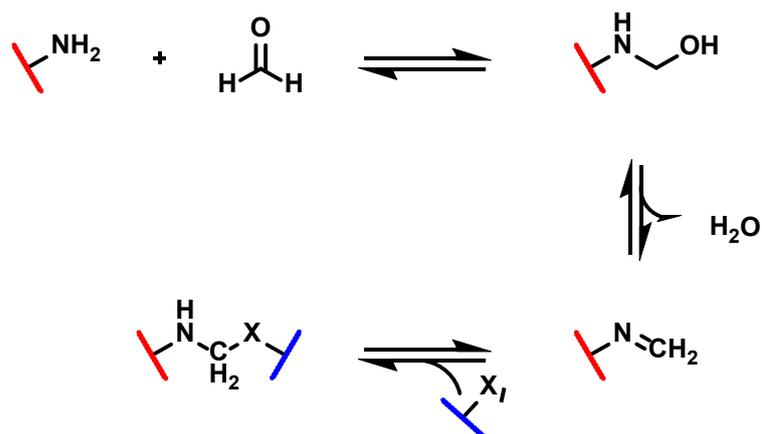
### 1.2 FFPE sample collection and storage

For comparable results in translational research, the sample collection and archival of tissue biopsies need to undergo optimized conditions. The preanalytical treatment of tissues (surgical resection, fixation, paraffin embedding, freezing and tissue banking) is an important step for standardized molecular downstream processing. Formalin fixation, paraffin embedding, freezing and tissue banking are well defined in the literature, except the time between surgical resection and grossing/fixation as well as fixation time and temperature.<sup>107</sup> One problem is the transfer of resected tissue from the surgical theater to the pathology laboratories, which is normally done by

conservation in formalin solution. The high toxicity of formalin leads to strict handling guidelines such as the requirement to handle it under the fume hood, which is not feasible in the surgery theater. Therefore alternative methods for the transfer of resected tissue into the pathology lab were developed. Bussolati and coworkers presented a short-time preservation method for tissues under vacuum to avoid the exposure to formalin, and to provide nucleic acids of acceptable quality.<sup>108</sup> Nevertheless, rapid formalin fixation leads to a good quality preservation of proteins, while shock-freezing provides high quality preservation for DNA and RNA but not for proteins due to degradation and ischemia while thawing, which also leads to a change in post-translational modification (PTM) patterns, like phosphorylation.<sup>109</sup> To achieve standardization for the preanalytical tissue treatment, in 2005 the U.S. National Cancer Institute (NCI), the European Union (EU) and other organizations, developed best practices for biospecimen resources, which supports the need for a system of quality control combined with a system for auditing.<sup>110, 111</sup> In addition to quality control of sample collection methods, the storage conditions have a huge influence on sample quality. Especially for antigen preservation, inadequate tissue processing resulting in retention of endogenous water as well as high humidity in storage can result in significant protein degradation and a loss of antigenicity.<sup>103, 112</sup> Tissue storage over several years can result in reduced protein recovery, as well as over-fixation for longer than 24 h, caused by irreversible formalin crosslinks.<sup>103</sup> Qiagen developed a commercially available buffer system (EXB plus, Qproteome FFPE Tissue Kit, Qiagen) to overcome low protein recovery from long-time stored and/or over-fixed FFPE tissues, which improves standardization for tissue samples taken at different time points. Improving standardization reveals access to very large patient cohorts for several, different diseases and enables proteomics studies with powerful statistics.

### 1.3 Crosslinking reaction within FFPE treatment

Mechanistically formalin fixation of proteins occurs at free  $\epsilon$ -amino groups of lysines, in which formaldehyde forms the Schiff base in the first place followed by acid catalyzed cross linkage between these amino groups and reactive CH-groups of phenolic and imidazole rings which can be found on side chains from phenylalanine, tyrosine and proline.<sup>113, 114</sup> In addition to amino groups and reactive CH groups derived from amino acids containing phenolic and imidazole rings, various nucleophilic side chains, such as amino-, thiol-, and hydroxy- groups, can participate in the subsequent formation of the crosslinking adducts. Formalin fixation introduces a covalent bond linking functional groups of two different macromolecules, such as proteins and DNA/RNA.<sup>115</sup> Figure 3 shows the general mechanism of formalin fixation. As already mentioned, the first step is the formation of the methylol adduct, which forms the Schiff base (methylene adduct) after dehydration in the second step. The crosslinking occurs at the methylene adduct by nucleophilic attack of another macromolecule carrying any nucleophilic side chain (NH<sub>2</sub>-R, SH-R, Phe-R). The global crosslinking of all macromolecules within a tissue sample maintains tissue morphology and cellular architecture over long periods, even stored at room temperature.



X : nucleophilic side chain from other proteins or nucleic acids (SH-R, NH<sub>2</sub>-R, Phe-R)

Figure 3: General mechanism of formalin fixation in tissue samples. In the first step an amine-group of a macromolecule and formaldehyde forms the methylol adduct, followed by dehydration and the formation of the Schiff base. A nucleophilic attack of another macromolecule carrying any nucleophilic side chain, like thiol- amino- or phenyl-residues, enables the crosslinking between these two macromolecules.

The reversion of these crosslinks is the greatest challenge for subsequent protein and/or DNA/RNA analysis from FFPE specimens. Incomplete reversal can result in higher numbers of modified peptides arising from FFPE treatment, decreasing overall protein identification.<sup>116</sup> Due to the fact, that formalin can form a covalent bond between two macromolecules it can be used for interaction studies concerning protein-protein interaction via proteomic studies using mass spectrometry and protein-DNA interactions using CHIP sequencing techniques.<sup>117, 118</sup> One of the main advantages of formaldehyde crosslinking is its partial reversibility under mild conditions, which allows for the recovery of native proteins after analysis, a feature particularly useful in studies where the preservation of protein function and structure is critical. However, formaldehyde crosslinking has its disadvantages. Formaldehyde can crosslink any two proteins in close proximity, regardless of whether they are functionally related, leading to high background noise and complicating result interpretation. While formaldehyde crosslinking is partially reversible, complete reversal is not always achievable, which can hinder the recovery of fully functional proteins and interfere with certain types of analyses. Additionally, prolonged exposure to formaldehyde can lead to protein denaturation, affecting the structural integrity and functional properties of proteins, which might result in artifacts in the data obtained from crosslinked samples.

## 1.4 Protein recovery method for FFPE tissues

### 1.4.1 Heat induced antigen retrieval (HIAR)

In 1991 Shi and coworkers published the first approach of heat induced antigen retrieval from over-fixed and/or long-term stored FFPE tissues for immunohistochemical staining.<sup>119</sup> They tested 52 monoclonal and polyclonal antibodies on FFPE tissues treated with the presented antigen retrieval protocol, in which the slide staining without antigen retrieval failed due to over-fixation and/or long-term storage or for antibodies, that are typically unreactive due to formalin

fixation. With this method, 39 antibodies showed an increased immunohistochemical staining, nine antibodies showed no change and four antibodies showed a decrease staining. Furthermore, after antigen retrieval incubation times for immunohistochemical staining could be reduced and the dilution of primary antibodies could be increased as well as pre-digestion of tissues could be omitted. The single steps from the procedure of the first antigen retrieval protocol are shown in Figure 4.

1. Tissue sections were deparaffinized and rehydrated to water.
2. Endogenous peroxidase was blocked with 3% H<sub>2</sub>O<sub>2</sub> for 5 min.
3. Slides were washed with distilled water for 5 min.
4. Slides were then placed in plastic Coplin jars containing either distilled water, a metal solution of saturated lead thiocyanate, or 1% zinc sulfate.
5. Jars were covered with loose-fitting screw caps and heated in the microwave oven for either 5 or 10 min. Sometimes a 10-min heating time was divided into two 5-min cycles with an interval of 1 min between cycles to check on the fluid level in the jars.
6. After heating, the Coplin jars were removed from the oven and allowed to cool for 15 min.
7. Slides were then rinsed in distilled water twice and in PBS for 5 min.
8. Treated slides were immunostained as described below

Figure 4: Protocol for the first heat-induced antigen retrieval proposed by Shi *et al.*<sup>119</sup>

One of the crucial steps was the addition of a lead solution while doing the HIAR protocol, which was introduced based on the hypothesis that heavy metals form insoluble complexes with polypeptides to improve protein preservation.<sup>120</sup> They have shown, that even for the retrieval of antigens, the addition of metal ions restored immunostaining for partially formalin sensitive antibodies, like vimentin and reversed the deleterious effects on antibody epitopes by formalin fixation.<sup>119</sup> For the 20<sup>th</sup> anniversary of the HIAR method, Shi *et al.* summarized all applications in which FFPE material was used, since the introduction of the method and showed that formalin fixation has become a standard method (“Gold standard”) for tissue preservation and immunohistochemical applications.<sup>121</sup> Further applications, like Western Blot, SDS-PAGE and LC-MS or IMS based Proteomics requires the extraction of the proteins from FFPE tissue slides, in which the most crucial factor is the heating step to ensure complete protein recovery.<sup>122-125</sup> Next to the heating step in the FFPE tissue protein extraction protocol, the choice of an efficient buffer system is decisive for successful protein recovery. Jiang *et al.* showed an 13-fold increase in protein concentration by incubating FFPE homogenates for 30 min at 100 °C compared to no heating and yielded the highest protein recovery with buffers containing chaotropes like guanidine-HCl and detergents like SDS, for which the results are summarized in Table 2.<sup>126</sup> Even if they showed higher protein concentrations while identifying the same number of proteins in the guanidine-containing compared to the SDS-containing buffer, guanidine-HCl is commonly

avoided for protein extraction in MS-based proteomics, because it hampers trypsin activity.<sup>127</sup> For in-solution digestion it is inevitable to remove or reduce the concentration of chaotropes and/or detergents because the most ingredients can decrease trypsin activity, which results in low digestion efficiencies and poor sequence coverage. The reduction of chaotropes, like guanidine-HCl, without protein loss is a difficult challenge, why the most procedures introduce a dilution step of the extraction buffer before tryptic digestion.<sup>126-128</sup> Dilution may be problematic for samples with low protein content because of limited tissue material from the region of interest, therefore different digestion strategies are required for sample lysates that contain guanidine-HCl. In 2013 *Poulsen et al.* published an efficient protein digestion protocol for sample lysates containing guanidine-HCl, in which digestion can be completed within 30 min with endopeptidase Lys-C and even with incubation at higher temperatures (95 °C) no chemical artifacts were observed.<sup>129</sup> Another important approach for the recovery of proteins is the replacement of detergents with organic solvents, like methanol, acetonitrile, TFA or TFE in the extraction buffer.<sup>130-133</sup> Mostly this replacement took place, when the goal of the procedure was to selectively extract proteins or peptides of interest, for which denaturing conditions does not lead to precipitation but solvation, like for membrane proteins or low molecular weight proteins and peptides.<sup>130, 131, 133</sup>

*Table 2: Comparison of different protein extraction buffers for FFPE tissue disruption protocols and comparison of protocols using and avoiding heat treatment. Comparison on the level of the protein concentration and the number of identified proteins. Data extracted from Jiang et al. (2007).<sup>126</sup>*

<i>Extraction buffer/method</i>	<i>Material</i>	<i>Protein concentration</i>	<i>No. of identified proteins</i>
<i>6 M guanidine-HCl without heating</i>	fresh frozen	17.7 mg/mL	976
<i>6 M guanidine-HCl without heating</i>	FFPE	0.82 mg/mL	130
<b><i>2% SDS with heating</i></b>	<b>FFPE</b>	<b>6.4 mg/mL</b>	<b>820</b>
<i>Direct digestion of tissue homogenate</i>	FFPE	-	331
<b><i>6 M guanidine-HCl with heating</i></b>	<b>FFPE</b>	<b>10.9 mg/mL</b>	<b>827</b>
<i>The pellet with CNBr treatment</i>	FFPE	-	526

While these methods exclude hydrophilic proteins to improve protein extraction of hydrophobic proteins, these are not useful for shot-gun proteomics, in which the goal is to identify a broad range of proteins in the proteome. On the other hand, the selective extraction of proteins/peptides with hydrophilic behavior improve sample preparation, for example for shot-gun membrane proteome analysis.<sup>133-138</sup> Trifluoroethanol (TFE) exposed to be an organic solvent, which is capable to extract peptides with a significant higher mean grand average of hydrophobicity (GRAVY), compared to methanol as organic solvent for membrane protein extraction (-0.107, resp. -0.465).<sup>133</sup> The most protocols, using organic solvents, like TFE for protein extraction, compare these results to the universally applicable urea-based protein extraction protocol but as shown in Table 2, results from SDS-based protein extraction for FFPE tissues assumes that this method is potentially as efficient as protein extraction based on organic solvents. Wang and coworkers compared proteomic results achieved by the three top extraction methods (urea-based, SDS-based and TFE-based protein extraction) by using moderate amount of fresh frozen

tissue samples and low amounts (~1 million MCF-7 cells) of cultured human cell samples.<sup>132</sup> Results, summarized in Table 3, showed increased identification of peptides and proteins using the TFE-method compared to detergent-based extraction methods for both, moderate and low sample amounts, and higher reproducibility between replicas. Likewise, this TFE-based protein extraction protocol was applied to very low sample amounts, containing ~5000 MCF-7 cells, which resulted in identification of  $104 \pm 24$  proteins.<sup>132</sup> Even though the TFE method is a very efficient protein extraction method avoiding additional cleaning steps caused by the volatility of TFE, the method has rarely been used for FFPE tissue preparation so far.<sup>139-141</sup>

Table 3: Comparison of the top protein extraction methods (**urea-based, SDS-based and TFE-based protein extraction**) for moderate sample amounts (~5 mg mouse brain tissue sample) and low sample amounts (~1 million MCF-7 cells). Data summarized from Wang et al. (2005).<sup>132</sup>

Method	Sample type and amount	Recovered peptides [ $\mu$ g]	Different peptides identified	Proteins identified
CHAPS method (urea-based)	~4.7 mg fresh mouse brain tissue	$244 \pm 25$	$573 \pm 75$	$257 \pm 41$
TFE method	~4.7 mg fresh mouse brain tissue	$231 \pm 73$	$760 \pm 32$	$323 \pm 5$
SDS method	<b>~1 million MCF-7 cells</b>	122	$396 \pm 100$	$214 \pm 48$
TFE method	<b>~1 million MCF-7 cells</b>	125	$457 \pm 30$	$251 \pm 16$

Coscia, et al. compared three different protein extraction protocols for FFPE tissue proteomics on macrodissected Glioma tissue samples, in which SDS-containing buffer was finally compared to TFE-containing buffer besides commercially available RapiGest buffer.<sup>141</sup> This comparison showed no difference between all three buffers, in which around 4,800 proteins were quantified in all tissue pieces independently from the extraction buffer used for protein extraction. In addition to the buffer components, effective antigen retrieval is highly dependent on pH. Elevated pH levels tend to enhance protein recovery, whereas lower pH levels result in decreased protein recovery.<sup>142</sup> A basic pH range of 8.0 to 9.5 is recommended for optimal protein retrieval.<sup>116</sup> Several different variations concerning the HIAR protocol are available, but in general three main components for successful antigen retrieval from FFPE tissues can get summarized: 1.) Heating treatment, 2.) Detergent-based buffer (preferably SDS), 3.) Basic pH for extraction (8.0 - 9.5). Furthermore, the HIAR protocol is applicable on a broad range of different FFPE tissue samples and enables a straightforward de-crosslinking of formalin bridges and an efficient extraction of proteins. For the field of FFPE tissue analysis, Shi et al. concluded in 2019, that 'the simple method of HIAR is the gold key to open the door of this treasure' and until now, this protocol seems invincible for this purpose.<sup>143</sup>

## 1.5 Proteomic sample preparation

### 1.5.1 2D gel electrophoresis

2D gel electrophoresis is a powerful technique, developed in 1975, which is widely used for separating and analyzing complex protein mixtures based on their isoelectric point (pI), by isoelectro focusing (IEF) as first dimension and molecular weight, by SDS-PAGE as second dimension.<sup>144</sup> This method provides a way to visualize the proteome by creating a two-dimensional map of proteins within a complex sample, enabling the detection of changes in protein expression levels, post-translational modifications, and protein isoforms. The first decade of the 20<sup>th</sup> century the 2D gel electrophoresis has been widely used for protein purification prior LC-MS/MS-based proteomic analysis and was intensively reviewed.<sup>145-148</sup> Since the last decade, other methods are favorable used for LC-MS/MS-based proteomic sample preparation, such as FASP and Capture coaggregation and are further described in chapter 1.5.2 and 1.5.3. Nevertheless, the fractionation power of 2D gel electrophoresis enables other application options for the field of proteomics. The 2D gel electrophoresis is involved in special applications for proteomic analysis, like for samples with high dynamic ranges, PTM analysis and/or native proteomics.<sup>149</sup>

### 1.5.2 Filter-Aided Sample Preparation (FASP)

FASP, is a widely used protein purification method for proteomic sample preparation of several sample types. It allows the efficient and reproducible processing of protein samples for mass spectrometry-based proteomic analysis with small sample loss. Proteins in the sample are denatured using chaotropic agents, like urea and guanidine hydrochloride, followed by reduction of disulfide bonds and alkylation of cysteine residues with a reductant like tris(2-carboxyethyl)phosphine (TCEP) or dithiothreitol (DTT). Subsequent digestion by trypsin yields peptides, which are then separated from larger molecules via filtration and recovered for LC-MS/MS analysis. FASP offers several advantages for proteomic sample preparation, including reduced sample loss, improved reproducibility, and compatibility with a wide range of biological samples.<sup>4, 150-153</sup> For FFPE tissue proteomics FASP has been widely used,<sup>4, 8, 151</sup> but Föll et al. showed, that direct trypsinization of the tissue slides result in similar peptide recovery compared to FASP and with similar reproducibility.<sup>8</sup> Within FASP an unexpected side reaction of formaldehyde peptide derivatization was found by Tang et al., which lowers overall protein identification and is a phenomenon coming from the commercial available filters.<sup>154</sup> Further known contaminants coming from commercial available filters are polymers, such as polyethylene glycol (PEG), polypropylene glycol (PPG), polysiloxanes ect.<sup>155</sup> These contaminations are highly ionizable and can reduce peptide identification and reproducibility. Experiences from our group are that several different commercially available filters for FASP have high quality variations within different product batches which can affect LC-MS/MS analysis due to contaminations and reduce reproducibility and peptide identification. Quality variations pose risks of sample loss or inappropriate sample preparation, particularly for small sample amounts or rare samples, thereby

impacting protein identification. However, in enrichment methods for modified peptides like N-glycoproteomics or phosphoproteomics, FASP represents a critical cleanup step preceding affinity enrichment.<sup>156</sup> For glycopeptide enrichment using a lectin affinity approach, it is essential for the peptide mixture to be free of proteinases after digestion, which can be provided by FASP.<sup>157</sup>

### 1.5.3 Capture coaggregation / SP3

SP3 (Single-Pot, Solid-Phase-enhanced Sample Preparation), is a robust capture coaggregation method for proteomic sample preparation that enables the selective enrichment of target proteins or protein complexes from complex biological samples using paramagnetic beads.<sup>158, 159</sup> SP3 combines the principles of capture coaggregation with enzymatic digestion and has gained popularity due to its simplicity, efficiency, and reproducibility. The critical steps of the SP3 method are shown in Figure 5, starting from a protein mixture, typically cell or tissue lysates containing contaminants from cell debris and/or buffers or after in-solution reduction and alkylation. To remove the containing contaminants, a bead-containing solution is added to the protein mixture. These beads have paramagnetic properties and are usually surface-functionalized with carboxyl- or amine-groups.

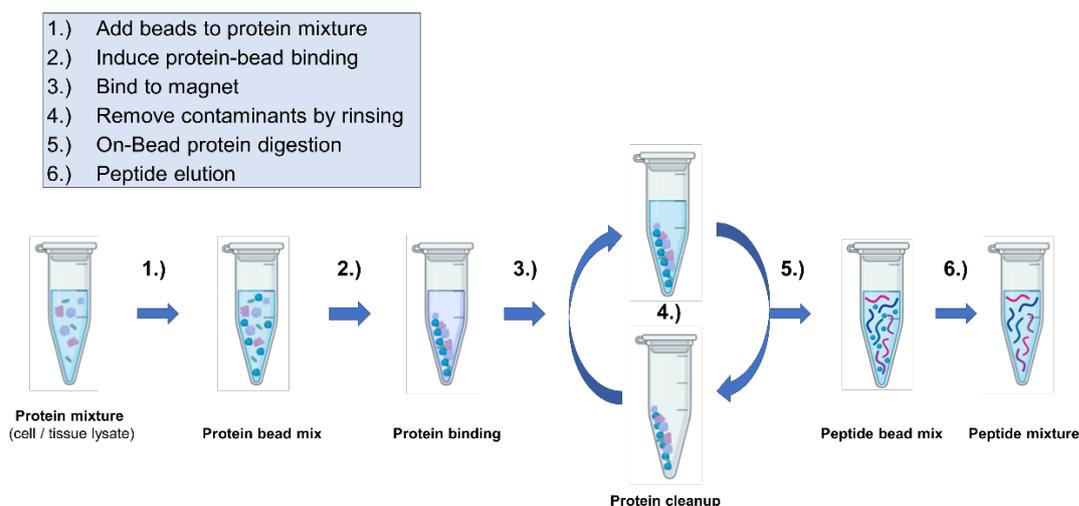


Figure 5: Critical steps for the SP3 method. Designed with BioRender.com and adapted from Hughes et al.<sup>159</sup>

Due to these properties proteins aggregate on the surface of the magnetic beads after adding an organic solvent, such as acetonitrile, methanol or ethanol up to final concentrations of 50 - 80%.<sup>159</sup> After bead binding the magnetic beads will then localize on the tube wall by using an external magnet and the remaining supernatant can be removed. This allows to remove contaminants by adding several washing steps with the benefit of minimal sample loss. After protein cleanup, a simple solvent to buffer change enables on-bead digestion of the bead-bound proteins. After digestion the remaining supernatant contains the digested peptides, that can be obtained by separation of the supernatant from the magnetic beads, either by using the external magnet. Further evaporation of the digestion buffer and resolving of the remaining peptides in 0.1% TFA,

the peptide solution can get subjected to LC-MS/MS analysis. SP3 is compatible with a wide range of detergents, salts, chaotropes and solvents with the ability to tolerate moderate to high concentrations.<sup>159</sup> Therefore many different proteomic sample preparation workflows benefit from the simplicity of the SP3 method. Several applications of the SP3 method are used in the field of scarce sample proteomics, because this method enables protein preparation with minimal sample loss.<sup>160, 161</sup> Sielaff et al. showed, that SP3 outperforms especially for low sample amounts (1 µg HeLa protein amount) against other methods, such as FASP and in-StageTip digestion.<sup>160</sup>

## 1.6 MS Analysis methods

### 1.6.1 DDA

Data-dependent acquisition (DDA) is a pivotal mass spectrometry technique extensively employed in the analysis of formalin-fixed paraffin-embedded (FFPE) tissue proteomics.<sup>1-8, 141</sup> This approach offers both notable benefits and inherent challenges in elucidating the complex protein landscapes within FFPE samples. One of the primary advantages of DDA lies in its ability to provide comprehensive coverage of the proteome with low expenditure. By dynamically selecting precursor ions based on their intensity or abundance, DDA facilitates the acquisition of tandem mass spectra from a diverse range of peptides within a single analytical run. This dynamic prioritization optimizes the detection of abundant peptides while still capturing lower abundance species, thereby enhancing the depth and sensitivity of proteomic analysis in FFPE tissues. Moreover, DDA enables the identification of low-abundance proteins, which is crucial for biomarker discovery and understanding disease mechanisms. Despite the challenges posed by protein cross-linking and modifications induced during formalin fixation, DDA's efficient precursor ion selection and fragmentation mechanisms make it possible to uncover previously undetected proteins, thus expanding our knowledge of FFPE tissue proteomes.<sup>7</sup> However, DDA is not without limitations. Variability in peptide identification reproducibility across replicate analyses can hinder the reliability of results, particularly in FFPE tissue proteomics where sample heterogeneity and preparation variability are prevalent. Additionally, DDA may miss detecting low-abundance proteins, especially those masked by highly abundant species or affected by chemical modifications induced during tissue fixation. Dynamic range compression is another challenge associated with DDA, as the prioritization of abundant precursor ions for fragmentation can lead to reduced sensitivity for detecting low-abundance peptides. This compression may obscure biologically relevant changes in protein expression levels, particularly in FFPE tissue samples with complex protein profiles. Furthermore, quantitative accuracy can be compromised by factors such as variability in precursor ion selection, ionization efficiency, and fragmentation patterns, impacting the reliability of quantitative measurements derived from DDA data. In conclusion, while DDA represents a powerful tool for FFPE tissue proteomics, researchers must carefully consider its benefits and limitations in experimental design and data interpretation. Optimizing DDA parameters and integrating complementary analytical approaches can enhance the depth, sensitivity, and reliability of proteomic studies in FFPE tissues, ultimately advancing our understanding of disease mechanisms and facilitating biomarker discovery efforts.

## 1.6.2 TMT

The tandem mass tag (TMT) labeling approach has emerged as a powerful tool for quantitative proteomics, offering unique advantages and facing specific challenges when applied to FFPE tissue samples. This chapter delves into the technical intricacies, benefits, and limitations of employing the TMT labeling approach in FFPE tissue proteomics. TMT labeling provides a robust and multiplexed method for quantifying protein abundance in different types of complex biological samples, including FFPE tissues. By incorporating isobaric mass tags into peptides from different samples, TMT enables simultaneous quantification of multiple samples within a single mass spectrometry (MS) experiment. This multiplexing capability allows for efficient comparison of protein expression levels across various experimental conditions or disease states, enhancing the throughput and efficiency of proteomic analyses in FFPE tissues. Moreover, TMT labeling offers excellent precision, facilitating reliable comparisons of protein abundance across different samples. This quantitative precision is particularly valuable in FFPE tissue proteomics, where sample heterogeneity and variability in protein extraction efficiency can pose challenges to quantitative analyses. Additionally, TMT labeling enhances the dynamic range of protein quantification in FFPE tissue proteomics. By incorporating multiple channels within a single TMT experiment, researchers can quantify proteins spanning a broad range of abundances, from highly abundant to low-abundance species. This expanded dynamic range allows for the detection and quantification of both major signaling pathways and rare biomarkers within scarce FFPE tissues, providing comprehensive insights into disease mechanisms and potential therapeutic targets.<sup>162</sup> However, despite its numerous benefits, the TMT labeling approach also presents certain limitations when applied to FFPE tissue proteomics. One major challenge is the potential for incomplete labeling of peptides, particularly in FFPE samples with extensive protein cross-linking or modifications induced by formalin fixation. Incomplete labeling can result in inaccurate quantification and compromise the reliability of quantitative proteomic data derived from TMT experiments.<sup>163</sup> Furthermore, TMT labeling introduces additional complexity to data analysis and interpretation in FFPE tissue proteomics. The multiplexed nature of TMT experiments requires sophisticated computational methods for data processing, including accurate peptide identification, quantification, and normalization across different TMT channels.<sup>164</sup> Additionally, variability in labeling efficiency and differences in sample characteristics among TMT channels can introduce technical biases and confound the interpretation of quantitative results in FFPE tissue proteomics. In conclusion, while the TMT labeling approach offers numerous technical benefits for quantitative proteomic analysis of FFPE tissues, including multiplexing, quantitative accuracy, and expanded dynamic range, it also poses specific challenges related to labeling efficiency, data analysis, and interpretation. By carefully addressing these limitations and optimizing experimental protocols and data analysis pipelines, researchers can harness the full potential of TMT labeling in FFPE tissue proteomics to advance our understanding of disease mechanisms and identify novel biomarkers for clinical diagnosis and therapy. Optimized condition for limited sample material, like for single-cell proteomics is for example the usage of a carrier proteome in one TMT channel to boost peptide signals coming

from low abundant peptides.<sup>165</sup> Another step after TMT labeling, that can be optimized on-line and off-line is the high-pH fractionation which offers further potential for deeper insights into the proteome of scarce FFPE tissue proteomics.<sup>162</sup>

### 1.6.3 DIA

DIA is a mass spectrometric method that aims to overcome the limitations of traditional Data Dependent Acquisition (DDA) by collecting comprehensive MS data. Unlike DDA, which selects precursor ions for fragmentation based on their abundance, DIA simultaneously fragments all ions within a selected mass-to-charge ( $m/z$ ) range. This results in a more unbiased and inclusive dataset, capturing a broader spectrum of peptides and proteins. DIA improves reproducibility between runs because it does not rely on stochastic selection of ions, which is crucial for clinical applications and longitudinal studies where comparability between samples is essential. Additionally, DIA enables the detection of low-abundance proteins and post-translational modifications that might be missed by DDA, providing more accurate and consistent quantitation, which is particularly important for biomarker discovery and validation.<sup>10, 166</sup> The application of DIA to FFPE tissues has revolutionized proteomic analysis in clinical and research settings. Using DIA, all peptide ions within predefined  $m/z$  windows are fragmented and analyzed. Advanced instruments, such as the Orbitrap or Q-TOF, provide high-resolution data essential for distinguishing closely related peptides.<sup>167</sup> Bioinformatics tools like DIA-NN, Spectronaut™ and Skyline™ are employed to interpret the complex DIA datasets, allowing for accurate peptide identification and quantification.<sup>168-171</sup> One of the most promising applications of DIA in FFPE proteomics is the discovery of biomarkers for diseases such as cancer, for which DIA is extensively used in the last years.<sup>96, 141, 172-175</sup> DIA's comprehensive data collection allows for the identification of novel protein biomarkers that could be missed by other methods, providing insights into disease mechanisms, prognosis, and therapeutic targets. DIA has the potential to improve clinical diagnostics by providing detailed proteomic profiles from FFPE samples, leading to more accurate disease classification, personalized treatment plans, and monitoring of treatment efficacy. The vast archives of FFPE samples make them ideal for retrospective studies. DIA represents a significant advancement in the field of proteomics, particularly for the analysis of FFPE tissues. Its high reproducibility, sensitivity, and quantitative accuracy make it an invaluable tool for biomarker discovery, clinical diagnostics, and retrospective studies. As DIA technology continues to evolve, its application to FFPE tissue proteomics will likely expand, providing deeper insights into human health and disease.

## 2 Summary

Formalin fixation, introduced in 1893, remains a cornerstone of immunohistochemistry (IHC) and tissue preservation. Tissue banks worldwide house millions of samples annually, facilitating -omic research and biomarker discovery. Formalin fixation followed by paraffin embedding (FFPE) is the gold standard, offering high availability and precise morphology for microdissection. However, FFPE poses challenges for proteomic analysis due to crosslinking and low protein recovery.

Various extraction methods have been developed to overcome these hurdles and enable comparable results to fresh frozen tissue. Optimizing tissue collection and processing is crucial for translational research. While formalin fixation and paraffin embedding are well-established methods, challenges arise during tissue transfer and storage. Alternative preservation methods like vacuum preservation have been explored to mitigate formalin's toxicity. However, rapid formalin fixation preserves proteins effectively, while shock-freezing maintains DNA and RNA integrity. Standardization efforts by organizations like the NCI and EU aim to ensure quality control and auditing of biospecimen resources. Proper storage conditions are vital to prevent protein degradation and antigen loss. Qiagen's EXB plus buffer system addresses low protein recovery from long-term stored FFPE tissues, enhancing standardization and enabling large-scale proteomic studies for diverse diseases. Formalin fixation of proteins occurs via the formation of methylol adducts and subsequent crosslinking between amino groups and reactive CH groups. This mechanism maintains tissue morphology but poses challenges for subsequent protein and DNA/RNA analysis due to incomplete reversal of crosslinks. Formaldehyde crosslinking enables interaction studies like protein-protein and protein-DNA interactions but can lead to background noise and protein denaturation. While partially reversible, formaldehyde crosslinking may hinder the recovery of fully functional proteins and interfere with certain analyses, posing limitations in data interpretation. Shi and colleagues pioneered heat-induced antigen retrieval (HIAR) from over-fixed or long-term stored FFPE tissues in 1991, enhancing immunohistochemical staining for numerous antibodies. The method restored staining for partially formalin-sensitive antibodies like vimentin, proving its efficacy. Over the years, HIAR has become a standard in tissue preservation and immunohistochemical applications. For other analyses like Western Blot, SDS-PAGE, and proteomics, efficient protein extraction from FFPE tissues is crucial. Heating steps and buffer selection play vital roles in maximizing protein recovery. Various extraction protocols exist, with organic solvents like trifluoroethanol (TFE) showing promise. TFE-based extraction methods have shown increased peptide and protein identification compared to detergent-based methods, even for low sample amounts. Effective antigen retrieval also depends on pH, with a basic pH range recommended for optimal protein retrieval. The HIAR protocol, with its heating treatment, detergent-based buffer, and basic pH, is widely applicable and efficient for protein extraction from FFPE tissues. Shi et al. emphasized HIAR's simplicity and effectiveness, dubbing it the "gold key" for FFPE tissue analysis. For downstream processing a suitable proteomic sample preparation method for LC-MS/MS analysis is important for the overall success to achieve sensitive protein analysis. 2D gel electrophoresis, developed in 1975, separates complex protein mixtures based on their isoelectric point (pI) and molecular weight. It offers insights into protein expression levels, post-translational modifications, and isoforms. While once widely used for protein purification, other methods like FASP and Capture coaggregation are now favored for LC-MS/MS-based proteomic sample preparation. FASP efficiently purifies proteins for mass spectrometry analysis, but commercial filters may introduce unexpected side reactions and contaminants. Capture coaggregation, or SP3, is a robust method for selective enrichment of target proteins or complexes, offering simplicity, efficiency, and minimal sample loss. It's

particularly useful for scarce sample proteomics, outperforming other methods like FASP in low sample amounts. Last but not least, the most suitable MS acquisition method is important for the success in proteomic analysis within low sample amounts coming from scarce FFPE tissue samples. Data-dependent acquisition (DDA) is widely used in formalin-fixed paraffin-embedded (FFPE) tissue proteomics for its ability to comprehensively cover the proteome with minimal cost. While it efficiently selects precursor ions for fragmentation, reproducibility across replicates can be challenging. DDA can miss low-abundance proteins and suffers from dynamic range compression. Tandem mass tag (TMT) labeling offers precise and multiplexed protein quantification in FFPE tissues, but incomplete labeling and complex data analysis are concerns. Despite challenges, TMT enables broad dynamic range quantification and identification of rare biomarkers. Data-independent acquisition (DIA) overcomes DDA limitations by fragmenting all ions within a selected mass-to-charge range, resulting in unbiased and inclusive datasets. DIA improves reproducibility, detects low-abundance proteins, and facilitates biomarker discovery in FFPE tissues. Its application in clinical diagnostics and retrospective studies holds promise for advancing our understanding of disease mechanisms.

### 3 References

- (1) Fox, C. H.; Johnson, F. B.; Whiting, J.; Roller, P. P. Formaldehyde fixation. *Journal of Histochemistry & Cytochemistry* **1985**, *33* (8), 845-853.
- (2) Stanta, G. *Archive Tissues*. 2011; pp 3-4.
- (3) Bevilacqua, G.; Bosman, F.; Dassel, T.; Höfler, H.; Janin, A.; Langer, R.; Larsimont, D.; Morente, M. M.; Riegman, P.; Schirmacher, P.; et al. The role of the pathologist in tissue banking: European Consensus Expert Group Report. *Virchows Archiv* **2010**, *456* (4), 449-454.
- (4) Gustafsson, O. J. R.; Arentz, G.; Hoffmann, P. Proteomic developments in the analysis of formalin-fixed tissue. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2015**, *1854* (6), 559-580.
- (5) Ilyin, S. E.; Belkowski, S. M.; Plata-Salamán, C. R. Biomarker discovery and validation: technologies and integrative approaches. *Trends in Biotechnology* **2004**, *22* (8), 411-416.
- (6) Palmer-Toy, D. E.; Krastins, B.; Sarracino, D. A.; Nadol, J. B.; Merchant, S. N. Efficient Method for the Proteomic Analysis of Fixed and Embedded Tissues. *Journal of Proteome Research* **2005**, *4* (6), 2404-2411.
- (7) Ikeda, K.; Monden, T.; Kanoh, T.; Tsujie, M.; Izawa, H.; Haba, A.; Ohnishi, T.; Sekimoto, M.; Tomita, N.; Shiozaki, H. Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections. *Journal of Histochemistry & Cytochemistry* **1998**, *46* (3), 397-403.
- (8) Wolff, C.; Schott, C.; Porschewski, P.; Reischauer, B.; Becker, K.-F. Successful Protein Extraction from Over-Fixed and Long-Term Stored Formalin-Fixed Tissues. *Plos One* **2011**, *6* (1), e16353.
- (9) Addis, M. F.; Tanca, A.; Pagnozzi, D.; Crobu, S.; Fanciulli, G.; Cossu-Rocca, P.; Uzzau, S. Generation of high-quality protein extracts from formalin-fixed, paraffin-embedded tissues. *PROTEOMICS* **2009**, *9* (15), 3815-3823.
- (10) Bravo, S.; Garcia-Vence, M.; Chantada-Vazquez, M. d. P.; Sosa-Fajardo, A.; Agra, R.; Barcia de la Iglesia, A.; Otero-Glez, A.; Garcia-Gonzalez, M. A.; Cameselle-Teijeiro, J.; Nuñez, C. Protein extraction from FFPE kidney tissue samples: A review of the literature and characterization of techniques. *Frontiers in Medicine* **2021**, *8*, 553.
- (11) Föll, M. C.; Fahrner, M.; Oria, V. O.; Kühs, M.; Biniossek, M. L.; Werner, M.; Bronsert, P.; Schilling, O. Reproducible proteomics sample preparation for single FFPE tissue slices using acid-labile surfactant and direct trypsinization. *Clinical Proteomics* **2018**, *15* (1).
- (12) Shi, S.-R.; Taylor, C. R.; Fowler, C. B.; Mason, J. T. Complete solubilization of formalin-fixed, paraffin-embedded tissue may improve proteomic studies. *PROTEOMICS – Clinical Applications* **2013**, *7* (3-4), 264-272.

- (13) Stanta, G. *Guidelines for molecular analysis in archive tissues*; Springer Science & Business Media, 2011.
- (14) Bussolati, G.; Chiusa, L.; Cimino, A.; D'Armento, G. Tissue transfer to pathology labs: under vacuum is the safe alternative to formalin. *Virchows Archiv* **2008**, *452* (2), 229-231.
- (15) Blow, N. Tissue issues. *Nature* **2007**, *448* (7156), 959-960.
- (16) Hallmans, G.; Vaught, J. B. Best practices for establishing a biobank. In *Methods in Biobanking*, Springer, 2011; pp 241-260.
- (17) Pitt, K. E.; Campbell, L. D.; Skubitz, A. P.; Aamodt, R. L.; Anouna, A.; Baird, P.; Beck, J. C.; Bledsoe, M. J.; De Souza, Y.; Grizzle, W. E. Best practices for repositories I: Collection, storage, and retrieval of human biological materials for research. *Cell Preservation Technology* **2005**, *3* (1), 5-48.
- (18) Xie, R.; Chung, J.-Y.; Ylaya, K.; Williams, R. L.; Guerrero, N.; Nakatsuka, N.; Badie, C.; Hewitt, S. M. Factors Influencing the Degradation of Archival Formalin-Fixed Paraffin-Embedded Tissue Sections. *Journal of Histochemistry & Cytochemistry* **2011**, *59* (4), 356-365.
- (19) Fraenkel-Conrat, H.; Brandon, B. A.; Olcott, H. S. The reaction of formaldehyde with proteins: IV. Participation of indole groups. Gramicidin. *Journal of Biological Chemistry* **1947**, *168* (1), 99-118.
- (20) Fraenkel-Conrat, H.; Olcott, H. S. Reaction of formaldehyde with proteins: VI. Cross-linking of amino groups with phenol, imidazole, or indole groups. *Journal of Biological Chemistry* **1948**, *174* (3), 827-843.
- (21) Hoffman, E. A.; Frey, B. L.; Smith, L. M.; Auble, D. T. Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J Biol Chem* **2015**, *290* (44), 26404-26411.
- (22) Magdeldin, S.; Yamamoto, T. Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues. *PROTEOMICS* **2012**, *12* (7), 1045-1058.
- (23) Sutherland, B. W.; Toews, J.; Kast, J. Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions. *Journal of Mass Spectrometry* **2008**, *43* (6), 699-715.
- (24) Orlando, V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends in biochemical sciences* **2000**, *25* (3), 99-104.
- (25) Shi, S. R.; Key, M. E.; Kalra, K. L. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. *Journal of Histochemistry & Cytochemistry* **1991**, *39* (6), 741-748.
- (26) Jones, M.; Banks, P.; Caron, B. Transition metal salts as adjuncts to formalin for tissue fixation. *Lab Invest* **1981**, *44* (32).
- (27) Shi, S.-R.; Shi, Y.; Taylor, C. R. Antigen retrieval immunohistochemistry: review and future prospects in research and diagnosis over two decades. *Journal of Histochemistry & Cytochemistry* **2011**, *59* (1), 13-32.
- (28) Chu, W.-S.; Liang, Q.; Liu, J.; Wei, M. Q.; Winters, M.; Liotta, L.; Sandberg, G.; Gong, M. A nondestructive molecule extraction method allowing morphological and molecular analyses using a single tissue section. *Laboratory investigation* **2005**, *85* (11), 1416-1428.
- (29) Prieto, D. A.; Hood, B. L.; Darfler, M. M.; Guiel, T. G.; Lucas, D. A.; Conrads, T. P.; Veenstra, T. D.; Krizman, D. B. Liquid Tissue™: proteomic profiling of formalin-fixed tissues. *BioTechniques* **2005**, *38* (6S), S32-S35.
- (30) Yamashita, S.; Okada, Y. Mechanisms of heat-induced antigen retrieval: analyses in vitro employing SDS-PAGE and immunohistochemistry. *Journal of Histochemistry & Cytochemistry* **2005**, *53* (1), 13-21.
- (31) Fowler, C. B.; O'leary, T. J.; Mason, J. T. Techniques of protein extraction from FFPE tissue/cells for mass spectrometry. *Antigen Retrieval Immunohistochemistry Based Research and Diagnostics* **2010**, 333-346.
- (32) Jiang, X.; Jiang, X.; Feng, S.; Tian, R.; Ye, M.; Zou, H. Development of Efficient Protein Extraction Methods for Shotgun Proteome Analysis of Formalin-Fixed Tissues. *Journal of Proteome Research* **2007**, *6* (3), 1038-1047.
- (33) Proc, J. L.; Kuzyk, M. A.; Hardie, D. B.; Yang, J.; Smith, D. S.; Jackson, A. M.; Parker, C. E.; Borchers, C. H. A Quantitative Study of the Effects of Chaotropic Agents, Surfactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin. *Journal of Proteome Research* **2010**, *9* (10), 5422-5437.
- (34) Hervey; Strader, M. B.; Hurst, G. B. Comparison of Digestion Protocols for Microgram Quantities of Enriched Protein Samples. *Journal of Proteome Research* **2007**, *6* (8), 3054-3061.

- (35) Poulsen, J. W.; Madsen, C. T.; Young, C.; Poulsen, F. M.; Nielsen, M. L. Using Guanidine-Hydrochloride for Fast and Efficient Protein Digestion and Single-step Affinity-purification Mass Spectrometry. *Journal of Proteome Research* **2013**, *12* (2), 1020-1030.
- (36) Ferro, M.; Seigneurin-Berny, D.; Rolland, N.; Chapel, A.; Salvi, D.; Garin, J.; Joyard, J. Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins. *ELECTROPHORESIS* **2000**, *21* (16), 3517-3526.
- (37) Chertov, O.; Biragyn, A.; Kwak, L. W.; Simpson, J. T.; Boronina, T.; Hoang, V. M.; Prieto, D. A.; Conrads, T. P.; Veenstra, T. D.; Fisher, R. J. Organic solvent extraction of proteins and peptides from serum as an effective sample preparation for detection and identification of biomarkers by mass spectrometry. *PROTEOMICS* **2004**, *4* (4), 1195-1203.
- (38) Wang, H.; Qian, W.-J.; Mottaz, H. M.; Clauss, T. R. W.; Anderson, D. J.; Moore, R. J.; Camp, D. G.; Khan, A. H.; Sforza, D. M.; Pallavicini, M.; et al. Development and Evaluation of a Micro- and Nanoscale Proteomic Sample Preparation Method. *Journal of Proteome Research* **2005**, *4* (6), 2397-2403.
- (39) Zhang, H.; Lin, Q.; Ponnusamy, S.; Kothandaraman, N.; Lim, T. K.; Zhao, C.; Kit, H. S.; Arijit, B.; Rauff, M.; Hew, C.-L.; et al. Differential recovery of membrane proteins after extraction by aqueous methanol and trifluoroethanol. *PROTEOMICS* **2007**, *7* (10), 1654-1663.
- (40) Wigley, W. C.; Vijayakumar, S.; Jones, J. D.; Slaughter, C.; Thomas, P. J. Transmembrane Domain of Cystic Fibrosis Transmembrane Conductance Regulator: Design, Characterization, and Secondary Structure of Synthetic Peptides m1-m6. *Biochemistry* **1998**, *37* (3), 844-853.
- (41) Blonder, J.; Goshe, M. B.; Moore, R. J.; Pasa-Tolic, L.; Masselon, C. D.; Lipton, M. S.; Smith, R. D. Enrichment of Integral Membrane Proteins for Proteomic Analysis Using Liquid Chromatography-Tandem Mass Spectrometry. *Journal of Proteome Research* **2002**, *1* (4), 351-360.
- (42) Goshe, M. B.; Blonder, J.; Smith, R. D. Affinity Labeling of Highly Hydrophobic Integral Membrane Proteins for Proteome-Wide Analysis. *Journal of Proteome Research* **2003**, *2* (2), 153-161.
- (43) Deshusses, J. M. P.; Burgess, J. A.; Scherl, A.; Wenger, Y.; Walter, N.; Converset, V.; Paesano, S.; Corthals, G. L.; Hochstrasser, D. F.; Sanchez, J.-C. Exploitation of specific properties of trifluoroethanol for extraction and separation of membrane proteins. *PROTEOMICS* **2003**, *3* (8), 1418-1424.
- (44) Zuobi-Hasona, K.; Crowley, P. J.; Hasona, A.; Bleiweis, A. S.; Brady, L. J. Solubilization of cellular membrane proteins from *Streptococcus mutans* for two-dimensional gel electrophoresis. *ELECTROPHORESIS* **2005**, *26* (6), 1200-1205.
- (45) Tian, Y.; Gurley, K.; Meany, D. L.; Kemp, C. J.; Zhang, H. N-linked glycoproteomic analysis of formalin-fixed and paraffin-embedded tissues. *Journal of proteome research* **2009**, *8* (4), 1657-1662.
- (46) Sprung, R. W., Jr.; Brock, J. W. C.; Tanksley, J. P.; Li, M.; Washington, M. K.; Slebos, R. J. C.; Liebler, D. C. Equivalence of Protein Inventories Obtained from Formalin-fixed Paraffin-embedded and Frozen Tissue in Multidimensional Liquid Chromatography-Tandem Mass Spectrometry Shotgun Proteomic Analysis. *Molecular & Cellular Proteomics* **2009**, *8* (8), 1988-1998.
- (47) Coscia, F.; Doll, S.; Bech, J. M.; Schweizer, L.; Mund, A.; Lengyel, E.; Lindebjerg, J.; Madsen, G. I.; Moreira, J. M.; Mann, M. A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis. *J Pathol* **2020**, *251* (1), 100-112.
- (48) Shi, S. R.; Imam, S. A.; Young, L.; Cote, R. J.; Taylor, C. R. Antigen retrieval immunohistochemistry under the influence of pH using monoclonal antibodies. *Journal of Histochemistry & Cytochemistry* **1995**, *43* (2), 193-201.
- (49) Shi, S.-R.; Shi, Y.; Taylor, C. R.; Gu, J. New Dimensions of Antigen Retrieval Technique: 28 Years of Development, Practice, and Expansion. *Applied Immunohistochemistry & Molecular Morphology* **2019**, *27* (10), 715-721.
- (50) O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry* **1975**, *250* (10), 4007-4021.
- (51) Sitek, B.; Sipos, B.; Alkatout, I.; Poschmann, G.; Stephan, C.; Schulenburg, T.; Marcus, K.; Luttgies, J.; Dittert, D. D.; Baretton, G.; et al. Analysis of the pancreatic tumor progression by a quantitative proteomic approach and immunohistochemical validation. *J Proteome Res* **2009**, *8* (4), 1647-1656.
- (52) Sitek, B.; Luttgies, J.; Marcus, K.; Kloppel, G.; Schmiegel, W.; Meyer, H. E.; Hahn, S. A.; Stuhler, K. Application of fluorescence difference gel electrophoresis saturation labelling for the

- analysis of microdissected precursor lesions of pancreatic ductal adenocarcinoma. *Proteomics* **2005**, *5* (10), 2665-2679.
- (53) Rabilloud, T.; Lelong, C. Two-dimensional gel electrophoresis in proteomics: A tutorial. *Journal of Proteomics* **2011**, *74* (10), 1829-1841.
- (54) Rabilloud, T. Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains. *PROTEOMICS* **2002**, *2* (1), 3-10.
- (55) Lee, P. Y.; Saraygord-Afshari, N.; Low, T. Y. The evolution of two-dimensional gel electrophoresis - from proteomics to emerging alternative applications. *Journal of Chromatography A* **2020**, *1615*, 460763.
- (56) Zhang, Z.; Dubiak, K. M.; Huber, P. W.; Dovichi, N. J. Miniaturized Filter-Aided Sample Preparation (MICRO-FASP) Method for High Throughput, Ultrasensitive Proteomics Sample Preparation Reveals Proteome Asymmetry in *Xenopus laevis* Embryos. *Analytical Chemistry* **2020**, *92* (7), 5554-5560.
- (57) Wiśniewski, J. R.; Zougman, A.; Mann, M. Combination of FASP and StageTip-Based Fractionation Allows In-Depth Analysis of the Hippocampal Membrane Proteome. *Journal of Proteome Research* **2009**, *8* (12), 5674-5678.
- (58) Wiśniewski, J. R.; Ostasiewicz, P.; Mann, M. High Recovery FASP Applied to the Proteomic Analysis of Microdissected Formalin Fixed Paraffin Embedded Cancer Tissues Retrieves Known Colon Cancer Markers. *Journal of Proteome Research* **2011**, *10* (7), 3040-3049.
- (59) Wiśniewski, J. R. Quantitative Evaluation of Filter Aided Sample Preparation (FASP) and Multienzyme Digestion FASP Protocols. *Analytical Chemistry* **2016**, *88* (10), 5438-5443.
- (60) Ni, M.-w.; Wang, L.; Chen, W.; Mou, H.-z.; Zhou, J.; Zheng, Z.-g. Modified filter-aided sample preparation (FASP) method increases peptide and protein identifications for shotgun proteomics. *Rapid Communications in Mass Spectrometry* **2017**, *31* (2), 171-178.
- (61) Tang, L.; Wu, Z.; Wang, J.; Zhang, X. Formaldehyde Derivatization: An Unexpected Side Reaction During Filter-Aided Sample Preparation. *Analytical Chemistry* **2020**, *92* (18), 12120-12125.
- (62) Gomez-Zepeda, D.; Michna, T.; Ziesmann, T.; Distler, U.; Tenzer, S. HowDirty: An R package to evaluate molecular contaminants in LC-MS experiments. *PROTEOMICS* **2023**, *n/a* (n/a), 2300134.
- (63) Ostasiewicz, P.; Zielinska, D. F.; Mann, M.; Wiśniewski, J. R. Proteome, Phosphoproteome, and N-Glycoproteome Are Quantitatively Preserved in Formalin-Fixed Paraffin-Embedded Tissue and Analyzable by High-Resolution Mass Spectrometry. *Journal of Proteome Research* **2010**, *9* (7), 3688-3700.
- (64) Wiśniewski, J. R.; Zielinska, D. F.; Mann, M. Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method. *Analytical Biochemistry* **2011**, *410* (2), 307-309.
- (65) Hughes, C. S.; Foehr, S.; Garfield, D. A.; Furlong, E. E.; Steinmetz, L. M.; Krijgsveld, J. Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular Systems Biology* **2014**, *10* (10), 757.
- (66) Hughes, C. S.; Moggridge, S.; Müller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols* **2019**, *14* (1), 68-85.
- (67) Sielaff, M.; Kuharev, J.; Bohn, T.; Hahlbrock, J.; Bopp, T.; Tenzer, S.; Distler, U. Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *Journal of Proteome Research* **2017**, *16* (11), 4060-4072.
- (68) Kassem, S.; van der Pan, K.; de Jager, A. L.; Naber, B. A.; de Laat, I. F.; Louis, A.; van Dongen, J. J.; Teodosio, C.; Díez, P. Proteomics for Low Cell Numbers: How to Optimize the Sample Preparation Workflow for Mass Spectrometry Analysis. *Journal of Proteome Research* **2021**.
- (69) Patel, V.; Hood, B. L.; Molinolo, A. A.; Lee, N. H.; Conrads, T. P.; Braisted, J. C.; Krizman, D. B.; Veenstra, T. D.; Gutkind, J. S. Proteomic Analysis of Laser-Captured Paraffin-Embedded Tissues: A Molecular Portrait of Head and Neck Cancer Progression. *Clinical Cancer Research* **2008**, *14* (4), 1002-1014.
- (70) Craven, R. A.; Cairns, D. A.; Zougman, A.; Harnden, P.; Selby, P. J.; Banks, R. E. Proteomic analysis of formalin-fixed paraffin-embedded renal tissue samples by label-free MS: Assessment of overall technical variability and the impact of block age. *PROTEOMICS – Clinical Applications* **2013**, *7* (3-4), 273-282.

- (71) Johann, D. J.; Rodriguez-Canales, J.; Mukherjee, S.; Prieto, D. A.; Hanson, J. C.; Emmert-Buck, M.; Blonder, J. Approaching Solid Tumor Heterogeneity on a Cellular Basis by Tissue Proteomics Using Laser Capture Microdissection and Biological Mass Spectrometry. *Journal of Proteome Research* **2009**, *8* (5), 2310-2318.
- (72) Umar, A.; Kang, H.; Timmermans, A. M.; Look, M. P.; Meijer-van Gelder, M. E.; den Bakker, M. A.; Jaitly, N.; Martens, J. W. M.; Luider, T. M.; Foekens, J. A.; et al. Identification of a Putative Protein Profile Associated with Tamoxifen Therapy Resistance in Breast Cancer\*. *Molecular & Cellular Proteomics* **2009**, *8* (6), 1278-1294.
- (73) Longuespée, R.; Fléron, M.; Pottier, C.; Quesada-Calvo, F.; Meuwis, M.-A.; Baiwir, D.; Smargiasso, N.; Mazzucchelli, G.; De Pauw-Gillet, M.-C.; Delvenne, P. Tissue proteomics for the next decade? Towards a molecular dimension in histology. *Omic: a journal of integrative biology* **2014**, *18* (9), 539-552.
- (74) Djuric, U.; Rodrigues, D. C.; Batruch, I.; Ellis, J.; Shannon, P.; Diamandis, P. Spatiotemporal proteomic profiling of human cerebral development. *Molecular & Cellular Proteomics* **2017**, *16* (9), 1548-1562.
- (75) Griesser, E.; Wyatt, H.; Ten Have, S.; Stierstorfer, B.; Lenter, M.; Lamond, A. I. Quantitative Profiling of the Human Substantia Nigra Proteome from Laser-capture Microdissected FFPE Tissue. *Mol Cell Proteomics* **2020**, *19* (5), 839-851.
- (76) Zecha, J.; Satpathy, S.; Kanashova, T.; Avanesian, S. C.; Kane, M. H.; Clauser, K. R.; Mertins, P.; Carr, S. A.; Kuster, B. TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach\*[S]. *Molecular & Cellular Proteomics* **2019**, *18* (7), 1468-1478.
- (77) Pappireddi, N.; Martin, L.; Wühr, M. A review on quantitative multiplexed proteomics. *Chembiochem* **2019**, *20* (10), 1210-1224.
- (78) Cheung, T. K.; Lee, C.-Y.; Bayer, F. P.; McCoy, A.; Kuster, B.; Rose, C. M. Defining the carrier proteome limit for single-cell proteomics. *Nature Methods* **2021**, *18* (1), 76-83.
- (79) Ye, Z.; Mao, Y.; Clausen, H.; Vakhrushev, S. Y. Glyco-DIA: a method for quantitative O-glycoproteomics with in silico-boosted glycopeptide libraries. *Nature Methods* **2019**, *16* (9), 902-910.
- (80) Bekker-Jensen, D. B.; Bernhardt, O. M.; Högberg, A.; Martinez-Val, A.; Verbeke, L.; Gandhi, T.; Kelstrup, C. D.; Reiter, L.; Olsen, J. V. Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nature Communications* **2020**, *11* (1), 787.
- (81) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis\*. *Molecular & Cellular Proteomics* **2012**, *11* (6), O111.016717.
- (82) Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods* **2020**, *17* (1), 41-44.
- (83) Gandhi, T.; Verbeke, L.; Bernhardt, O. M.; Muntel, J.; Müller, S.; Bruderer, R. M.; and Lukas Reiter, Y. X. Best of Both Worlds: Spectronaut Combines the Depth of Resource with iRT-Precision of Project Specific Data.
- (84) Pino, L. K.; Searle, B. C.; Bollinger, J. G.; Nunn, B.; MacLean, B.; MacCoss, M. J. The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrometry Reviews* **2020**, *39* (3), 229-244.
- (85) Zhang, F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T. Data-independent acquisition mass spectrometry-based proteomics and software tools: a glimpse in 2020. *Proteomics* **2020**, *20* (17-18), 1900276.
- (86) Kim, Y. J.; Sweet, S. M. M.; Egertson, J. D.; Sedgewick, A. J.; Woo, S.; Liao, W.-I.; Merrihew, G. E.; Searle, B. C.; Vaske, C.; Heaton, R.; et al. Data-Independent Acquisition Mass Spectrometry To Quantify Protein Levels in FFPE Tumor Biopsies for Molecular Diagnostics. *Journal of Proteome Research* **2019**, *18* (1), 426-435.
- (87) Stanojevic, A.; Samiotaki, M.; Lygirou, V.; Marinkovic, M.; Nikolic, V.; Stojanovic-Rundic, S.; Jankovic, R.; Vlahou, A.; Panayotou, G.; Fijneman, R. J. A.; et al. Data independent acquisition mass spectrometry (DIA-MS) analysis of FFPE rectal cancer samples offers in depth proteomics characterization of response to neoadjuvant chemoradiotherapy. *medRxiv* **2023**, 2023.2005.2012.23289671.
- (88) Weke, K.; Kote, S.; Faktor, J.; Al Shboul, S.; Uwugiaren, N.; Brennan, P. M.; Goodlett, D. R.; Hupp, T. R.; Dapic, I. DIA-MS proteome analysis of formalin-fixed paraffin-embedded glioblastoma tissues. *Analytica Chimica Acta* **2022**, *1204*, 339695.

- (89) Keam, S. P.; Gulati, T.; Gamell, C.; Caramia, F.; Huang, C.; Schittenhelm, R. B.; Kleifeld, O.; Neeson, P. J.; Haupt, Y.; Williams, S. G. Exploring the oncoproteomic response of human prostate cancer to therapeutic radiation using data-independent acquisition (DIA) mass spectrometry. *The Prostate* **2018**, *78* (8), 563-575.
- (90) Pongratanakul, P.; Bremmer, F.; Pauls, S.; Poschmann, G.; Kresbach, C.; Parmaksiz, F.; Skowron, M. A.; Fuß, J.; Stephan, A.; Paffenholz, P.; et al. Assessing the risk to develop a growing teratoma syndrome based on molecular and epigenetic subtyping as well as novel secreted biomarkers. *Cancer Letters* **2024**, 216673.

## Author contributions to Manuscript 1

### **Uncover the archived treasures – Challenges and opportunities in FFPE tissue proteomics**

Stella Pauls

Prepared figures and tables, made literature research, wrote and edited the manuscript.

Anja Stefanski

Conceived and supervised the study.

Kai Stühler

Conceived and supervised the study, acquired funding, reviewed/edited the manuscript.



## 6 Manuscript 2:

Proteome analysis of microdissected pancreas tissue by improved HIAR protocol



# Proteome analysis of microdissected pancreas tissue by improved HIAR protocol

Stella Pauls<sup>1</sup>, Anja Stefanski<sup>2</sup>, Friederike Opitz<sup>3</sup>, Sandra Biskup<sup>3</sup>, Irene Esposito<sup>3</sup>, Gereon Poschmann<sup>1</sup>, Thomas Lenz<sup>1</sup>, Kai Stühler<sup>1,2</sup>

<sup>1</sup>Proteome Research, Institut für Molekulare Medizin, Medizinische Fakultät, Heinrich-Heine-Universität, Universitätsstraße 1, D-40225 Düsseldorf

<sup>2</sup>Molecular proteomics laboratory (MPL), Institut für Molekulare Medizin, Medizinische Fakultät, Heinrich-Heine-Universität, Universitätsstraße 1, D-40225 Düsseldorf

<sup>3</sup>Institut für Pathologie, Universitätsklinikum Düsseldorf, Medizinische Fakultät, Heinrich-Heine-Universität, Moorenstr. 5, D-40225 Düsseldorf

## Abstract:

Proteomic analysis of different tissue samples is important for the detection of proteins that are involved in several different diseases, to find new disease-specific biomarker to improve early diagnostics or to find therapeutical targets to give therapeutical options for disease-treatment. In this study we optimized and applied sensitive methods combining cell-specific microdissection of pancreatic FFPE tissue material and quantitative, label-free proteomics to improve protein identification and quantification to gain deeper understanding of biological mechanisms underlying disease progression. The long-term known HIAR protocol for the efficient tissue lysis and protein extraction from FFPE tissues is improved by adding freeze-thaw steps and by extension of extraction rounds from two to three rounds and using small extraction volumes. Using the CaproMag™-SP3 for protein cleanup and digestion enables highly reproducible and sensitive results in protein identification. By using a data independent acquisition (DIA) method for LC-MS/MS analysis the goal is a deeper insight into low abundant proteins and was achieved by protein-ID increase of 70% in comparison to data dependent acquisition (DDA). Besides Protein-ID increase, DIA enables proteomic data with low numbers of missing values and highly reproducible quantification results with low scattering. This improved workflow is a simple and fast approach to reach standardization within scarce FFPE tissue proteomics.

## 1 Introduction

Formalin-Fixed Paraffin-Embedded (FFPE) tissues have long been regarded as the gold standard for preserving clinical tissue samples.<sup>116, 121</sup> Their widespread adoption in pathology laboratories worldwide stems from their ability to maintain tissue morphology and cellular architecture over long periods, providing invaluable resources for retrospective studies and diagnostic purposes. Despite their advantages, the utility of FFPE tissues in molecular analyses, particularly proteomics, has been hampered by the challenges posed by formalin-induced crosslinking and the resultant low protein recovery. The intricate crosslinking chemistry introduced by formalin fixation renders proteins less accessible for extraction and analysis, leading to suboptimal results

and limited insights into the tissues molecular landscape.<sup>104, 176</sup> Although routinely utilized in pathology departments for tissue archiving, the formalin fixation process, involving cross-linking with nucleic acids, polysaccharides, and protein side chains, poses challenges for protein and RNA/DNA extraction.<sup>99, 116, 176, 177</sup> Reversal of these cross-links, often achieved through heating ("heat induced antigen retrieval"), is necessary for subsequent biochemical analysis. However, the pressing need to unlock the proteomic information contained within FFPE samples has encouraged the development of several protocols aimed at overcoming these challenges.<sup>8, 102-106</sup> In addition to heating, diverse extraction buffers containing Tris, SDS, and other detergents such as RapiGest, chaotropes like guanidine-HCl, and organic solvents like trifluoroethanol and acetonitrile are commonly used to solubilize proteins from FFPE tissues.<sup>99, 105, 126, 132, 178</sup> Tissue sections, being a heterogeneous mix of various cell populations with differing cell numbers, present challenges for protein analysis on distinct cell types. Laser-capture microdissection (LCM) emerges as a powerful tool for visualizing and isolating specific cell subpopulations within heterogeneous tissue while maintaining morphological context. LCM facilitates cell enrichment and enables protein expression analysis in targeted cell subsets. Protein extraction and proteomic sample preparation workflows for FFPE tissue specimens must be optimized for maximum sensitivity to enable analysis on low cell counts. In this study, we embark on a comprehensive comparison and development endeavor to enhance the proteomic analysis of FFPE tissues. Building upon existing methodologies, such as the heat-induced antigen retrieval protocol (HIAR) and leveraging advancements in analytical techniques, we focus our efforts on refining and improving upon current protocols to achieve higher protein recovery and quality from FFPE samples. Our aim is to address the limitations of existing approaches and establish a robust method that is compatible with the unique properties of FFPE tissues with high sensitivity for the analysis of limited FFPE tissue material, such as for microdissected FFPE cell subsets.

## 4 Material and Methods

### 4.1 Sample collection

Tissue biopsies were formalin-fixed and paraffin embedded, morphomolecular characterized and microdissected by the Institute of pathology of the Universitätsklinikum Düsseldorf.

### 4.2 Sample preparation for LC-MS/MS analysis

A modified FFPE tissue lysis protocol of Ikeda and colleagues was applied.<sup>102</sup> Microdissected FFPE tissue pieces were transferred in methanolic suspension into PCR-strips, were centrifuged and the methanolic supernatant was removed followed by air-drying of the residual solvent. The remaining tissue was resuspended in 60  $\mu$ L lysis buffer (300 mM TRIS/HCl, 2% SDS, pH 8.0), shock-frozen in liquid nitrogen and immediately heated for 25 min at 99°C and 350 rpm. For complete lysis, samples were ultrasonicated on ice for 20 min with 30 s on/off cycles and then shook for 2 h at 80°C and 500 rpm, followed by a second ultrasonication step. After centrifugation, the supernatant was transferred into a new 0.5 mL Lobind Eppendorf tubes and the pellet was

resuspended in 60  $\mu\text{L}$  buffer for a second extraction (resp. third), similar to the first. The resulting supernatants were combined and a modified magnetic bead-based sample preparation protocol according to Hughes and colleagues were applied.<sup>159</sup> Briefly, 50  $\mu\text{L}$  per sample lysate were reduced by adding 1.25  $\mu\text{L}$  300 mM DTT (dithiothreitol) and shaking for 20 min at 56°C and 1000 rpm, followed by alkylation with the addition of 6.67  $\mu\text{L}$  100 mM IAA and incubation for 15 min in the dark. A 20  $\mu\text{g}/\mu\text{L}$  bead stock of 1:1 Sera-Mag SpeedBeads was freshly prepared and 1.25  $\mu\text{L}$  were added to each sample. Afterwards, for protein aggregation capture 160  $\mu\text{L}$  ethanol was added to a final concentration of 73% and incubated for 15 min at 20°C. After three rinsing steps with 80% EtOH and one rinsing step with 100% ACN, beads were resuspended in 20  $\mu\text{L}$  of 50 mM TEAB buffer and digested with final 0.05  $\mu\text{g}$  trypsin at 37°C and 1,000 rpm overnight. Extra-digestion was carried out by adding 2  $\mu\text{L}$  of 0.025  $\mu\text{g}/\mu\text{L}$  trypsin stock and shaking at 37°C and 1000 rpm for 4 h. The supernatants were collected and the remaining solvent was removed by evaporation. Samples were

### 4.3 Liquid chromatography-mass spectrometry (LC–MS/MS) analysis

For the LC-MS acquisition an Orbitrap Fusion Lumos Tribrid Mass Spectrometer with FAIMS (Thermo Fisher Scientific) coupled to an Ultimate 3000 Rapid Separation liquid chromatography system (Thermo Fisher Scientific, Idstein, Germany) equipped with an Acclaim PepMap 100 C18 column (75  $\mu\text{m}$  inner diameter, 25 cm length, 2 mm particle size from Thermo Fisher Scientific) as separation column and an Acclaim PepMap 100 C18 column (75  $\mu\text{m}$  inner diameter, 2 cm length, 2 mm particle size from Thermo Fisher Scientific) as trap column. A LC-gradient of 180 min was applied and the mass spectrometer operated in positive data independent acquisition mode with a scan range of 380 – 985  $m/z$  at a resolution of 60,000 for the MS1 and a scan range of 145 – 1450  $m/z$  with a defined isolation window of 10  $m/z$  with an overlap of 1  $m/z$  and a resolution of 15000 for the MS2. FAIMS operated with a carrier gas flow of 4.5 L/min and the compensation voltage (CV) was set to -50 V. The capillary temperature was set to 275°C, the source voltage to 2.0 kV, the normalized AGC target was set to 200% and the maximum injection time was 40 ms. HCD collision energy was set to 30%.

### 4.4 MS data analysis

#### 4.4.1 Proteomics

Data analysis was carried out with DIA-NN (version 1.8.1, by Demichev, Ralser and Lilley labs, available on <https://github.com/vdemichev/DiaNN>).<sup>168</sup> All RAW files were searched against the human proteome UniProt KB dataset (UP000005640, download on 12.01.2023, 81837 entries) and the Maxquant Contaminant database (download on 03.05.2022), using the deep learning tool to generate an *in silico* spectral library, which is implemented in DIA-NN. The digestion enzyme was set to trypsin, the maximum number of missed cleavages was set to two and the peptide length was 7 – 30 amino acids. Mass accuracy was optimized by DIA-NN using the first run in the experiment. As variable modifications were methionine oxidation, N-terminal methionine loss and methylation of lysines (only for FFPE tissue samples) defined. Fixed modification was

carbamidomethylation of cysteines. All samples of the study were analyzed in a match between run (MBR) search.

## 5 Results and Discussion

### 5.1 Sample collection

Samples used for the optimization process are microdissected from tissue biopsies coming from pancreatic tissue. The microdissected area is 2.5 mm<sup>2</sup> or 0.5 mm<sup>2</sup> of the remaining healthy tissue nearby a pancreatic precursor lesion of PDAC. All samples are laser microdissected using a Zeiss PALM MicroBeam™ and collected into PCR tube caps. The microdissection is performed in the institute of pathology of the Uniklinikum Düsseldorf (UKD). All samples are stored at -80 °C until tissue lysis.

### 5.2 Comparison of analysis platforms for scarce FFPE tissue proteomics

In 1991, the first FFPE tissue disruption protocol was introduced by Shi et al., in which reversion of crosslinks from formalin fixation is induced by heating treatments.<sup>119</sup> Since then, the tissue disruption protocol is optimized from several working groups to maximize complete antigen retrieval and to simplify sample handling. Nevertheless, most investigators in the field of FFPE tissue proteomics highlights the heating treatment at high temperatures (80 - 100°C) as the most critical step for efficient protein extraction from FFPE tissues.<sup>122-125</sup> The first attempts for the efficient reversion of crosslinks within FFPE treated tissue samples was the efficient antigen detection via immunohistochemistry, why the method is termed heat-induced antigen retrieval (HIAR).<sup>119, 142</sup> Nowadays the HIAR protocol is widely used for efficient protein extraction from FFPE tissue samples for LC-MS/MS based proteomics.<sup>1, 5, 8, 9, 99, 103-106, 116, 123, 125, 141, 143, 162, 176, 179-182</sup> Several different buffer ingredients are tested for the HIAR protocol, such as SDS, CHAPS (urea-based), TFE, guanidine-HCl, glycine, β-octylglycoside and many more, but the most working groups declares SDS as the critical component for extraction success.<sup>99, 105, 126, 132, 178</sup> Efficient protein extraction from FFPE tissue samples is highly dependent on pH. Higher pH levels typically boost protein recovery, while lower pH levels tend to diminish protein recovery.<sup>142</sup> It is advisable to maintain a basic pH range between 8.0 to 9.5 for optimal protein retrieval.<sup>116, 142, 143</sup> In some studies, it is recommended to use high pressure in combination with heating to increase total protein extraction.<sup>124, 183</sup> In 2019, Shi et al. concluded after 28 years of development, the straightforward HIAR method serving as the pivotal tool to unlock the potential of FFPE tissue proteomics.<sup>143</sup> Several working groups used different varieties of the HIAR method for following proteomic sample preparation and analysis of scarce FFPE specimens and the results are listed in Table 4.<sup>1-7</sup> This study employs an optimized version of the HIAR method tailored for microdissected FFPE tissue samples with low cell counts. This enhanced approach improves sensitivity for analyzing limited FFPE tissues. The results achieved with this improved HIAR protocol are compared to the recent published results for FFPE tissue samples. For the literature comparability all samples during the optimization process are measured in a data dependent

acquisition (DDA) mode and searched against human Uniprot database using Proteome Discoverer (PD). One advantage is the usage of the MSPepSearch node, included in the PD Workflow options, which allows confirmation of peptide sequences identified by database search and to identify unidentified, recurring mass spectra. These changes in the database search algorithm increased protein identification by 11% for 2.5 mm<sup>2</sup> and by 31% for 0.5 mm<sup>2</sup> tissue sections. Comparison of the optimized proteomic sample preparation platform for LCM FFPE tissues to the recent literature in the field of tissue proteomics (fresh frozen and FFPE) points out the need for more sensitive proteomic preparation methods in tissue proteomics. In Table 4 the results of this study are compared to different studies, using a DDA approach for LC-MS/MS acquisition but with several combinations of other tissue lysis and/or sample preparation methods. Finding the right proteomic platform to analyze a distinct group of samples, like FFPE tissues, is highly important to produce comparable results and to step forward to automation for high-throughput analyses. As shown in Table 4, results in this study enables increased total protein identification per mm<sup>2</sup> and per cells compared to other studies in the field of scarce tissue proteomics. Due to the high variety of several different tissue types, it is necessary to align sample preparation methods to ensure complete tissue lysis and for efficient protein cleanup and digest. Here we report a sample preparation platform for LCM FFPE tissues with the ability to increase protein identification and to ensure reproducibility. The final optimization step is based on the selection of the MS acquisition method, in which all steps related to sample preparation for labeling approaches, such as TMT or iTRAQ, were minimized. Therefore, a label-free approach was chosen for the optimization of the MS method.

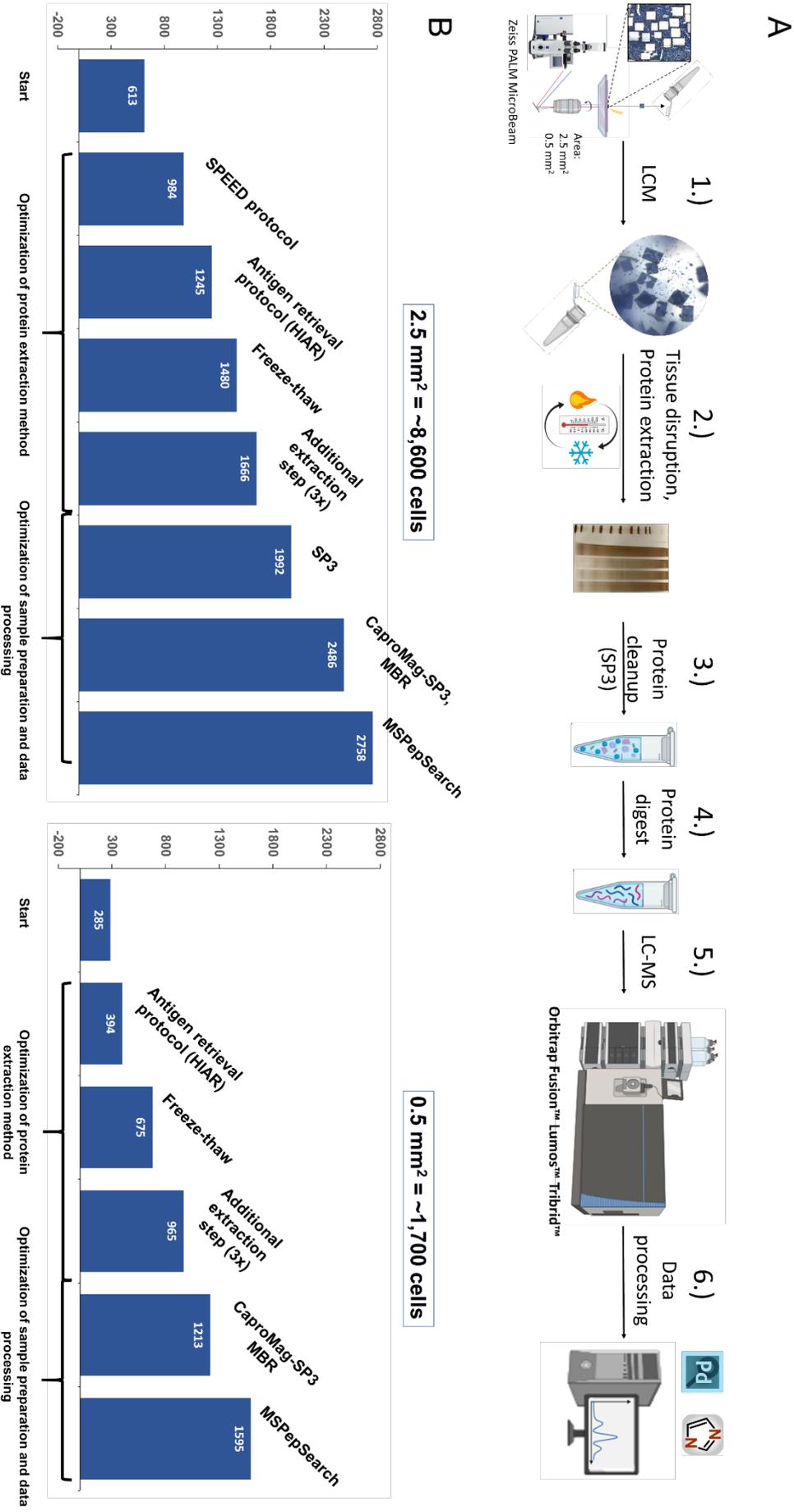
Table 4: Recent studies in the field of scarce tissue proteomics using fresh frozen tissue (2009) and FFPE tissue (after 2011). Every listed publication used LC-MS/MS and label-free quantification, which is comparable to the MS method used in this study. <sup>1-9</sup>

Paper	Year	Microdissected tissue	Processing Method	MS Method	Proteins	Proteins per area/cells
Patel, et al.	2008	Single cells (~20,000 cells, FFPE tissue) of human squamous epithelium from head and neck	Liquid Tissue MS Protein Prep kit	Nano-LC-ESI-MS/MS	115 proteins	6 proteins / 1000 cells
Donald, et al.	2009	Metastatic single cells (~50,000 cells, fresh frozen) of breast carcinoma metastasis from human lymph nodes	Pooled to ~50,000 cells, In-solution digest	LC-MS	531 proteins	11 proteins / 1000 cells
Umar, et al.	2009	0.5 mm <sup>2</sup> (~4,000 cells, fresh frozen) from human breast cancer tissue	Pooled to ~25,000 cells (~6 x 0.5 mm <sup>2</sup> ), In-solution digest	Nano-LC-FTICR MS	2,556 proteins	818 proteins / mm <sup>2</sup> 102 proteins / 1000 cells
Wiśniewski, et al.	2011	20,000 cells of normal colonic mucosa and colon cancer FFPE-tissue	FFPE-FASP-SAX	Nano-LC-MS/MS	3,600 – 4,400 proteins	180 – 220 proteins / 1000 cells
Craven, et al.	2012	5 cm <sup>2</sup> kidney FFPE tissue	FASP	Nano-LC-MS/MS	2,000 proteins	4 proteins / mm <sup>2</sup>
Longuespée, et al.	2016	2,700 cells of breast cancer FFPE tissue	2-D Clean-Up Kit, In-solution digest	Nano-LC-MS/MS	1,400 proteins	519 proteins / 1000 cells
Djuric, et al.	2017	8 – 15 mm <sup>2</sup> pathological normal fetal brain tissue	In-solution digest	Nano-LC-MS/MS	2,200 – 2,400 proteins	160 – 275 proteins / mm <sup>2</sup>
Föll, et al.	2018	118 mm <sup>2</sup> FFPE tonsil tissue 293 mm <sup>2</sup> FFPE tonsil tissue	Direct tissue trypsinization (DTR)	Nano-LC-MS/MS	1,841 proteins	15.6 proteins / mm <sup>2</sup>
Friedrich, et al.	2021	150 mm <sup>2</sup> FFPE lung ADC tissue	SDS-SP3	Nano-LC-MS/MS	2,917 proteins	19.4 proteins/mm <sup>2</sup>
<b>This study</b>	2021	2.5 mm <sup>2</sup> (~8,600 cells, FFPE tissue) of normal pancreatic tissue	SDS-SP3, single run (s.r.) and match between runs (MBR) of 2 x 2.5 mm <sup>2</sup>	Nano-LC-MS/MS	1,992 proteins/single run 2,758 proteins (MBR, 2 x 2.5 mm <sup>2</sup> )	797 proteins / mm <sup>2</sup> (s.r.) 1103 proteins / mm <sup>2</sup> (MBR) 230 proteins / 1000 cells (s.r.) 320 proteins / 1000 cells (MBR)
<b>This study</b>	2021	0.5 mm <sup>2</sup> (~1,700 cells, FFPE tissue) of normal pancreatic tissue	SDS-SP3, single run (s.r.) and match between runs (MBR) of 2 x 2.5 mm <sup>2</sup>	Nano-LC-MS/MS	1213 proteins/single run 1595 proteins (MBR, 2 x 2.5 mm <sup>2</sup> )	2426 proteins / mm <sup>2</sup> (s.r.) 3190 proteins / mm <sup>2</sup> (MBR) 710 proteins / 1000 cells (s.r.) 930 proteins / 1000 cells (MBR)

### 5.3 Optimization of disruption/extraction method

We present a simple and fast proteomic sample preparation platform for the preparation of small-size microdissected FFPE tissues. The resulting sample preparation protocol uses the HIAR protocol from Zecha, Satpathy, Kanashova, Avanesian, Kane, Clauser, Mertins, Carr and Kuster<sup>163</sup> with small adjustments and a SP3-based protein cleanup and digestion method using the CaproMag<sup>TM</sup>.<sup>184</sup> The general workflow of the optimized method is shown in Figure 6A. Starting from microdissected pancreatic tissue, areas of 2.5 mm<sup>2</sup> and 0.5 mm<sup>2</sup> are isolated in PCR tubes, followed by the HIAR-based tissue disruption protocol. Therefore, 60 µL of the extraction buffer (300 mM TRIS/HCl, 2% SDS, pH 8.0) is added to the microdissected pieces for each extraction step. Before every boiling step, the tissue pieces are freeze in the remaining extraction buffer by cooling in liquid nitrogen and are then directly subjected to the 99 °C pre-heated thermocycler for 25 min and 350 rpm. After ultrasonication on ice for 20 min with 30 s on/off cycles, the extraction suspension is subjected to a thermocycler for 2 h at 80 °C, followed by a second ultrasonication step for 20 min with 30 s on/off cycles. The suspension is then centrifuged and the remaining supernatant is collected. After collecting all supernatant from three extraction rounds 1/3 of the total volume of the tissue lysate is subjected to CaproMag<sup>TM</sup>-SP3. The total volume of the remaining FFPE tissue lysate is in average around 150 µL. Small adjustments of the HIAR protocol for scarce FFPE tissue samples, in sizes of 0.5 mm<sup>2</sup> and 2.5 mm<sup>2</sup> are the introduction of freeze-thaw treatments before each boiling step and using small volumes (60 µL) of extraction buffers with three extraction rounds increased total protein extraction in our study (**Error! Reference source not found.**). First attempts using the so-called SPEED (Sample Preparation by Easy Extraction and Digestion) protocol, according to Doellinger, Schneider, Hoeller and Lasch<sup>185</sup>, resulted in low protein recovery and reproducibility and we decided to use the classical approach, the HIAR protocol for optimization toward scarce FFPE tissue proteomics. Several important optimization steps affect downstream sample preparation and are further explained in detail in the following chapters.

Figure 6: A) General workflow for sample preparation to analyse sections of microdissected FFPE pancreas tissue: 1.) laser microdissection of FFPE pancreatic tissue 2.) Tissue disruption (3 x) with alternating freeze-thaw cycles 3.) protein clean-up using SP3-approach (optional reduction/alkylation) 4.) protein digest (tryptic) 5.) LC-MS measurement 6.) Data processing. B) Number of identified proteins in every optimization step for laser micro-dissected normal pancreas FFPE tissue sections with low section size / cell numbers (2.5 mm<sup>2</sup> = 8,600 and 0.5 mm<sup>2</sup> = 1,700 cells)



## 5.4 Determination of protein concentration

For the determination of the protein concentration of a remaining FFPE tissue lysate the Pierce™ BCA Protein Assay from Thermo fisher Scientific is applicably for whole FFPE slides (~10 mm<sup>2</sup>). This assay was tested with a calibration standard of BSA in the remaining protein extraction buffer and a concentration limit of detection of 0.25 µg/µL is observed. For the analysis of scarce FFPE tissue samples, especially microdissected FFPE tissues these concentrations cannot be achieved with the extraction volumes used in the optimized HIAR protocol. To estimate protein concentration an SDS-PAGE is used with silver staining for higher sensitivity towards these small amounts of protein. Normal pancreas tissue, microdissected to 2.5 mm<sup>2</sup> is used for the determination of protein concentration.

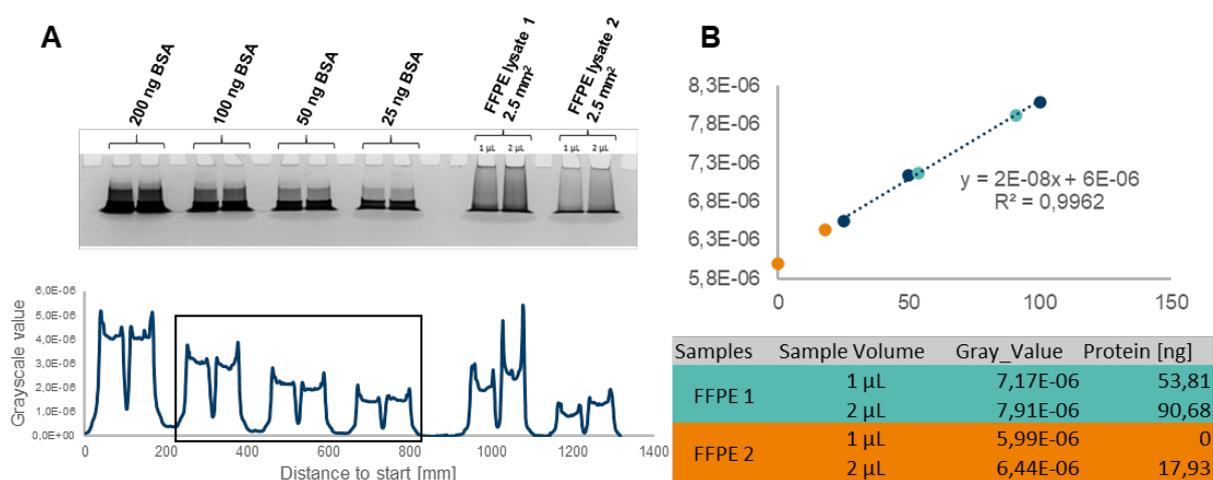


Figure 7: Determination of protein concentration for low-concentrated FFPE tissue lysates, extracted from 2.5 mm<sup>2</sup> tissue samples. A: Picture of the SDS-PAGE gel used for calibration and determination of protein concentration. Below, the distance to the start point (left side of the gel) is plotted against the remaining grayscale values of each lane. The calibration samples used for linear regression are marked with a black box, for which the average value of the densitometric peak is used. B) Linear regression curve of the calibration samples (dark blue) and the position of the measured grayscale values from FFPE tissue lysates in two different amounts (1 µL and 2 µL). The remaining protein concentrations of the FFPE tissue lysates, calculated via regression function are presented in the box below.

Via densitometry the obtained lanes of a calibration series containing different amounts of BSA can be analyzed on its grayscale using ImageJ to plot the remaining values in a linear regression curve. This calibration curve can then be used for the determination of the protein concentration of the remaining FFPE tissue lysates. Observations during SDS-PAGE needs to be questioned critically, because proteins extracted from FFPE tissue lysates behave different in comparison to proteins extracted from samples, that did not undergo FFPE treatment. It has to be noticed, that the running behavior of FFPE proteins in the SDS-PAGE shows tailing, which can be explained by residual FFPE crosslinks. Nevertheless, it can be observed, that heating of the SDS-Sample and following ultrasonication prevents sample from sticking to the start position. However, the running behavior of proteins extracted from FFPE sample lysates is not reproducible, why it is not possible to determine protein concentration with high precision. Under these circumstances protein concentration is estimated to be in a range between 20 – 50 ng/µL for the remaining 2.5 mm<sup>2</sup> FFPE lysates obtained by the extraction protocol presented in chapter 5.3. Using 1/3

(50  $\mu$ L) of the total lysate volume of 150  $\mu$ L reveals the possibility to analyze each tissue lysate in a triplicate with around 1.0 – 2.5  $\mu$ g of total protein amount, before sample digestion for LC-MS/MS analysis.

## 5.5 Optimization of MS sample preparation

SP3 is well established for MS preparation of routine sample. Therefore, SP3 was tested for analysis of scarce sample amounts extracted from LCM FFPE tissues.

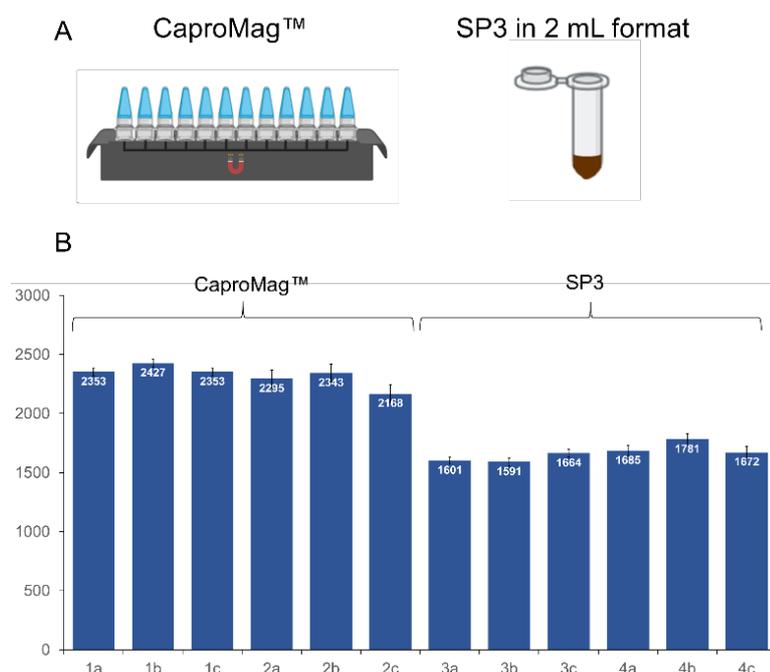


Figure 8: A) Schematic figure of both SP3 methods, CaproMag™-SP3 and commercially available SP3 in 2 mL format. (Designed with BioRender.com) B) Triplicates of 2.5 mm<sup>2</sup> tissue lysates (a, b, c) from different microdissected normal pancreas tissues (1, 2, 3, 4), further processed with CaproMag™-SP3 or conventional SP3.

The CaproMag™ device (Figure 8A) enables simple handling of small volumes and small amounts of magnetic beads, due to the use of the PCR-stripe format and the collection of the beads in the caps. CaproMag™-SP3 is a rapid and highly reproducible protein preparation method and increases protein identification by 20% in our study and is more reproducible, especially for very small amounts of protein (Figure 8) and applicably for LCM FFPE tissue lysates. According to Hughes et al.<sup>159</sup> SP3 sample preparation starts with bead binding of the previously in-solution reduced and alkylated FFPE sample lysate. It is recommended to use denaturing reagents for in-solution reduction and alkylation, such as SDS, which is included in the lysis buffer of the HIAR procedure. For 50  $\mu$ L sample lysate the estimated protein amount is between 1.0 – 2.5  $\mu$ g and the required protein-to-bead ratio for SP3 is 1:10. For optimal bead binding, even at samples with the highest reachable protein concentration, resulting in a protein amount of 2.5  $\mu$ g in 50  $\mu$ L sample lysate, the amount of beads is set to 25  $\mu$ g. This excess of beads avoid bead clumping and ensures complete protein-bead binding for small variations in protein concentration of large numbers of different LCM FFPE samples. For efficient bead binding an organic solvent is used to trigger the formation of protein-bead aggregates, to capture proteins

on the beads for subsequent cleanup of the remaining FFPE tissue lysate. According to the SP3 guidelines from Hughes, Moggridge, Müller, Sorensen, Morin and Krijgsveld <sup>159</sup>, the minimum solvent concentration for efficient bead binding is 50% and suitable solvents are ethanol, methanol or acetonitrile. We tested different concentrations of ethanol for the efficient bead binding of proteins coming from FFPE tissue lysates to ensure complete protein recovery.

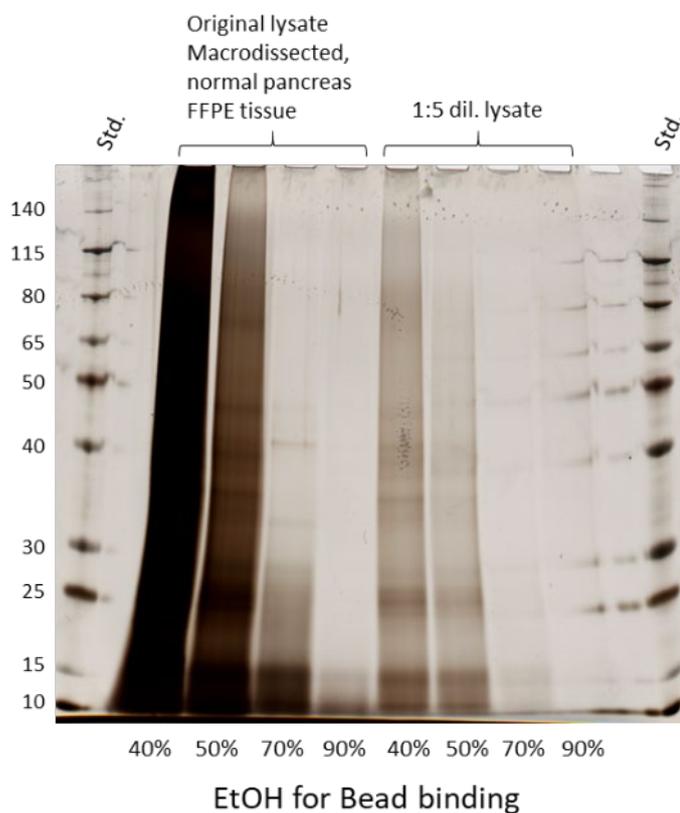


Figure 9: Supernatants of the sample lysates after bead binding triggered by different solvent concentrations using ethanol (40% - 90%). On the left side the original sample lysate of the microdissected pancreatic tissue is used and on the right side the 1:5 diluted lysate is used to mimic scarce FFPE sample lysates.

Therefore, we use minimum ethanol concentration of 75% for efficient protein-bead binding, suggesting a large excess of beads for samples with small protein amounts. The sample cleanup is realized by three rinsing steps using 80% ethanol and one last rinsing step with 100% acetonitrile. After complete removal of the remaining solvent the beads are resuspended in 20  $\mu$ L 50 mM TEAB with final 1:50 Trypsin (0.05  $\mu$ g) related to the highest possible protein concentration for these FFPE tissue lysates (2.5  $\mu$ g). Griesser, Wyatt, Ten Have, Stierstorfer, Lenter and Lamond <sup>162</sup> inserted a peptide cleaning step, by acidification of the remaining digestion solution, followed by addition of high solvent concentrations (EtOH or ACN). The suggestion is, that peptides are then bound to the magnetic beads and can be cleaned by rinsing steps with solvents like ethanol or acetonitrile. In our study, we tested the afterwards peptide cleaning and found low reproducibility and a loss of some hydrophobic and small peptides, that are then washed away by these rinsing steps. We recommend evaporating the digestion buffer and to remove remaining buffer salts by acidification, which is achieved by dissolving the lyophilized peptides in 0.1% TFA. We used TEAB as sample digestion buffer over ammonium bicarbonate

buffer, because it is more volatile after acidification, to ensure complete removal of the remaining buffer salts. The remaining peptide solution can then get subjected to LC-MS/MS analysis.

## 5.6 Crosslinks from formalin fixation

Formalin fixation is mechanistically a nucleophilic addition of a carbonyl compound (formaldehyde) to an aliphatic or aromatic amine coming from a macromolecule (proteins, DNA/RNA).

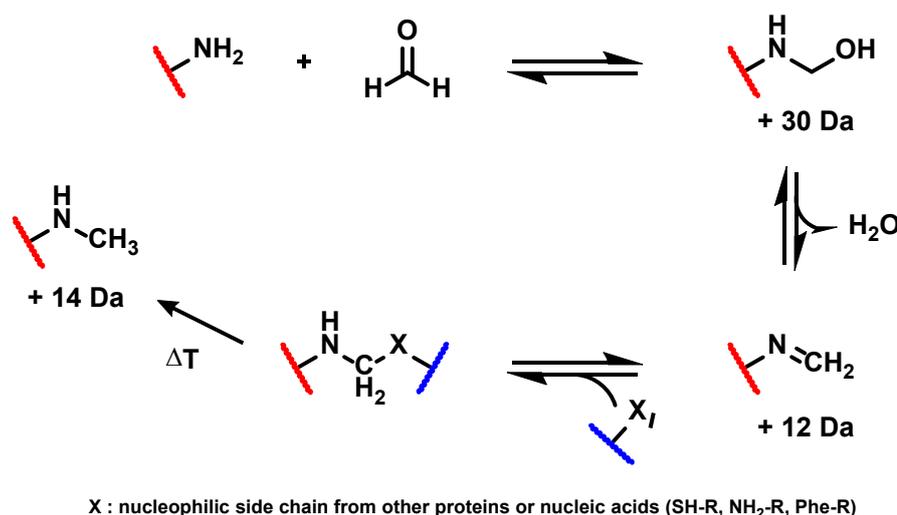


Figure 10: General mechanism of formalin fixation of macromolecules, like proteins and nucleic acids (RNA/DNA), within a tissue slide. Mass shift for each intermediate is shown below the concerning intermediate. After heating treatments (from HIAR protocol) a possible decomposition is the formation of a methylated amine residue (+14 Da).

After nucleophilic attack of any nucleophilic side chain from another macromolecule the crosslinked product is formed, which is necessary for the great preservation efficiency achieved by formalin fixation. Another side effect of formalin fixation is the dehydration and shrinkage of the remaining tissue piece. This enables the static fixation of any process within a cell and/or tissue and prevents proteins and DNA/RNA from decomposition by hydrolysis. To proof the efficiency of the HIAR protocol a modification analysis is achieved by introducing the mass shift of +30 Da for peptides carrying the hydroxymethyl residue (methylol adduct) and +12 Da for peptides carrying the methylene residue (methylene adduct). Heat-induced decomposition of the crosslink adduct can result in the formation of a methylamine adduct, for which the existence is also proofed by modification analysis using the mass shift of +14 Da. Table 5 shows the results of the modification analysis, revealing a low number of peptides carrying the methylol and/or methylene adduct. Methylation at lysine is more frequently observed in modification analysis compared to samples that did not undergo formalin fixation, with about 17% of the peptides carrying the modification. However, this does not impact the overall protein identification, as the highest number of protein IDs is found in the standard search without these modifications, and the results do not differ significantly.

Table 5: Results of Database search with Proteome Discoverer of 6 x 2.5 mm<sup>2</sup> tissue pieces and modification analysis to detect residual modifications arising from FFPE treatment.

	Normal search	Search with Methylene/Methylol mod. at Lysine	Search with Methyl mod. at Lysine
Proteins	2,482	2,354	2,449
PSM	117,222	106,388	118,005
Peptides	22,214	20,802	22,801
Modified Peptides	-	309 (1.5%)	3,951 (17%)

## 5.7 Optimization of MS acquisition method

In the recent years several software tools for the processing of mass spectra measured in different label-free acquisition modes to DDA, such as data independent acquisition (DIA) have been improved, why DIA has gained more popularity in all fields of proteomics. For DIA RAW file processing a spectral library is needed to match the fragment spectra for peptide identification and required the previous generation of sample specific spectral libraries in DDA mode. Generation of these spectral libraries is often time-consuming and the resulting libraries do not improve identification markedly compared to DDA or other MS-methods. The emerging artificial intelligence technology had also made its way to proteomic analysis applications. The *in silico* prediction of spectral libraries within the help of neural networks combined with deep learning algorithms, implemented in DIA-NN and Spectronaut™ enables simple DIA data processing. The easy access to open-source analysis tools, such as DIA-NN makes DIA analysis suitable to any kind of proteomic laboratory. To further improve proteomic analysis platform for LCM FFPE tissue on small cell numbers, DDA is replaced by DIA in the last optimization step of our study. For this LC-MS/MS optimization a FFPE tissue lysate of a macrodissected, pancreatic FFPE tissue is diluted to mimic the 2.5 mm<sup>2</sup> and the 0.5 mm<sup>2</sup> tissue lysate to exclude other factors that could influence protein identification during method optimization. Results for the comparison of 4 x 2.5 mm<sup>2</sup> tissue analogues measured in DDA or DIA mode and processed with Proteome Discoverer or DIA-NN in a match between run search are shown in Table 6.

Table 6: Results of MS-method comparison between DDA and DIA on FFPE tissue analogues for 2.5 mm<sup>2</sup> tissue sections from diluted macrodissected pancreatic FFPE tissue lysate. Results refer to 4 x 2.5 mm<sup>2</sup> tissue sections in a match between run search.

	DDA (PD)	DIA (DIA-NN)
PSM (1% FDR)	65,676	78,678
Peptides % of PSM	11,310 17%	19,234 24%
Proteins	1,710	2,888
Proteins (>1 peptide)	1,653	2,826
Protein (min. 3/4 values)	1,618	2,697
Pearson Correlation	0.96	0.98
StD (%)	2%	0.7%

On several levels total identification is increased by using the DIA MS method in comparison to DDA. On the one hand the number of PSM increases by 20% and on the other hand PSM usage increases by identifying 7% more peptides from all spectra. The number of identified proteins raised from 1,710 to 2,888 proteins (increase of 70%) by using DIA instead of DDA. To test method reproducibility 4 x 2.5 mm<sup>2</sup> tissue analogues are analyzed and the remaining Pearson Correlation of the comparison show that results for DIA are more reproducible between different technical replicas (0.98, resp. 0.96). The number of identified Proteins with more than 1 peptide is higher for DIA due to the efficient acquisition of several low abundant peptides owed by the method, that where otherwise excluded by DDA. This affects the number of missing values observed for the DIA method compared to DDA, having the issue of high numbers of missing values. Visualized in Figure 11, DIA outperforms for both tissue lysate analogues (0.5 mm<sup>2</sup> and 2.5 mm<sup>2</sup>) in comparison to DDA owning higher protein ID rates and less missing values. It has to be noticed, that even the 0.5 mm<sup>2</sup> tissue analogues has higher protein ID rates compared to DDA 2.5 mm<sup>2</sup> tissue analogues (1,600, resp. 1,300 proteins). That leads to the assumption, that especially for lysates with low protein amounts, DIA holds higher sensitivity compared to DDA. Going further into detail of the heatmap it can be observed, that high abundant proteins are high abundant for both methods (DDA and DIA), but low abundant proteins are identified with less missing values for DIA, which supports the assumption of higher sensitivity for the DIA method.

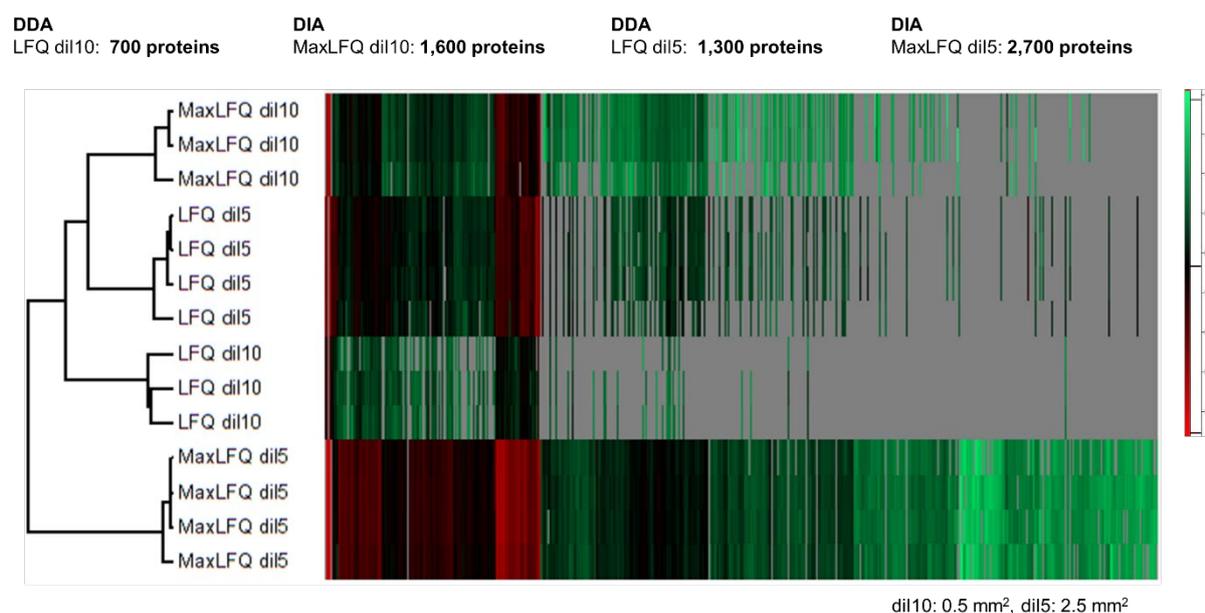
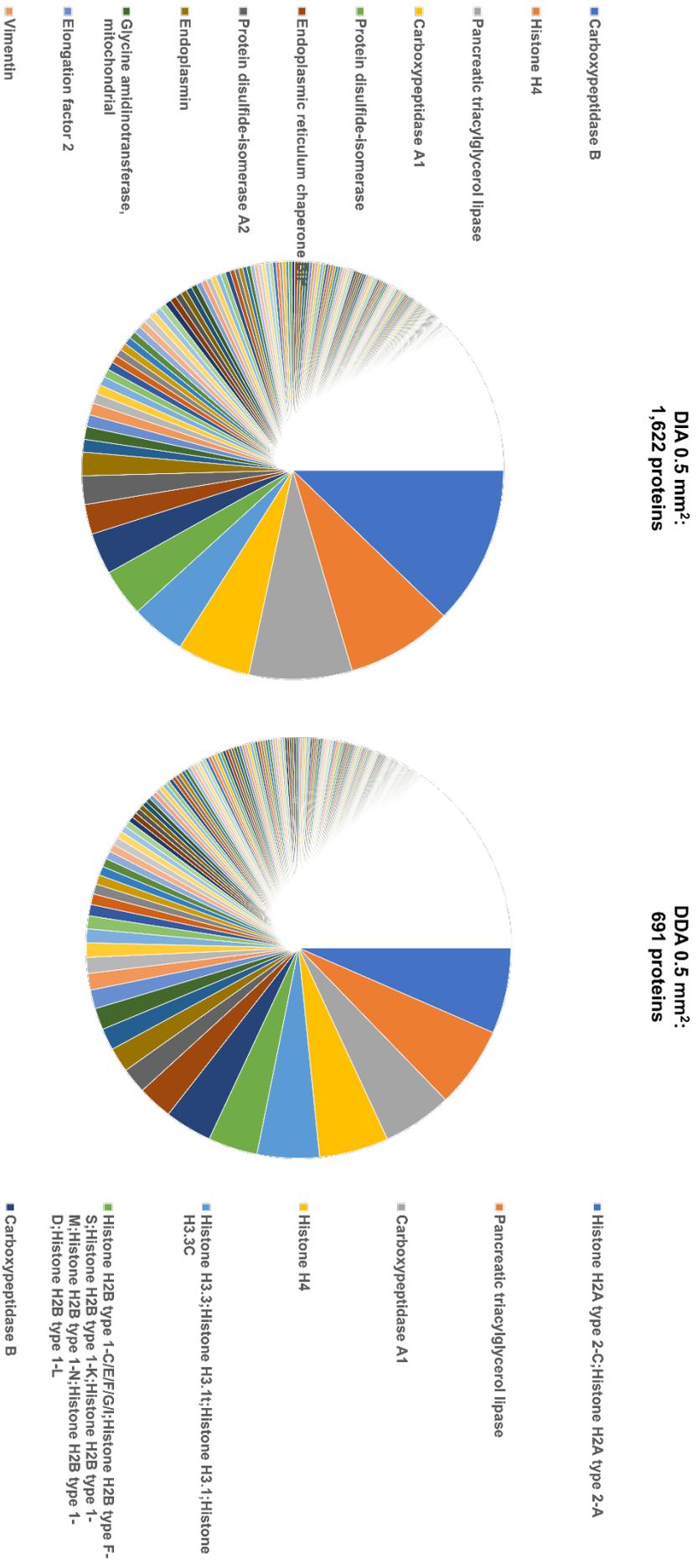


Figure 11: Heatmap clustering of macrodissected pancreatic tissue lysate diluted to mimic 2.5 mm<sup>2</sup> and 0.5 mm<sup>2</sup> tissue analogues and measured in DDA and DIA mode and searched by DIA-NN or MaxQuant. High abundant proteins are colored in red and low abundant proteins are colored in green. Missing values are colored in grey.

Nevertheless, similar proteins and distributions are observed for both methods and validates the results observed for DIA, when DDA is the current state of the art label-free quantitation method. Proteins with the highest abundances, which make up about ~50% of the total intensity are shown in Figure 12 for DIA and DDA data processing concerning the 0.5 mm<sup>2</sup> tissue analogues. Characteristic pancreatic enzymes, such as carboxypeptidase A1 and B and pancreatic

triacylglycerol lipase are high abundant in both MS acquisition methods. Histone H4 is found to be high abundant for both methods, but in DDA several other histones like H3, H2A and H2B are listed in the TopHits. These Histones are either found in the DIA data but are not under the TopHits due to several other high abundant proteins that are found in DIA and not in DDA. As shown, endoplasmic reticulum related proteins, such as endoplasmic reticulum chaperone BiP are identified with high abundances only in the TopHits of DIA. Furthermore, mitochondrial ATP synthase subunit alpha and mitochondrial glycine amidinotransferase are under the TopHit proteins with high abundances for DIA. One reason could be, that DIA is able to collect more peptide fragment spectra of proteins, that are, due to their physicochemical properties, less ionizable in MS acquisition, like for example membrane proteins carrying many hydrophobic peptides. DDA only uses the top20 peptide-ions with the highest intensities from the current MS1 spectrum for further fragmentation. Therefore, the method lacks in sensitivity for samples with high dynamic ranges. The DIA method overcomes these problems, by collecting nearly all peptide-ions and to create co-spectra, that are analyzable with the help of AI-based *in-silico* predicted spectra or the use of pre-measured spectral libraries. The increase of protein ID rate within the DIA method is up to 70% and supports the recommendation for DIA as label-free LC-MS/MS method especially in clinical proteomics and for low sample amounts due to limited number of clinical sample material. In contrast to the need for extensive sample preparation steps, which can lead to loss of material, it is crucial to minimize these steps when analyzing small sample amounts. This reduction helps to preserve the integrity of the sample, making labeling approaches less desirable. Mass spectrometric analyses with labeling approaches have the benefit of high data completeness, which was not reachable for label-free quantitation methods, such as DDA. DIA enable the label-free identification of high numbers of protein identifications with a solid data completeness and without prior labeling treatments during sample preparation, like iTRAQ or TMT.

Figure 12: Comparison of the TopHit proteins with the highest abundances for each acquisition mode (DDA and DIA). Comparison is made on the search results of 0.5 mm<sup>2</sup> tissue analogues with DIA-NN or MaxQuant.



## 6 Summary

The availability of a sufficient number of patient samples is often critical in high-throughput omics approaches. Retrospective collections of formalin-fixed paraffin-embedded (FFPE) tissues provide a valuable source to increase the number of patient samples, offering the opportunity to perform multi-omics approaches. Proteomic analysis of scarce tissue samples has proven to be a successful tool for identifying cell-type specific proteins involved in tumor progression and used for tumor diagnosis. Here, we developed a freeze-thaw lysis protocol for LCM FFPE tissues and an optimized SP3 approach to analyze small cell numbers of microdissected FFPE tissues. This approach resulted in a minimal sample handling method for LC-MS/MS sample preparation and is recommended for small sample amounts especially for scarce FFPE tissue proteomics and/or single-cell proteomics. Automation of tissue lysis and sample preparation is the next step to increase reproducibility to its maximum and to reduce contamination from sample handling. Therefore, sample preparation methods need to get implemented in several automation devices, that are not affordable to every proteomic laboratory. Several new approaches are to integrate tissue microarrays (TMA), often coupled to automatic pipetting devices, to maximize sample throughput and to ensure maximal reproducibility.<sup>186-188</sup> To make scarce FFPE tissue proteomics accessible to a wide range of different laboratories we present a proteomic platform, that can easily be established in the most laboratories with low expenditure. Coupling of our tissue lysis and sample preparation methods to DIA LC-MS/MS method in combination with the open-access analysis tool DIA-NN, makes simple DIA data processing accessible to everyone. The opportunity to build large DIA-based databases gives a chance to further develop more software solutions for different kinds of DIA data processing methods and may be the beginning of a new era in the field of proteomics, especially combined with the upcoming AI-technologies. Using our optimized sample preparation protocol in combination with a data independent acquisition (DIA) method and data processing by DIA-NN led to the identification of 70% more protein IDs in low sample amounts compared to a DDA method.

## 7 References

- (1) Patel, V.; Hood, B. L.; Molinolo, A. A.; Lee, N. H.; Conrads, T. P.; Braisted, J. C.; Krizman, D. B.; Veenstra, T. D.; Gutkind, J. S. Proteomic Analysis of Laser-Captured Paraffin-Embedded Tissues: A Molecular Portrait of Head and Neck Cancer Progression. *Clinical Cancer Research* **2008**, *14* (4), 1002-1014.
- (2) Johann, D. J.; Rodriguez-Canales, J.; Mukherjee, S.; Prieto, D. A.; Hanson, J. C.; Emmert-Buck, M.; Blonder, J. Approaching Solid Tumor Heterogeneity on a Cellular Basis by Tissue Proteomics Using Laser Capture Microdissection and Biological Mass Spectrometry. *Journal of Proteome Research* **2009**, *8* (5), 2310-2318.
- (3) Umar, A.; Kang, H.; Timmermans, A. M.; Look, M. P.; Meijer-van Gelder, M. E.; den Bakker, M. A.; Jaitly, N.; Martens, J. W. M.; Luider, T. M.; Foekens, J. A.; et al. Identification of a Putative Protein Profile Associated with Tamoxifen Therapy Resistance in Breast Cancer\*. *Molecular & Cellular Proteomics* **2009**, *8* (6), 1278-1294.
- (4) Wiśniewski, J. R.; Ostasiewicz, P.; Mann, M. High Recovery FASP Applied to the Proteomic Analysis of Microdissected Formalin Fixed Paraffin Embedded Cancer Tissues Retrieves Known Colon Cancer Markers. *Journal of Proteome Research* **2011**, *10* (7), 3040-3049.

- (5) Craven, R. A.; Cairns, D. A.; Zougman, A.; Harnden, P.; Selby, P. J.; Banks, R. E. Proteomic analysis of formalin-fixed paraffin-embedded renal tissue samples by label-free MS: Assessment of overall technical variability and the impact of block age. *PROTEOMICS – Clinical Applications* **2013**, *7* (3-4), 273-282.
- (6) Longuespée, R.; Fléron, M.; Pottier, C.; Quesada-Calvo, F.; Meuwis, M.-A.; Baiwir, D.; Smargiasso, N.; Mazzucchelli, G.; De Pauw-Gillet, M.-C.; Delvenne, P. Tissue proteomics for the next decade? Towards a molecular dimension in histology. *Omics: a journal of integrative biology* **2014**, *18* (9), 539-552.
- (7) Djuric, U.; Rodrigues, D. C.; Batruch, I.; Ellis, J.; Shannon, P.; Diamandis, P. Spatiotemporal proteomic profiling of human cerebral development. *Molecular & Cellular Proteomics* **2017**, *16* (9), 1548-1562.
- (8) Föll, M. C.; Fahrner, M.; Oria, V. O.; Kühs, M.; Biniossek, M. L.; Werner, M.; Bronsert, P.; Schilling, O. Reproducible proteomics sample preparation for single FFPE tissue slices using acid-labile surfactant and direct trypsinization. *Clinical Proteomics* **2018**, *15* (1).
- (9) Friedrich, C.; Schallenberg, S.; Kirchner, M.; Ziehm, M.; Niquet, S.; Haji, M.; Beier, C.; Neudecker, J.; Klauschen, F.; Mertins, P. Comprehensive micro-scaled proteome and phosphoproteome characterization of archived retrospective cancer repositories. *Nature Communications* **2021**, *12* (1), 3576.
- (10) Shi, S.-R.; Shi, Y.; Taylor, C. R. Antigen retrieval immunohistochemistry: review and future prospects in research and diagnosis over two decades. *Journal of Histochemistry & Cytochemistry* **2011**, *59* (1), 13-32.
- (11) Magdeldin, S.; Yamamoto, T. Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues. *PROTEOMICS* **2012**, *12* (7), 1045-1058.
- (12) O'Rourke, M. B.; Padula, M. P. Analysis of formalin-fixed, paraffin-embedded (FFPE) tissue via proteomic techniques and misconceptions of antigen retrieval. *BioTechniques* **2016**, *60* (5), 229-238.
- (13) Addis, M. F.; Tanca, A.; Pagnozzi, D.; Crobu, S.; Fanciulli, G.; Cossu-Rocca, P.; Uzzau, S. Generation of high-quality protein extracts from formalin-fixed, paraffin-embedded tissues. *PROTEOMICS* **2009**, *9* (15), 3815-3823.
- (14) Tanca, A.; Pagnozzi, D.; Addis, M. F. Setting proteins free: Progresses and achievements in proteomics of formalin-fixed, paraffin-embedded tissues. *PROTEOMICS – Clinical Applications* **2012**, *6* (1-2), 7-21.
- (15) Gustafsson, O. J. R.; Arentz, G.; Hoffmann, P. Proteomic developments in the analysis of formalin-fixed tissue. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2015**, *1854* (6), 559-580.
- (16) Ikeda, K.; Monden, T.; Kanoh, T.; Tsujie, M.; Izawa, H.; Haba, A.; Ohnishi, T.; Sekimoto, M.; Tomita, N.; Shiozaki, H. Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections. *Journal of Histochemistry & Cytochemistry* **1998**, *46* (3), 397-403.
- (17) Wolff, C.; Schott, C.; Porschewski, P.; Reischauer, B.; Becker, K.-F. Successful Protein Extraction from Over-Fixed and Long-Term Stored Formalin-Fixed Tissues. *Plos One* **2011**, *6* (1), e16353.
- (18) Bravo, S.; Garcia-Vence, M.; Chantada-Vazquez, M. d. P.; Sosa-Fajardo, A.; Agra, R.; Barcia de la Iglesia, A.; Otero-Glez, A.; Garcia-Gonzalez, M. A.; Cameselle-Teijeiro, J.; Nuñez, C. Protein extraction from FFPE kidney tissue samples: A review of the literature and characterization of techniques. *Frontiers in Medicine* **2021**, *8*, 553.
- (19) Shi, S.-R.; Taylor, C. R.; Fowler, C. B.; Mason, J. T. Complete solubilization of formalin-fixed, paraffin-embedded tissue may improve proteomic studies. *PROTEOMICS – Clinical Applications* **2013**, *7* (3-4), 264-272.
- (20) Wang, H.; Qian, W.-J.; Mottaz, H. M.; Clauss, T. R. W.; Anderson, D. J.; Moore, R. J.; Camp, D. G.; Khan, A. H.; Sforza, D. M.; Pallavicini, M.; et al. Development and Evaluation of a Micro- and Nanoscale Proteomic Sample Preparation Method. *Journal of Proteome Research* **2005**, *4* (6), 2397-2403.
- (21) Jiang, X.; Jiang, X.; Feng, S.; Tian, R.; Ye, M.; Zou, H. Development of Efficient Protein Extraction Methods for Shotgun Proteome Analysis of Formalin-Fixed Tissues. *Journal of Proteome Research* **2007**, *6* (3), 1038-1047.
- (22) Hwang, S. I.; Thumar, J.; Lundgren, D. H.; Rezaul, K.; Mayya, V.; Wu, L.; Eng, J.; Wright, M. E.; Han, D. K. Direct cancer tissue proteomics: a method to identify candidate cancer biomarkers from formalin-fixed paraffin-embedded archival tissues. *Oncogene* **2007**, *26* (1), 65-76.

- (23) Hughes, C. S.; Moggridge, S.; Müller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols* **2019**, *14* (1), 68-85.
- (24) Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods* **2020**, *17* (1), 41-44.
- (25) Shi, S. R.; Key, M. E.; Kalra, K. L. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. *Journal of Histochemistry & Cytochemistry* **1991**, *39* (6), 741-748.
- (26) Fowler, C. B.; O'leary, T. J.; Mason, J. T. Techniques of protein extraction from FFPE tissue/cells for mass spectrometry. *Antigen Retrieval Immunohistochemistry Based Research and Diagnostics* **2010**, 333-346.
- (27) Chu, W.-S.; Liang, Q.; Liu, J.; Wei, M. Q.; Winters, M.; Liotta, L.; Sandberg, G.; Gong, M. A nondestructive molecule extraction method allowing morphological and molecular analyses using a single tissue section. *Laboratory Investigation* **2005**, *85* (11), 1416-1428.
- (28) Yamashita, S.; Okada, Y. Mechanisms of heat-induced antigen retrieval: analyses in vitro employing SDS-PAGE and immunohistochemistry. *Journal of Histochemistry & Cytochemistry* **2005**, *53* (1), 13-21.
- (29) Prieto, D. A.; Hood, B. L.; Darfler, M. M.; Guiel, T. G.; Lucas, D. A.; Conrads, T. P.; Veenstra, T. D.; Krizman, D. B. Liquid Tissue™: proteomic profiling of formalin-fixed tissues. *BioTechniques* **2005**, *38* (6S), S32-S35.
- (30) Shi, S. R.; Imam, S. A.; Young, L.; Cote, R. J.; Taylor, C. R. Antigen retrieval immunohistochemistry under the influence of pH using monoclonal antibodies. *Journal of Histochemistry & Cytochemistry* **1995**, *43* (2), 193-201.
- (31) Shi, S.-R.; Shi, Y.; Taylor, C. R.; Gu, J. New Dimensions of Antigen Retrieval Technique: 28 Years of Development, Practice, and Expansion. *Applied Immunohistochemistry & Molecular Morphology* **2019**, *27* (10), 715-721.
- (32) Paulo, J. A.; Lee, L. S.; Banks, P. A.; Steen, H.; Conwell, D. L. Proteomic analysis of formalin-fixed paraffin-embedded pancreatic tissue using liquid chromatography tandem mass spectrometry. *Pancreas* **2012**, *41* (2), 175-185.
- (33) Nirmalan, N. J.; Hughes, C.; Peng, J.; McKenna, T.; Langridge, J.; Cairns, D. A.; Harnden, P.; Selby, P. J.; Banks, R. E. Initial development and validation of a novel extraction method for quantitative mining of the formalin-fixed, paraffin-embedded tissue proteome for biomarker investigations. *J Proteome Res* **2011**, *10* (2), 896-906.
- (34) Griesser, E.; Wyatt, H.; Ten Have, S.; Stierstorfer, B.; Lenter, M.; Lamond, A. I. Quantitative Profiling of the Human Substantia Nigra Proteome from Laser-capture Microdissected FFPE Tissue. *Mol Cell Proteomics* **2020**, *19* (5), 839-851.
- (35) Crockett, D. K.; Lin, Z.; Vaughn, C. P.; Lim, M. S.; Elenitoba-Johnson, K. S. J. Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS. *Laboratory Investigation* **2005**, *85* (11), 1405-1415.
- (36) Coscia, F.; Doll, S.; Bech, J. M.; Schweizer, L.; Mund, A.; Lengyel, E.; Lindebjerg, J.; Madsen, G. I.; Moreira, J. M.; Mann, M. A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis. *J Pathol* **2020**, *251* (1), 100-112.
- (37) Bauden, M.; Kristl, T.; Andersson, R.; Marko-Varga, G.; Ansari, D. Characterization of histone-related chemical modifications in formalin-fixed paraffin-embedded and fresh-frozen human pancreatic cancer xenografts using LC-MS/MS. *Laboratory Investigation* **2017**, *97* (3), 279-288.
- (38) Fowler, C.; O'Leary, T.; Mason, J. Improving the Proteomic Analysis of Archival Tissue by Using Pressure-Assisted Protein Extraction: A Mechanistic Approach. *Journal of proteomics & bioinformatics* **2014**, *7*, 151-157.
- (39) Zecha, J.; Satpathy, S.; Kanashova, T.; Avanesian, S. C.; Kane, M. H.; Clauser, K. R.; Mertins, P.; Carr, S. A.; Kuster, B. TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach\*[S]. *Molecular & Cellular Proteomics* **2019**, *18* (7), 1468-1478.
- (40) Lenz, T. Devices, systems and methods for separating magnetic particles. Google Patents: 2012.
- (41) Doellinger, J.; Schneider, A.; Hoeller, M.; Lasch, P. Sample Preparation by Easy Extraction and Digestion (SPEED) - A Universal, Rapid, and Detergent-free Protocol for Proteomics Based on Acid Extraction. *Molecular & Cellular Proteomics* **2020**, *19* (1), 209-222.
- (42) Takikita, M.; Chung, J.-Y.; Hewitt, S. M. Tissue microarrays enabling high-throughput molecular pathology. *Current Opinion in Biotechnology* **2007**, *18* (4), 318-325.

- (43) Hewitt, S. M. Design, construction, and use of tissue microarrays. In *Protein Arrays*, Springer, 2004; pp 61-72.
- (44) Espina, V.; Mehta, A. I.; Winters, M. E.; Calvert, V.; Wulfschlegel, J.; Petricoin III, E. F.; Liotta, L. A. Protein microarrays: Molecular profiling technologies for clinical specimens. *PROTEOMICS* **2003**, 3 (11), 2091-2100.

## Author contributions to Manuscript 2

### **Proteome analysis of microdissected pancreas tissue by improved HIAR protocol**

#### Stella Pauls

Planned and performed method optimization of HIAR protocol. Designed, performed and analyzed FFPE samples from normal pancreas tissue. Performed data analysis and post processing. Prepared figures and tables and wrote and edited the manuscript.

#### Anja Stefanski

Conceived and supervised the study.

#### Friederike Opitz

Coordinated and supervised the isolation of FFPE tissue samples and delivery.

#### Sandra Biskup

Performed isolation of FFPE tissue samples. Coordinated sample collection and delivery.

#### Gereon Poschmann

Helped performing data analysis and post processing.

#### Thomas Lenz

Offered CaproMag Device and helped performing data analysis and post processing.

#### Irene Esposito

Conceived and supervised the study, acquired funding.

#### Kai Stühler

Conceived and supervised the study, acquired funding, reviewed/edited the manuscript.



## 7 Manuscript 3:

Proteome analysis with combined DIA and Glyco DIA approach of laser microdissected precursor lesions from pancreatobiliary cancer to improve early cancer diagnostics



# Proteome analysis with combined DIA and Glyco-DIA approach of laser microdissected precursor lesions from pancreatobiliary cancer to improve early cancer diagnostics

Stella Pauls<sup>1</sup>, Anja Stefanski<sup>2</sup>, Friederike Opitz<sup>3</sup>, Sandra Biskup<sup>3</sup>, Irene Esposito<sup>3</sup>, Gereon Poschmann<sup>1</sup>, Thomas Lenz<sup>1</sup>, Marc-Daniel Driessen<sup>1</sup>, Kai Stühler<sup>1,2</sup>

<sup>1</sup>Proteome Research, Institut für Molekulare Medizin, Medizinische Fakultät, Heinrich-Heine-Universität, Universitätsstraße 1, D-40225 Düsseldorf

<sup>2</sup>Molecular proteomics laboratory (MPL), Institut für Molekulare Medizin, Medizinische Fakultät, Heinrich-Heine-Universität, Universitätsstraße 1, D-40225 Düsseldorf

<sup>3</sup>Institut für Pathologie, Universitätsklinikum Düsseldorf, Medizinische Fakultät, Heinrich-Heine-Universität, Moorenstr. 5, D-40225 Düsseldorf

## Abstract:

Proteomic analysis of different tissue samples offers the opportunity to directly detect proteins, that are involved in tumor progression, with the potential to find new disease-specific biomarker for improved diagnostics. For pancreatobiliary (PB) cancer a specific biomarker for diagnostics is not yet known, except the tumor markers CA 19-9 and CEA, which are used for follow-up purposes but cannot be used for early diagnosis of PB cancer due to their elevation also in tumor non-developing (TND) conditions, such as pancreatitis. The early stage in tumor progression is associated with a transition of normal cells into so-called precursor lesions, which seem to develop along similar pathways and show numerous morphological and molecular similarities. The group of Prof. Esposito at the Institute of Pathology developed methods for morphomolecular characterization of PB precursors as well as molecular subtyping of different precursor stages applied on formalin-fixed, paraffin-embedded (FFPE) tissues. For the characterization of PB cancer, several proteome studies have been attempted to detect proteins involved in tumor progression. In the last years, there has been an increased emphasis on molecular diagnostics for specific subpopulations, as the influence of surrounding cells (such as acinar, endocrine, stromal and inflammatory cells) has been minimized. In this study we optimized and applied sensitive methods combining microdissection and quantitative, label-free proteomics to gain a more comprehensive understanding in the development of PB cancer. By using a data independent acquisition (DIA) mode for LC-MS/MS analysis the goal is a combined proteomic and glycoproteomic (specific O-glycans) analysis to unravel differences in TD and TND tissue transformation of pancreatic precursor lesions. These findings support clinical decisions and improves early cancer diagnostics.

## 1 Introduction

The 12<sup>th</sup> most common cancer worldwide is pancreatic cancer, for which in 2020 over 495,000 new cases were reported.<sup>12</sup> For 2023 the American Cancer Society estimates 64,050 new cases

of pancreatic cancer and it was reported in 2020 as the 4<sup>th</sup> leading cause of cancer deaths at all ages and genders.<sup>13</sup> The main reason for the high mortality of pancreatic cancer is the lack of sufficient treatment, which has rarely changed over the past 30 years and consists resection of the tumor tissue, which is not viable in cases, where the tumor has expanded to other regions/organs. Another reason for the high mortality is the fact, that most patients do not recognize any symptoms and currently no early detection tests for pancreatic cancer are available. Therefore, over 50% of the patients are diagnosed, when the tumor already expanded to other organs (metastasis), while at this stage the 5-year overall survival rate is under 5%.<sup>13</sup> The important cells in the pancreas can be divided into two subclasses, the exocrine cells, that produce digestion enzymes for the duodenum and the endocrine cells, which produces hormones, like Insulin and Glucagon. Primarily pancreatic cancer develops in exocrine pancreatic cells, which are localized in the head of the pancreas, exposed to the pancreatobiliary ducts. About 95% of all pancreatic cancers are pancreatic ductal adenocarcinoma (PDAC).<sup>14</sup> While an environment of high concentrations of digestion enzymes of the bile liquid in the pancreatobiliary ducts keeps the risk of endogenous tissue damage, PDAC is developing from so-called precursor lesions, characterized by atypical tissue changes. Pancreatobiliary precursor lesions are possible starting points for the development of PDAC and cholangiocarcinoma (CCA), for which increasing morphological changes are reported at lesion progression.<sup>15-17</sup>

## 2 Material and Methods

### 2.1 Sample collection

Tissue biopsies were formalin-fixed and paraffin embedded, morphomolecular characterized and microdissected by the Institute of pathology of the Universitätsklinikum Düsseldorf. Staining was performed manually or using the Ventana BenchMark Ultra automated IHC/ISH slide staining system (Ventana Medical Systems Inc., Tucson, USA) (suppl. table 1a), using the H&E staining reagent. The cytoplasmic expression of TFF3 and MUCL3 was evaluated using the immunoreactivity scoring system (IRS) based on staining intensity (0: negative, 1: mild, 2: moderate, 3: intense) and the percentage of stained cells (0: no positive cells, 1: <10% positive cells, 2: 10-50% positive cells, 3: 51-80% positive cells, 4: >80% positive cells). The final score (0-12) was found by multiplying the positive cells proportion score (0-4) and the staining intensity score (0-3). The mean value of Ki67 proliferation rate of five randomly selected high power (40x) fields (HPFs) was calculated using the percentage of stained cells. A tissue collection of precursor lesions of PDAC obtained from 154 patients operated on in the years 2008–2021 was established. Samples, in which a tumor was already present are classified to tumor developing (TD) group and sample in which no tumor was observed are classified as tumor non-developing (TND) group. All different tissue types and the respective grades are characterized and documented in Table 7.

Table 7: Sample table of all samples used in this study. Grades are divided to normal (N), low-grade (LG), high-grade (HG) and Tumor (T). Some precursor lesions developed a tumor (w = with PDAC/CCA) or not (wo = without PDAC/CCA). Precursor lesions from PDAC are pancreatic intraepithelial neoplasm (PanIN) from pancreatic tissue type and intraductal papillary mucinous neoplasms (IPMN) from gastric and intestinal tissue type. Precursor lesions from CCA are biliary intraepithelial neoplasm (BillIN) from pancreatobiliary tissue type and intraductal papillary neoplasms of the bile duct (IPNB) from gastric, intestinal and pancreatobiliary tissue type.

Lesion	Tissue type	Number of all microdissected lesions								Overall
		N (TD)	N (TND)	LG (TD)	LG (TND)	HG (TD)	HG (TND)	T (TD)	Met (TD)	
<b>PanIN</b>	Pancreas	9	10	10	10	7	1	10	0	57
<b>IPMN</b>	Gastric	6	3	4	4	6	1	6	0	30
<b>IPMN</b>	Intestinal	2	6	2	6	4	1	2	0	23
<b>BillIN</b>	Pancreatobiliary	4	8	5	8	6	4	6	3	44
<b>IPNB</b>	Gastric	0	1	0	2	1	1	1	0	6
<b>IPNB</b>	Intestinal	1	4	0	1	3	4	3	0	16
<b>IPNB</b>	Pancreatobiliary	1	0	0	0	1	0	1	0	2

## 2.2 Sample preparation for LC-MS/MS analysis

A modified FFPE tissue lysis protocol of Ikeda and colleagues was applied.<sup>102</sup> Microdissected FFPE tissue pieces were transferred in methanolic suspension into PCR-strips, were centrifuged and the methanolic supernatant was removed followed by air-drying of the residual solvent. The remaining tissue was resuspended in 60  $\mu$ L lysis buffer (300 mM TRIS/HCl, 2% SDS, pH 8.0), shock-frozen in liquid nitrogen and immediately heated for 25 min at 99°C and 350 rpm. For complete lysis, samples were ultrasonicated on ice for 20 min with 30 s on/off cycles and then shook for 2 h at 80°C and 500 rpm, followed by a second ultrasonication step. After centrifugation, the supernatant was transferred into a new 0.5 mL Lobind Eppendorf tubes and the pellet was resuspended in 60  $\mu$ L buffer for a second extraction (resp. third), similar to the first. The resulting supernatants were combined and a modified magnetic bead-based sample preparation protocol according to Hughes and colleagues were applied.<sup>159</sup> Briefly, 50  $\mu$ L per sample lysate were reduced by adding 1.25  $\mu$ L 300 mM DTT (dithiothreitol) and shaking for 20 min at 56°C and 1000 rpm, followed by alkylation with the addition of 6.67  $\mu$ L 100 mM IAA and incubation for 15 min in the dark. A 20  $\mu$ g/ $\mu$ L bead stock of 1:1 Sera-Mag SpeedBeads was freshly prepared and 1.25  $\mu$ L were added to each sample. Afterwards, for protein aggregation capture 160  $\mu$ L ethanol was added to a final concentration of 73% and incubated for 15 min at 20°C. After three rinsing steps with 80% EtOH and one rinsing step with 100% ACN, beads were resuspended in 20  $\mu$ L of 50 mM TEAB buffer and digested with final 0.05  $\mu$ g trypsin at 37°C and 1,000 rpm overnight. Extra-digestion was carried out by adding 2  $\mu$ L of 0.025  $\mu$ g/ $\mu$ L trypsin stock and shaking at 37°C and 1000 rpm for 4 h. The supernatants were collected and the remaining solvent was removed by evaporation. Samples were subjected to LC-MS/MS by dissolving in 17  $\mu$ L 0.1% TFA.

## 2.3 Liquid chromatography-mass spectrometry (LC–MS/MS) analysis

For the LC-MS acquisition an Orbitrap Fusion Lumos Tribrid Mass Spectrometer with FAIMS (Thermo Fisher Scientific) coupled to an Ultimate 3000 Rapid Separation liquid chromatography system (Thermo Fisher Scientific, Idstein, Germany) equipped with an Acclaim PepMap 100 C18 column (75  $\mu\text{m}$  inner diameter, 25 cm length, 2 mm particle size from Thermo Fisher Scientific) as separation column and an Acclaim PepMap 100 C18 column (75  $\mu\text{m}$  inner diameter, 2 cm length, 2 mm particle size from Thermo Fisher Scientific) as trap column. A LC-gradient of 180 min was applied and the mass spectrometer operated in positive data independent acquisition mode with a scan range of 380 – 985  $m/z$  at a resolution of 60,000 for the MS1 and a scan range of 145 – 1450  $m/z$  with a defined isolation window of 10  $m/z$  with an overlap of 1  $m/z$  and a resolution of 15000 for the MS2. FAIMS operated with a nitrogen carrier gas flow of 4.5 L/min and the compensation voltage (CV) was set to -50 V. The capillary temperature was set to 275°C, the source voltage to 2.0 kV, the normalized automatic gain control (AGC) target was set to 200% and the maximum injection time was 40 ms. HCD collision energy was set to 30%.

## 2.4 MS data analysis

### 2.4.1 Proteomics

Data analysis was carried out with DIA-NN (version 1.8.1, by Demichev, Ralser and Lilley labs, available on <https://github.com/vdemichev/DiaNN>).<sup>168</sup> All RAW files were searched against the human proteome UniProt KB dataset (UP000005640, download on 12.01.2023, 81837 entries) and the Maxquant Contaminant database (download on 03.05.2022), using the deep learning tool to generate an *in silico* spectral library, which is implemented in DIA-NN. The digestion enzyme was set to trypsin, the maximum number of missed cleavages was set to two and the peptide length was 7 – 30 amino acids. Mass accuracy was optimized by DIA-NN using the first run in the experiment. As variable modifications were methionine oxidation, N-terminal methionine loss and methylation of lysines (only for FFPE tissue samples) defined. Fixed modification was carbamidomethylation of cysteines. All samples of the study were analyzed in a match between run (MBR) search.

During post processing in R (version 4.2.3), peptides were ungrouped and filtered to 1% FDR on protein and peptide level and to all proteins identified with  $\geq 2$  peptides. Contaminants were filtered out and the results were exported as excel sheet. Analysis of Variance (ANOVA) was carried out using the `aov()` function of the standard R package 'stats'. Benjamini-Hochberg correction of the obtained p-values was performed using the function `p.adjust()`, also a part of the standard R package 'stats'. To identify significant pairwise differences between group means after a significant result in an analysis of variance (ANOVA), a tukey post-hoc test was applied using the function `TukeyHSD()` from the standard R package 'stats'. Post-correction of the p-value cutoff was conducted via Bonferroni-correction, that involves dividing the desired significance level (usually 0.05 or 0.01) by the number of comparisons being made to control the overall Type I

error rate. Heatmap clustering was performed using the clustering algorithm and the heatmap function of the standard R package 'stats'.

For analysis of precursor progression, the Mfuzz<sup>189</sup> R package with a soft clustering algorithm was used to identify early-stage and late-stage up- and downregulation. The soft cluster analysis was performed on precursor lesions from tumor-developing (TD) and tumor non-developing (TND) tissue transformation for which three stages were included (N, LG, HG) and a clustering limit of 5 cluster were used for soft cluster calculation. For Gene ontology analysis a fisher exact test was performed in Perseus on SoftCluster Scores of each Cluster to report significant enriched terms.

## 2.4.2 Glycoproteomics (Glyco-DIA)

Glycoproteomic data analysis was carried out with Spectronaut (18.5.231110.55695) with a spectral library recently published in Ye, Mao, Clausen and Vakhrushev<sup>10</sup>. Glyco-DIA libraries are available on the O-Glycoproteome database (<http://glycoproteomics.somee.com>) or on github (<https://github.com/CCGMS/Glyco-DIA>). To combine individual libraries the rbind() function of the base R package was used to create one in silico library containing TnTn, TnT, TTn, TT, STnTn, STnT, NakedTn, NakedT, mSTTn, mSTT, dSTTn, dSTT individual libraries. Localization of modification were not implemented to analysis via site-specific library. This merged Glyco-DIA library was directly imported to Spectronaut and PanIN and IPMN RAW files were searched against this spectral library. Results for differential analysis of protein candidates and glyco modification analysis were exported from Spectronaut and were further processed in R to produce relevant figures, such as BoxPlots (package: ggplot2).

# 3 Results

## 3.1 Experimental setup for proteomic analysis of PB precursor lesions

For the characterization of tumor developing (TD) and non-developing, (TND) tissue transformation within pancreatobiliary tissues, several patient biopsies were used for the microdissection of the relevant precursor lesions of pancreatic ductal adenocarcinoma or cholangiocarcinoma (PDAC or CCA). Therefore, two patient groups with atypical tissue changes were analyzed in this study, the patient group of TND progression (Group 1), that mainly consists of patients with inflamed pancreas or liver, like in conditions for pancreatitis or cholangitis and the patient group of TD progression (Group 2), that developed PDAC or CCA. A tissue biopsy of the remaining tissue of each patient group are used for morphomolecular grading after FFPE treatment, followed by microdissection of 2.5 mm<sup>2</sup> of each tissue grade (Normal = N, low-grade = LG, high-grade = HG, Tumor = T). The general procedure is shown in *Figure 13*. The following precursor lesion types were included in this study: 1.) pancreatic intraepithelial neoplasm (PanIN), 2.) intraductal papillary mucinous neoplasms (IPMN) from gastric and

intestinal tissue type, 3.) biliary intraepithelial neoplasm (BillN), 4.) intraductal papillary neoplasms of the bile duct (IPNB) from gastric, intestinal and pancreatobiliary tissue type.

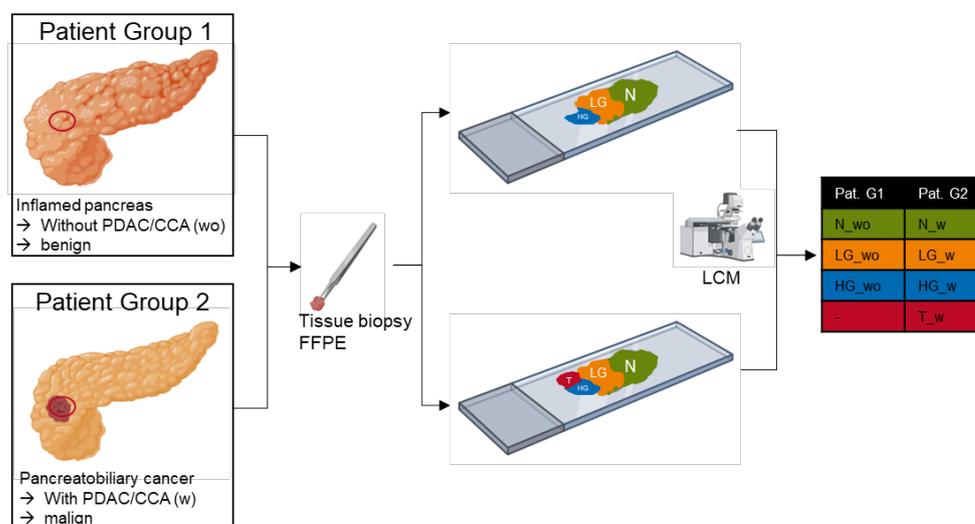


Figure 13: Schematic overview of the experimental setup for the current study (created with BioRender.com). Two patient groups with tissue abnormalities divided to a TND group (Group 1) without PDCA/CCA (wo) but with inflamed pancreas and a TD group (Group 2) that developed PDAC or CCA. For both patient groups tissue biopsies were taken, FFPE treated and underwent LCM. The remaining sample groups can be divided to the grades: Normal (N), low-grade (LG), high-grade (HG), tumor (T), with a TND progression (wo) or a TD progression (w).

With the current setup 5,123 proteins were identified in a DIA-NN search, using the UniProt protein database (81,837 entries), after filtering to >2 peptides and removal of contaminants. Overall, 179 samples are analyzed in a database search, including 57 PanIN samples, 53 IPMN samples, 44 BillN samples and 25 IPNB samples with different grades of lesion progression (normal, low-grade, high-grade, tumor, metastasis), listed in Table 8.

Table 8: Sample table of all samples used in this study. The number of microdissected precursor lesions used in this study in relation to all microdissected precursor lesion. Grades are divided to normal (N), low-grade (LG), high-grade (HG), Tumor (T) and metastasis (Met) with lesions that developed a tumor (w = with PDAC/CCA) or not (wo = without PDAC/CCA). Precursor lesions from PDAC are PanIN from pancreatic tissue type IPMN from gastric and intestinal tissue type. Precursor lesions from CCA are BillN from pancreatobiliary tissue type and IPNB from gastric, intestinal and pancreatobiliary tissue type.

		Number of valid samples / number of all microdissected lesions								
Lesion	Tissue type	N (w)	N (wo)	LG (w)	LG (wo)	HG (w)	HG (wo)	T (w)	Met (w)	Overall
<b>PanIN</b>	Pancreas	9/9	10/10	10 /10	10/10	6/7	1/1	9/10	0/0	55/57
<b>IPMN</b>	Gastric	6/6	3/3	4/4	0/4	6/6	0/1	6/6	0/0	25/30
<b>IPMN</b>	Intestinal	2/2	6/6	2/2	0/6	4/4	0/1	2/2	0/0	16/23
<b>BillN</b>	Pancreatobiliary	1/4	1/8	3/5	5/8	4/6	2/4	3/6	0/3	19/44
<b>IPNB</b>	Gastric	0/0	0/1	0/0	0/2	1/1	0/1	0/1	0/0	1/6
<b>IPNB</b>	Intestinal	0/1	0/4	0/0	1/1	2/3	3/4	2/3	0/0	8/16

---

<b>IPNB</b>	Pancreatobiliary	0/1	0/0	0/0	0/0	1/1	0/0	0/1	0/0	1/3
-------------	------------------	-----	-----	-----	-----	-----	-----	-----	-----	-----

---

### 3.2 Cluster analysis of precursor lesions from different PB tissue types

To understand the development of pancreatobiliary precursor lesions and to identify cancer-specific conditions, cluster analysis emerges a way to identify molecular similarities underlying lesion progression, as well as tissue and lesion type similarities and/or differences. For the cluster analysis all samples that pass outlier analysis are used and all missing values are filtered out. The results of cluster analysis are shown in the heatmap (Figure 14) for which distinct clustering of normal tissue and tumor tissue and mixed clusters with low-grade and high-grade precursor lesions can be observed. These mixed clusters of precursor lesion show a tendency of clustering in a TD cluster (w) and a TND cluster (wo).

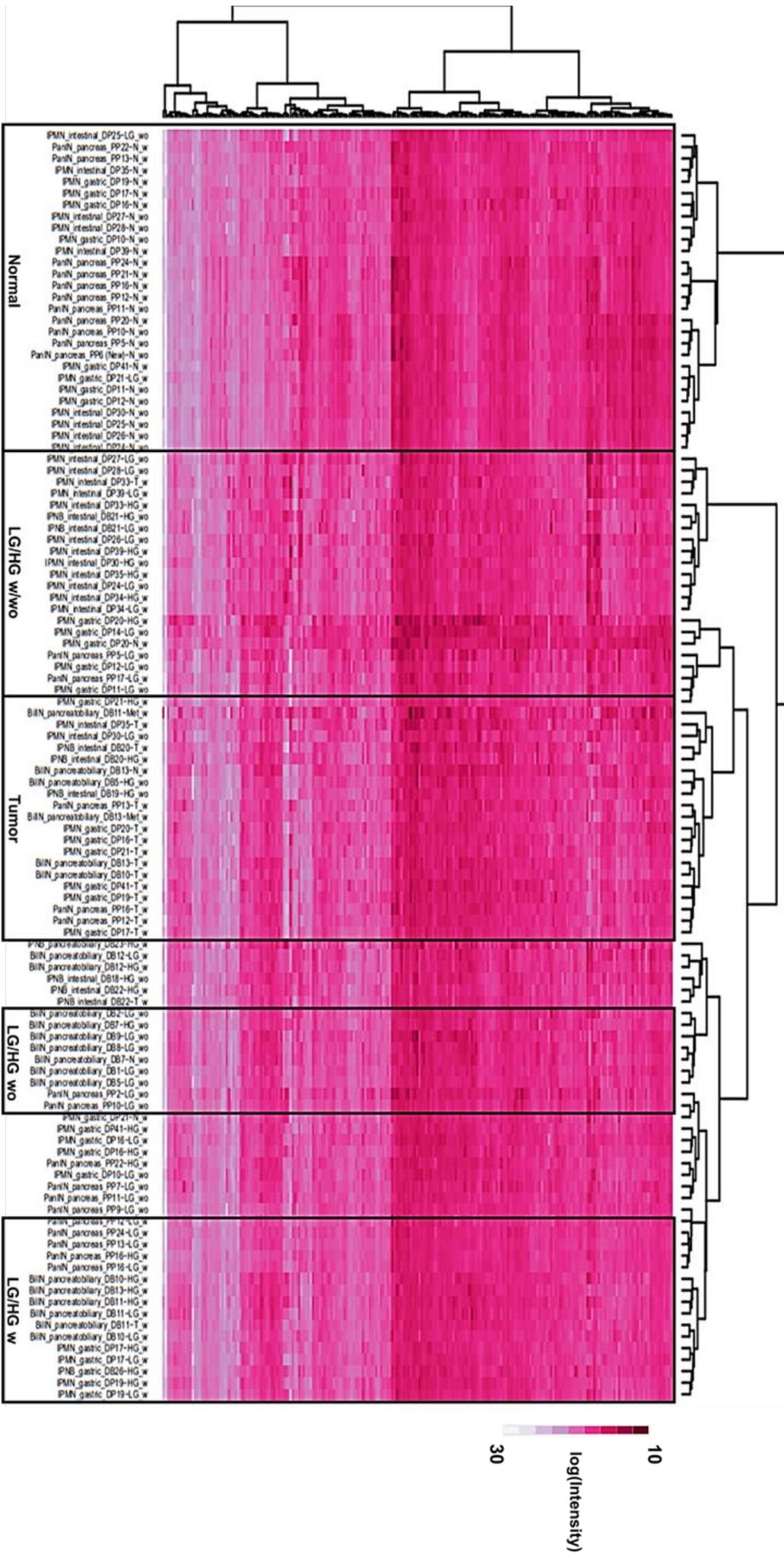


Figure 14: Heatmap clustering (method: ward.D1) of all precursor lesions from pancreaticobiliary tissue transformation. Grades are divided to normal (N), low grade (LG), high grade (HG) and Tumor (T) lesion grade. Some precursor lesions developed a tumor (w = with PDAC/CCA) or not (wo = without PDAC/CCA). Precursor lesions from PDAC are PanIN from pancreatic tissue type IPMN from gastric and intestinal tissue type. Precursor lesions from CCA are BiIN from pancreaticobiliary tissue type and IPNB from gastric, intestinal and nonpancreaticobiliary tissue type

### 3.3 SoftCluster profiles of TD and TND tissue progression

For analysis of precursor progression, a soft clustering algorithm is used to identify early-stage and late-stage up- and downregulation. The obtained cluster profiles are used to identify co-expression and/or co-regulation of proteins, involved in several biological pathways. The soft cluster analysis is performed on precursor lesions from TD and TND tissue transformation for which three stages are included (N, LG, HG) for each progression type (TD, TND) and a clustering limit of 5 cluster is used for soft cluster calculation. These 5 cluster profiles are classified into early downregulation, late downregulation, early upregulation, late upregulation and one cluster with constant upregulation.

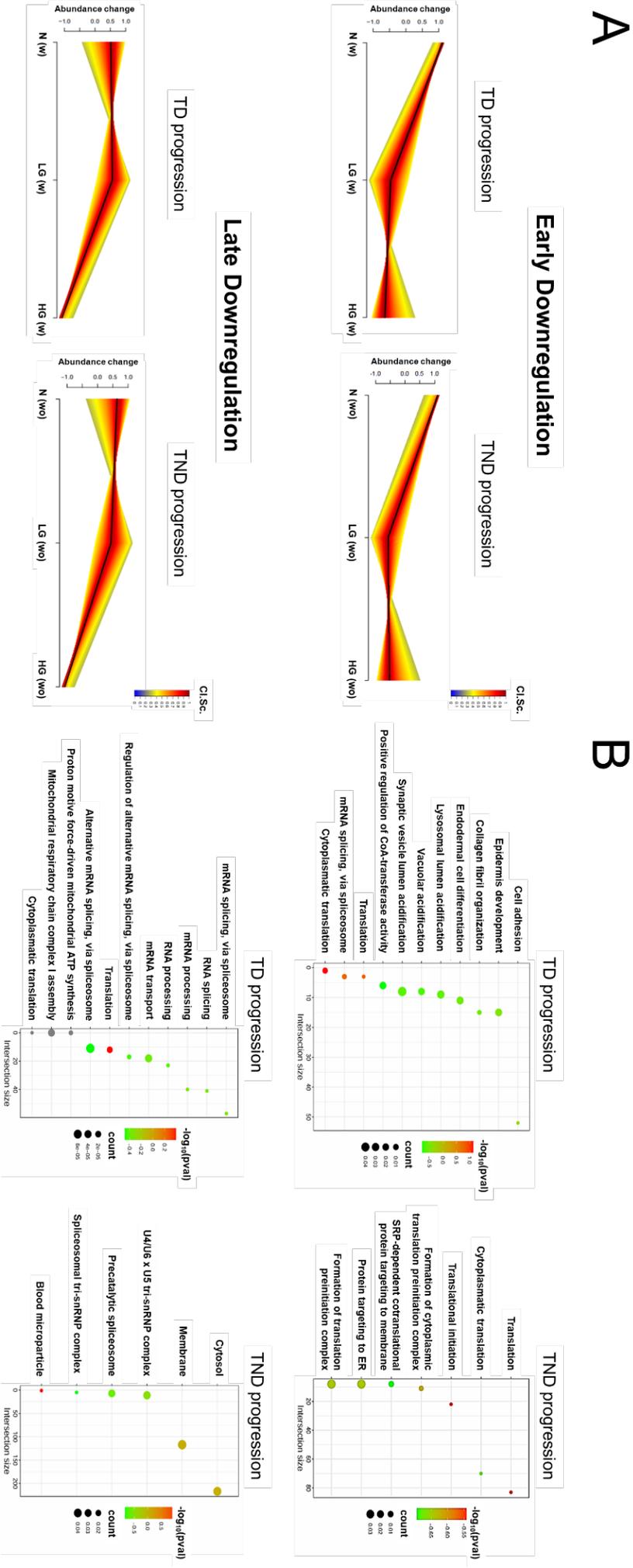
#### 3.3.1 Early and late downregulated pathways

Several biological pathways involved in Cytoplasmic translation, formation of translation preinitiation complex and translational initiation and protein targeting to ER are early downregulated in TND tissue transformation with high count and low p-value (Figure 15). In TD tissue transformation cytoplasmic translation is like in the TND progression early downregulated with low p-value, but in difference to the TND progression other biological pathways like mRNA splicing, positive regulation of CoA-transferase activity, lysosomal, vacuolar and synaptic vesicle lumen acidification are early downregulated in TD progression. Other structural relevant pathways like cell adhesion, epidermis development, collagen fibril organization and endodermal cell differentiation are also early downregulated uniquely in TD progression of early downregulation. For the progression profiles of late downregulated pathways in TD tissue transformation several pathways involved in RNA/mRNA splicing and regulation via spliceosome and mRNA transport follows this trend. Furthermore, pathways involved in the energy supply, such as proton motive force-driven mitochondrial ATP synthesis and the assembly of the mitochondrial of respiratory chain complex I are enriched in the late downregulation profile of TD progression. For the late downregulation of TND progression, no GO process terms are enriched, but GO components, like cytosol, membrane, blood microparticle and three locations in protein complexes associated with the spliceosome.

#### 3.3.2 Early, late and constant upregulated pathways

For upregulation, three profiles are calculated in the SoftCluster analysis, one for an early upregulation, one for late upregulation and one for constant upregulation (Figure 16). While pathways involved in cytoplasmic translation are enriched in the early downregulation profiles of both, TD and TND progression, some pathways regarding cytoplasmic translation are also early upregulated. While these pathways are subject to rigorous regulation and can contribute to various biological outcomes, these findings appear contradictory. This is because the majority of translational pathways are early downregulated, while some are early upregulated in the process. Translational processes and components appear to be involved in both, TD and TND progression, suggesting a similar effect on tissue transformation and the resulting effects of these pathways.

Figure 15: (A) Soft-clustering results of early and late downregulated pathways in TD and TND progression from normal tissue to low grade and high-grade precursor lesions. (B) Enrichment analysis of biological pathways that are enriched in the different progression profiles of early and late downregulated proteins.

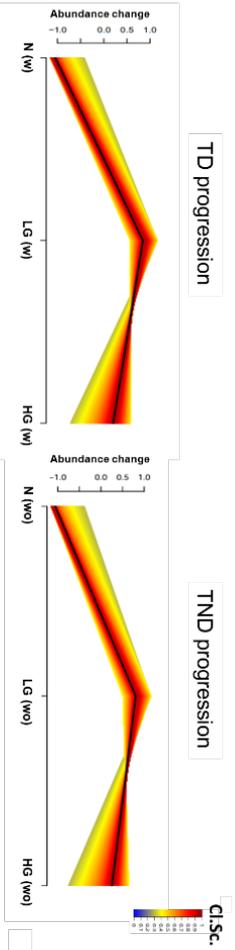


Variations in the early upregulation patterns between TD and TND progression involve pathways linked to the immune system, such as the innate immune response, defense response to bacteria and fungi, complement activation, and positive regulation of B-cell activation. For the TD progression with lowest p-value and highest strength endoplasmic reticulum to Golgi vesicle-mediated transport is early upregulated. These pathways are notably enriched in early upregulation profiles of TD progression and are not enriched in TND progression. For the TND progression the uniquely early upregulated pathways are linked to the actin cytoskeleton and collagen-containing extracellular matrix, which is also observed for the late upregulation profile of the TND progression. With low p-value mitochondrial proteins are late upregulated in TND progression. Pathways related to blood microparticle appear in late TND progression. Translational processes again appear in both, TND and TD progression of late upregulated pathways. In TD progression, more complex biological pathways, such as endoplasmic reticulum to Golgi vesicle-mediated transport, mRNA splicing and protein targeting to ER are late upregulated in the TD progression, suggesting as the turning point to proliferation. Protein (de)neddylation and N-glycosylation turn out as important modification involved in the regulation of proliferative behavior, because these pathways are also late upregulated in TD progression only. Furthermore, cell adhesion is late upregulated in TD progression with lowest p-value, indicating a high influence in proliferation. Constant upregulated in TD progression are many pathways involved in cellular respiration, such as aerobic respiration, proton motive force-driven ATP synthesis and several pathways involved in assembly and electron transport of mitochondrial, respiratory chain complex. All these pathways are uniquely enriched in the TD constant upregulation profile and do not appear in any upregulated (early, late, constant) profile of the TND progression. Due to the constant upregulation of ATP synthesis, TD cells produce high resources of energy, possibly used for proliferation and cell movement in cancer progression. Cytoplasmic translation, either important in the late upregulation profile, indicating a very complex interaction of several translational processes in both, the TD and the TND progression. Uniquely constant upregulated pathways in TND progression are nuclear migration, retrograde transport vesicle recycling within Golgi and several pathways involved in the innate immune answer, such as complement activation and B-cell receptor signaling pathway. SoftCluster analysis reveals inside into progression of TD and TND tissue transformation, showing high influence of translational processes in all profiles for both progression types. Thereby, many biological pathways undergo similar regulation, showing the high similarity of TND conditions, such as pancreatitis and cholangitis with TD conditions, such as pancreatic or hepatic cancer. Nonetheless for all profiles TD progression have higher biological activity due to the more complex results in network analysis and more enriched different biological pathways. The late upregulation profile can give insights to biological processes involved in the turning point to proliferation and shows that protein modifications, such as protein (de)neddylation and N-glycosylation and vesicle-mediated protein targeting from/to ER and/or Golgi seem to play an important role in TD progression and tumor development.

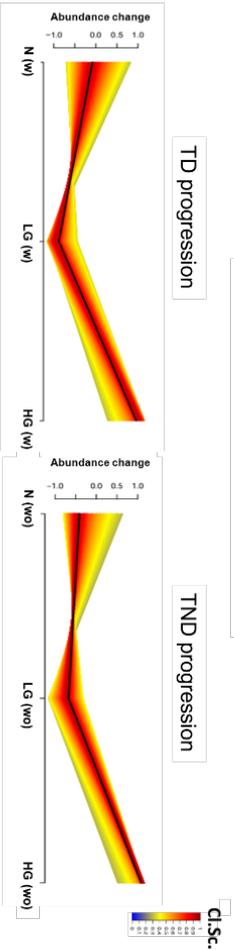
Figure 16: (A) Soft-clustering results of early, late and constant upregulated pathways in TD and TND progression from normal tissue to low grade and high-grade precursor lesions. (B) Enrichment analysis of biological pathways that are enriched in the different progression profiles of early and late downregulated proteins.

A

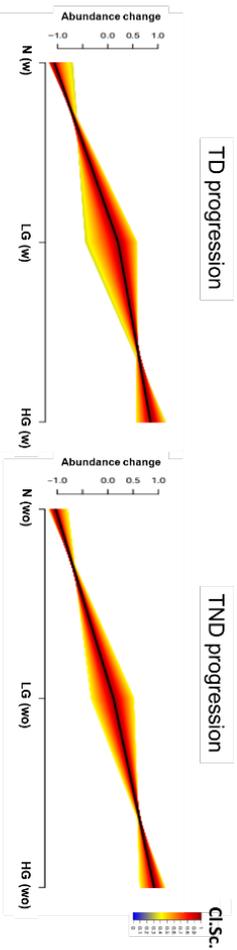
### Early Upregulation



### Late Upregulation

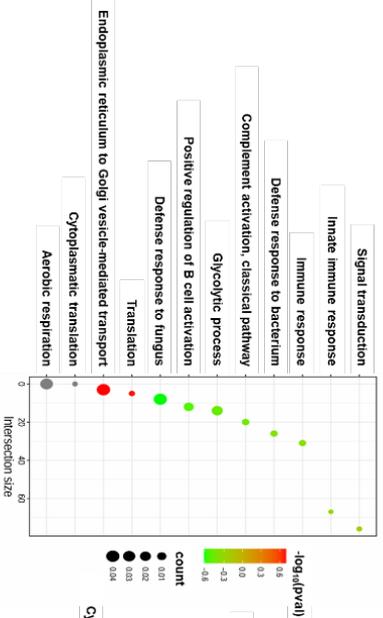


### Constant Upregulation

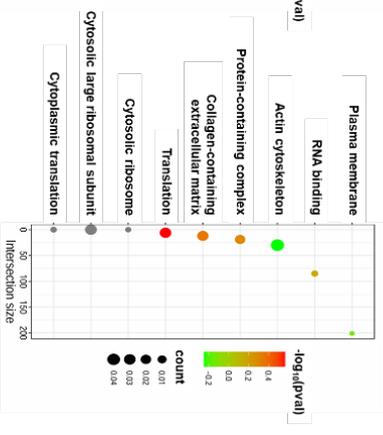


B

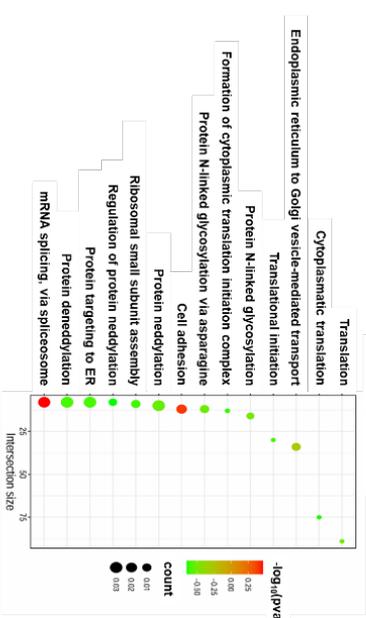
#### TD progression



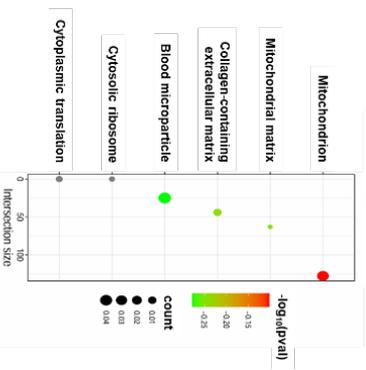
#### TND progression



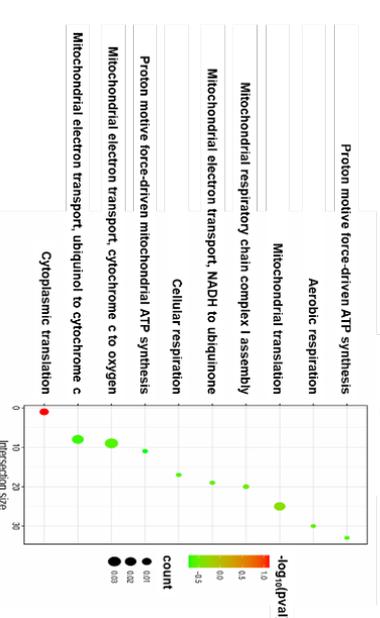
#### TD progression



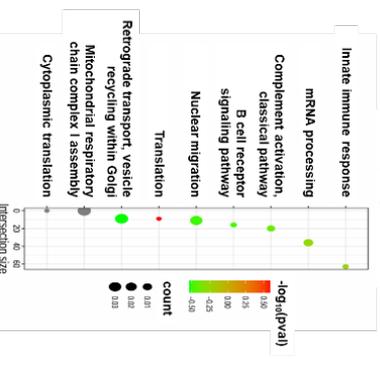
#### TND progression



#### TD progression



#### TND progression



### 3.4 Comparison of early events in TD and TND tissue transformation

Analysis of variance (ANOVA) reveals insights into several comparisons between the analyzed precursor lesions in this study. Early events in tissue transformation can be analyzed by comparing low-grade tissue protein abundances with the protein abundances found in normal tissue and can be used as a starting point for biomarker investigation. Differential regulated proteins are then compared between each progression type (TD, TND) and divided to upregulated and downregulated groups of proteins (Figure 17). For the TD progression 264 significant upregulated and 236 significant downregulated proteins are observed. For TND progression 445 significant upregulated and 399 significant downregulated proteins are observed. Comparison of TND and TD, significant upregulated proteins revealed 95 uniquely significant upregulated proteins in TD progression, that further underwent network analysis. Network analysis (created with ClueGO in Cytoscape, shown in Figure 17) of these early progression events uniquely upregulated in the TD progression reveal three connected pathway networks. One of these network pathways relate to nucleoside diphosphate metabolic processes, including nucleoside/nucleotide diphosphorylation, purine (ribo)nucleoside diphosphate metabolism, ADP metabolism and generation of ATP by pyruvate metabolic processes, pathways that were also observed in SoftCluster analysis for the constant upregulation profile of TD progression. While energy delivery is connected to carbohydrate metabolism and catabolism, these pathways are connected to the nucleoside diphosphate metabolism pathway. The second cluster of glycolytic processes, such as glucose-6-phosphate metabolism, pentose biosynthesis, oxidative pentose-phosphate shunt and hexose catabolism are processes uniquely upregulated in TD progression. Observations from early and late upregulated SoftCluster profile results support these observations of higher glycolytic activity from differential analysis, because in these, many glycolytic processes are activated in the TD progression only. The third network cluster contains pathways including alternative mRNA splicing and processing via spliceosome and its regulatory processes. SoftCluster analysis contributes to these results, due to the high relevance of mRNA splicing processes in several softclustering profiles (early/late downregulation and late upregulation) of the TD progression. Herein alternative mRNA splicing is found in both progression types and in upregulation and downregulation profiles, suggesting high regulative interactions of several splicing-associated proteins. Smaller cluster contains processes related to positive regulation of Notch signaling pathway, negative regulation of intrinsic apoptosis in response to DNA damage, regulation of RNA transcription from RNA polymerase II promotor in response to stress, positive regulation of mitophagy in response to mitochondrial depolarization, glutathione metabolism and activation of innate immune response. Results of the differential analysis of low-grade precursor lesions with the normal tissue to differ between TD and TND tissue transformation contributes several findings of the SoftCluster analysis and point out the most important biological pathways. The main biological differences between TD and TND lesion progression are related to pathways such as oxidative phosphorylation and ATP synthesis, alternative RNA splicing and glycolytic processes.





Further indications, like increased vesicle-mediated transport, Notch signaling and protein targeting to ER can relate to higher cell-cell communication activity in TD progression. Negative regulation of apoptotic signaling pathway in response to DNA damage could explain the switch to proliferative behavior of TD progression.

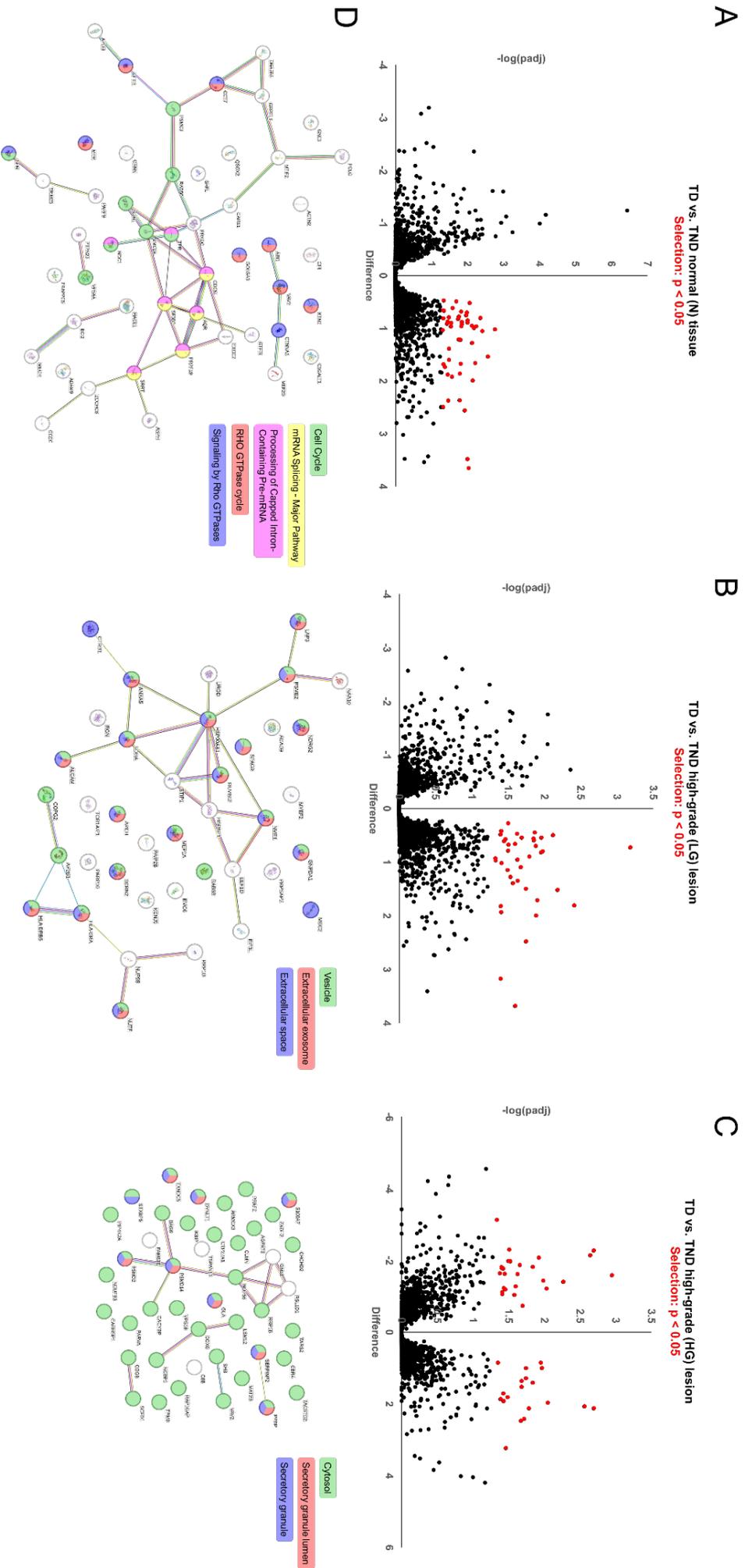
### 3.5 Comparison of corresponding TD and TND tissue grades

Another differential analysis arising from ANOVA is the comparison of the corresponding tissue grades in TD or TND progression and the results are shown in Figure 18. SoftCluster analysis showed high similarity of the most pathways during lesion progression whereby differential analysis results in a very small number of significant, differential regulated proteins. For biomarker investigation the focus lays on significant upregulated proteins in TD progression and for mainly secreted proteins, that are suitable to detect in body liquids and/or blood. Figure 18 shows the results of the differential analysis, focusing on upregulated proteins in TD progression for each grade (Normal, LG, HG), with the corresponding network analysis results from string-db.org. Only for the comparison of normal TD and TND tissue the network analysis revealed insights to GO biological processes, while for the low-grade and the high-grade comparisons only GO components are indicated. Therefore, this comparison reveals only limited insights to the biological pathways involved in TD progression. The main upregulated pathways regarding normal tissue on TD or TND progression are related to the cell cycle, mRNA splicing and Rho GTPase signaling. These pathways play a huge role in cell division and proliferation and are known to be involved in cancer-related processes. The GO components found for LG and HG comparisons of TD and TND lesions show upregulation of proteins involved in secretion, such as vesicles, extracellular exosomes and secretory granule. Higher secretory activity in TD progression holds the chance to find a secreted biomarker, which simplifies diagnostics.

#### 3.5.1 Significant upregulated proteins in TD progression

All of the significant upregulated proteins in TD progression are not significant upregulated at all grades, which would make them suitable for further biomarker validation. Some proteins switch the regulation from upregulated in normal to downregulated in low-grade and to upregulated in high-grade and this course excludes them as appropriate biomarkers as well. For some proteins the level is significant upregulated in TD normal tissue, but during lesion progression, protein levels equalize between TD and TND precursor lesions. All these proteins do not carry the potential for sufficient diagnostics. Upregulated proteins, that have higher levels in TD normal tissue (not significant) and the difference increases during lesion progression and they turn into significant upregulated proteins in TD low-grade and/or high-grade lesions, are more interesting candidates for biomarker validation. Another important criterion is that the candidates need to get identified with high confidence and low p-values.

Figure 19: Differential analysis of corresponding TD and TND tissue grades. A) Volcano Plot of ANOVA results from TD normal tissue to TND normal tissue. B) Volcano Plot of ANOVA results from TD low-grade lesion to TND low-grade lesion. C) Volcano Plot of ANOVA results from TD low-grade lesion to TND low-grade lesion. D) Network analysis of the significant ( $p < 0.05$ ) upregulated proteins in TD tissue grade



Protein candidates, which meet these criteria are shown in Table 9 and are checked on [proteinatlas.org](http://proteinatlas.org), if there is already literature evidence for these markers. All of the found protein candidates are already validated as markers (not prognostic) for many different cancers, such as pancreatic cancer, via RNA expression data and antibody staining (mAb) except Proteasome activator complex subunit 2 (PSME2), which is only validated via RNA expression data. Protein O-glucosyltransferase 1 (POGLUT1) is a candidate coming from differential analysis, because it is found for tissues from TD progression for all grades and not in TND progression in any grade. The complete absence of POGLUT1 in tissues from TND progression in our data makes it a selective marker for tumor. Reported data from antibody staining of POGLUT1 in pancreatic cancer tissues shows low staining efficiency and is in many cases not detected in pancreatic cancer tissue, which can be supported by our data. POGLUT1 is only found in IPMN precursor lesions and is not detected in PanIN precursor lesions, as well as rarely detected in tumor samples. It is an ER related protein and could possibly explain regulatory biological processes, such as higher activity of protein O-glycosylation in TD progression. Another interesting candidate arising from differential analysis and also related to ER, is ER membrane protein complex subunit 6 (EMC6), for which the protein abundance level is increased for all grades of TD compared to TND progression with sufficient number of valid values. EMC6 is not prognostic for any cancer type, because there is a 'low consistency between antibody staining and RNA expression data' (<https://www.proteinatlas.org/ENSG00000127774-EMC6/pathology>), even though increased expression is very specific for many cancer types.

Table 9: Biomarker candidates for early detection of TD tissue progression. The Column 'Protein atlas' indicates, if the entry is already validated as marker for pancreatic cancer via RNA expression data or monoclonal antibody staining (mAb) at [proteinatlas.org](http://proteinatlas.org).

Gene name	Entry	Protein names	Protein atlas
<b>PSMD2</b>	Q13200	26S proteasome non-ATPase regulatory subunit 2	RNA /mAb
<b>PSME2</b>	Q9UL46	Proteasome activator complex subunit 2	RNA
<b>RUVBL2</b>	Q9Y230	RuvB-like 2	RNA /mAb
<b>CARHSP1</b>	Q9Y2V2	Calcium-regulated heat-stable protein 1	RNA /mAb
<b>AP2B1</b>	P63010	AP-2 complex subunit beta	RNA /mAb
<b>NOP56</b>	O00567	Nucleolar protein 56	RNA /mAb
<b>BAG6</b>	P46379	Large proline-rich protein BAG6	RNA /mAb
<b>TACSTD2</b>	P09758	Tumor-associated calcium signal transducer 2 (Pancreatic carcinoma marker protein GA733-1)	RNA /mAb
<b>DDX6</b>	P26196	Probable ATP-dependent RNA helicase DDX6 (Oncogene RCK)	RNA /mAb
<b>TPM3</b>	P06753	Tropomyosin alpha-3 chain	RNA /mAb
<b>GNL3</b>	Q9BVP2	Guanine nucleotide-binding protein-like 3	RNA /mAb
<b>EMC6</b>	Q9BV81	ER membrane protein complex subunit 6	RNA /mAb
<b>TARS2</b>	Q9BW92	Threonine--tRNA ligase, mitochondrial	RNA /mAb
<b>ACAD9</b>	Q9H845	Complex I assembly factor ACAD9, mitochondrial	RNA /mAb
<b>MAT2B</b>	Q9NZL9	Methionine adenosyltransferase 2 subunit beta	RNA /mAb
<b>POGLUT1</b>	Q8NBL1	Protein O-glucosyltransferase 1	RNA /mAb
<b>ST6GAL1</b>	P15907	Beta-galactoside alpha-2,6-sialyltransferase 1	RNA /mAb

The data reliability assessment of EMC6 on proteinatlas is pending on external verification. EMC6 is detected in transcriptomic data with a current cut off of 3.21 FPKM and a 5-years overall survival rate of 47% for patients with higher expression than the expression cut off. Antibody staining of EMC6 on cancer tissues showed low staining efficiency, which results in the low consistency of the two validation methods. Data from our study support RNA expression data, because EMC6 is found to have higher levels in TD progression for all grades and is significant different in advances stages, such as for high-grade precursor lesions. Beta-galactoside alpha-2,6-sialyltransferase 1 (ST6GAL1) is another candidate arising from differential analysis, because it is, like POGLUT1, only observed in TD progression. Herein ST6GAL1 is only found in low-grade and high-grade TD precursor lesions and could explain increased sialylation, a common modification enriched in tumor tissues.<sup>190</sup> Nevertheless, evidence from proteinatlas.org shows that ST6GAL1 is not prognostic for pancreatic cancer and was rarely detected by antibody staining in cancer tissues. We can demonstrate, that ST6GAL1 expression is not found in normal tissues, as well as rarely found in tumor tissues and seems to be involved in TD progression but once the tumor tissue is developed, ST6GAL1 is rarely detectable anymore. Both, POGLUT1 and ST6GAL1, are proteins with glycosylation activity and influence glyco-modification type of several proteins involved in immune modulation and signaling pathways for cell-cell communication. The unique expression of POGLUT1 and ST6GAL1 in TD progression could result in different glycosylation patterns, that favors proliferative behavior and the modulation of the tumor microenvironment. The resulting glycoproteins are further analyzed on O-glycosylation patterns by the Glyco-DIA approach of Ye, Mao, Clausen and Vakhrushev<sup>10</sup> in the following chapter (3.6).

### 3.6 Protein O-glycosylation analysis via Glyco-DIA

The advantage of a DIA method is the analysis of nearly all detectable peptide-ions. This is the reason for the assumption, that modified peptides are even more collected in DIA compared to DDA. Therefore, a new approach is to directly characterize glyco-modified peptides from samples, without prior enrichment, but measured with a DIA method. Within the help of pre-measured spectral libraries containing these modified peptides, e.g., after enrichment, it is possible to detect these modified peptides in non-enriched DIA samples. In pancreatic cancer, O-glycosylation plays an important role in the generation of the T and Tn antigens.<sup>63-65</sup> Furthermore, O-glycosylation at serine/threonine residues is known as a phospho-site blocking mechanism, which can influence many regulatory processes.<sup>56-60</sup> In our proteome data we found higher abundances of proteins involved in O-glycosylation (POGLUT1) and sialylation (ST6GAL1) in TD progression, which consequences should further be validated by PTM analyses. Within the new approach of Glyco-DIA using the spectral libraries provided by Ye, Mao, Clausen and Vakhrushev<sup>10</sup> it is possible to get deeper insights to regulatory mechanisms in TD or TND tissue progression. One fact is, that glycosylation-patterns are highly dependent on different tissue types, why the sample set was reduced to only PDAC precursor lesions. Therefore, the PanIN and IPMN samples (Table 10) are further used for Glyco-DIA analysis to find differences in protein abundance and/or modification of glycosylated proteins to differ between TD and TND tissue progression. The

provided Glyco-DIA database contained 2,076 glycoproteins and 11,452 unique O-GalNAc type glycopeptides with the five most common core 1 O-glycan structures, such as Tn, T, SiaTn, SiaT and diSiaT.<sup>10</sup> Using this library for spectral-library based DIA search, overall, 372 glycoproteins were identified for the current sample set, shown in *Table 10*.

*Table 10: Sample table of all samples used for Glyco-DIA analysis. The number of microdissected precursor lesions used in this study in relation to all microdissected precursor lesion. Grades are divided to normal (N), low-grade (LG), high-grade (HG), Tumor (T) with lesions that developed a tumor (w = with PDAC/CCA) or not (wo = without PDAC/CCA). Precursor lesions from PDAC are PanIN from pancreatic tissue type IPMN from gastric and intestinal tissue type.*

Lesion	Tissue type	Number of valid samples / number of all microdissected lesions							Overall
		N (with)	N (wo)	LG (with)	LG (wo)	HG (with)	HG (wo)	T (with)	
<b>PanIN</b>	Pancreas	9/9	10/10	10 /10	10/10	6/7	1/1	9/10	55/57
<b>IPMN</b>	Gastric	6/6	3/3	4/4	0/4	6/6	0/1	6/6	25/30
<b>IPMN</b>	Intestinal	2/2	6/6	2/2	0/6	4/4	0/1	2/2	16/23

### 3.6.1 Cluster analysis of Glyco-DIA data

Analysis based on the highest abundant glycosylated proteins are used for differential analysis of TD and TND precursor lesions. All samples are further filtered to >50 valid values, which results in a higher number of outliers in the Glyco-DIA dataset in comparison to the whole proteome DIA dataset, but is expected due to the technical challenges of the analysis of glycosylated peptides. For clustering analysis, missing values are set to 0, because for modification analysis, besides the abundance level, another important information is the presence or absence of different glycoproteins. Cluster results of the distribution of protein Intensities coming from the Glyco-DIA dataset, for the current sample set (PDAC), are shown in the heatmap in *Figure 19*. Although the cluster results from the whole proteome dataset exhibit clear clustering based on grades, there is a slight inclination towards clustering within the TD (w) or TND (wo) cluster. Conversely, the Glyco-DIA cluster results demonstrate stronger differentiation between the TD and TND groups rather than among the different grades. Another observation is that especially on the level of normal tissue, more samples are excluded due to low identification rates of glycosylated proteins. This could be an indication that these proteins are higher glycosylated in TND normal tissue compared to TD normal tissue. Differential analysis within different precursor stages reveals insights to modification differences and/or differences on protein abundance level of O-glycosylated proteins.

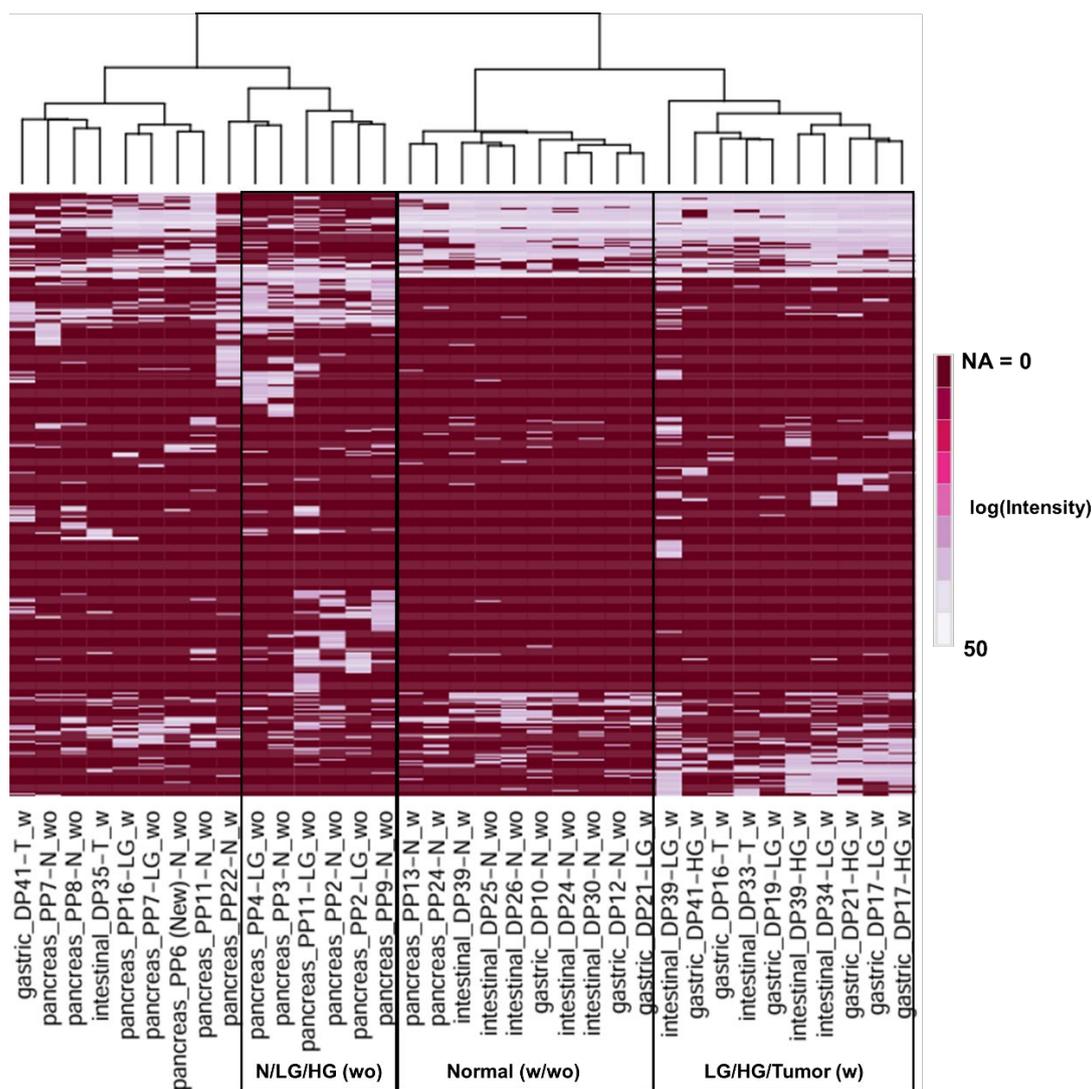


Figure 20: Heatmap clustering of proteomic results from Glyco-DIA analysis for the PDAC sample set.

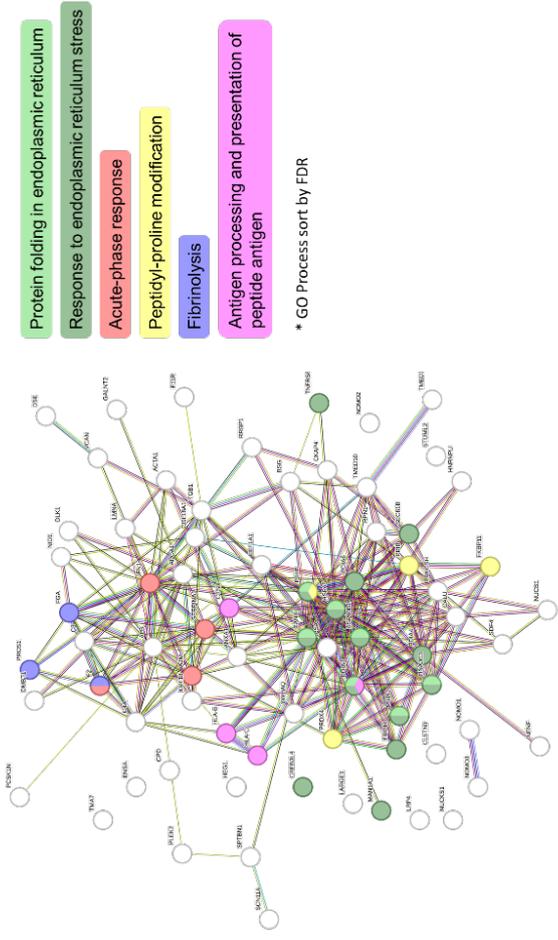
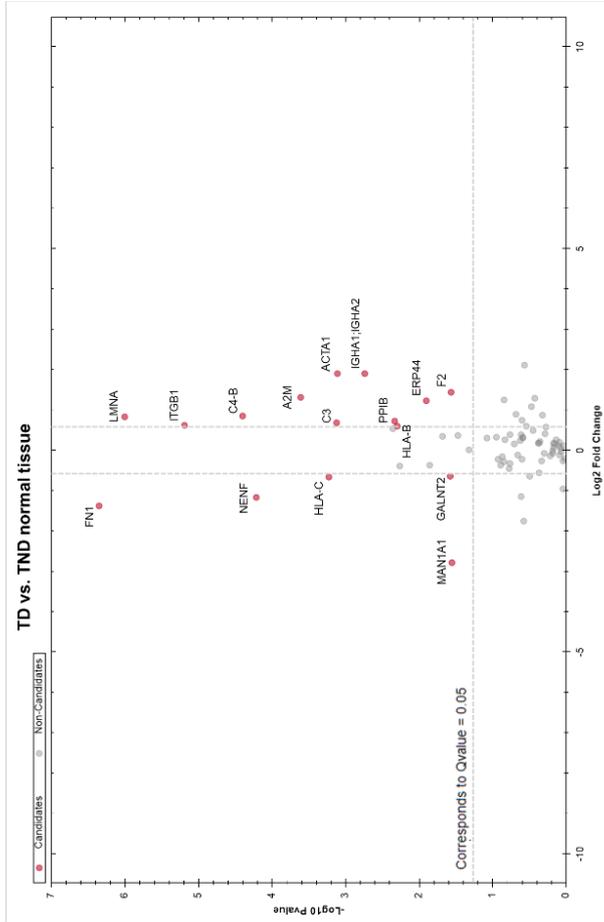
### 3.6.2 Differential analysis of Glyco-DIA data

Due to the small sample group size, only normal tissue and low-grade lesions are compared by differential analysis. Herein it is possible to compare protein levels of glycoproteins and the presence or absence of different modified peptides coming from these glycoproteins and to establish a relationship between these information. Using this approach many different regulated glycoproteins appear to be significant regulated. For normal tissue 16 different regulated glycoproteins are found to be significant differential abundant and 11 of these glycoproteins appear to be significant upregulated in TD normal tissue. Differential analysis of TD and TND low-grade lesions result in 21 significant deregulated proteins with 18 significant upregulated glycoproteins. Network analysis of all identified glycoproteins unveils the influence of biological pathways involved in tissue transformation on the level of glycoproteins and are shown in Figure 20. Herein, pathways involved in endoplasmic reticulum stress affecting protein folding in ER are found with low FDR for both comparisons.

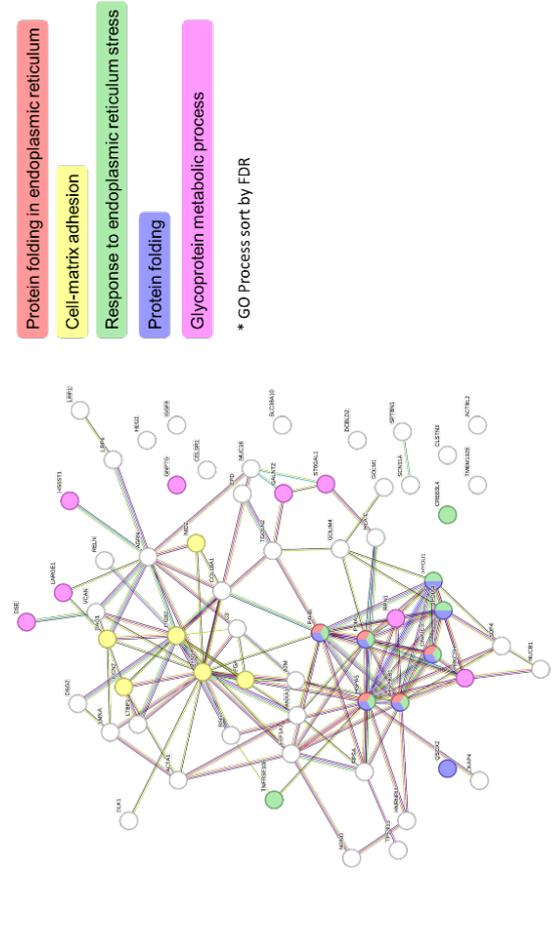
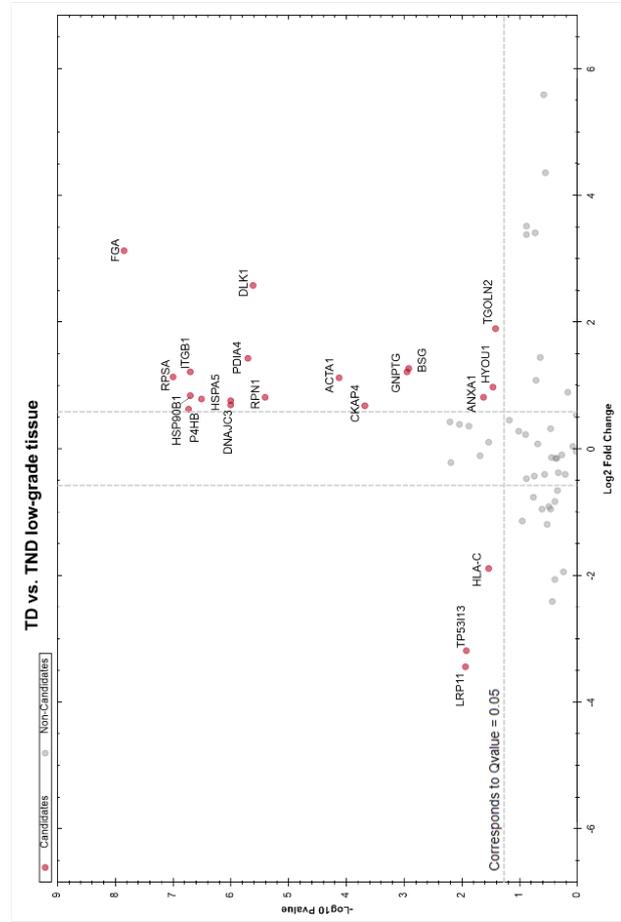
Figure 21: Protein candidates from differential analysis of Glyco-DIA data. A) Comparison of TD and TND normal tissue and the corresponding network analysis results. B) Comparison of TD and TND low-grade lesion and the corresponding network analysis results.

# A

## Protein Candidates



# B



Peptidyl-Proline modification is also involved in protein folding, affecting several different downstream processes, such as cell cycle regulation, DNA damage response etc. and is found in the Glyco-DIA data from comparison of TD to TND normal tissue. Pathways involved in acute-phase response including fibrinolysis are indicated to be involved in regulation for the comparison of TD and TND normal tissue. Pathways involved in the processing and presentation of peptide antigens have an impact on modulating tumor microenvironment and are indicated as regulated in the Glyco-DIA dataset on normal tissue grade. For the comparison of TD and TND low-grade precursor lesions the most significant upregulated glycoproteins are significant upregulated in TD progression. These proteins carry sufficient potential to be used as prognostic biomarker and can give insights to regulative processes modulated by glycoproteins. Besides protein folding and response to ER stress, cell-matrix adhesion and glycoprotein metabolic processes are indicated as differential regulated between TD and TND progression for the comparison on low-grade level. This indicated high regulatory activity in TD progression via O-glycosylation.

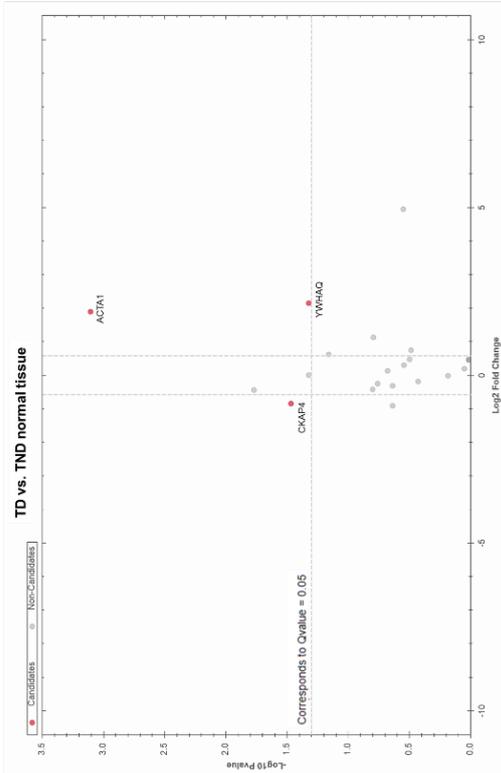
### 3.6.3 Differential modification analysis of Glyco-DIA data

To assess the impact of O-glycosylation on the regulation of tumor development in precursor lesions throughout disease progression, differential modification analysis is conducted. The presence of a glycoprotein in two different modification states can influence downstream processes and could reveal differences responsible for proliferative behavior. By analyzing modification difference, there is a bias if the protein abundance level also varies between TD and TND progression. Therefore, it would be recommended to focus on proteins whose abundance levels do not differ but exhibit a significant difference in peptide modification of this glycoprotein. Differential analysis on the modification level can further gets combined with the results coming from differential analysis on abundance level of the glycoprotein and can be visualized in a PTM fold change vs. PG fold change volcano plot (Figure 21). Glycoproteins with significant increase in one of the provided modifications from Glyco-DIA spectral libraries (HexNAc, Hex(1)HexNAc(1), Hex(1)HexNAc(1)NeuAc(1), Hex(1)HexNAc(1)NeuAc(2)) in TD progression are marked in red and all significant upregulated modifications on these glycoproteins are used for further analysis. Therefore, glycoproteins, that show increased overall abundance level are excluded for further modification analysis, because modification changes could be influenced by overall higher identification rate. These proteins are actin alpha skeletal muscle protein (ACTA1) and DnaJ homolog subfamily C member 3 (DNAJC3), which have increased protein abundance level and increased modification rate. Cytoskeleton associated protein (CKAP4) has decreased modification level on normal grade, but increased modification level for low-grade lesion grade and only shows differences in overall protein abundance for low-grade lesion. Due to the protein abundance level switch, which correlates to the modification switch for CKAP4, it is either excluded from further modification analysis. Observations concerning protein 14-3-3 theta indicate that there is a modification difference between normal grade and low-grade lesions.

Figure 22: A) Differential modification analysis of the comparison of TD and TND normal tissue and low-grade lesion. B) PTM fold change vs. PG fold change of the comparison of TD and TND normal tissue and low-grade lesion. C) Boxplot of protein abundance from four significant different modified glycoproteins (ACTA1, YWHAQ, DNAJC3, CKAP4) and their elevation in all grades.

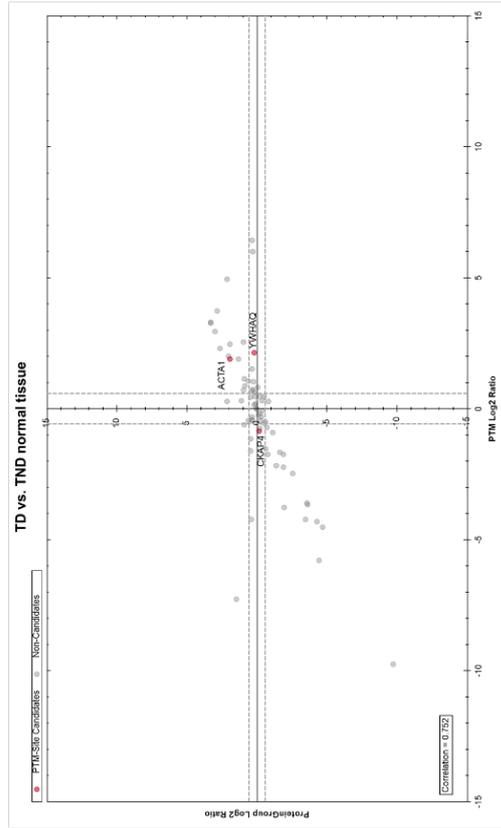
**A**

**Modification Candidates**



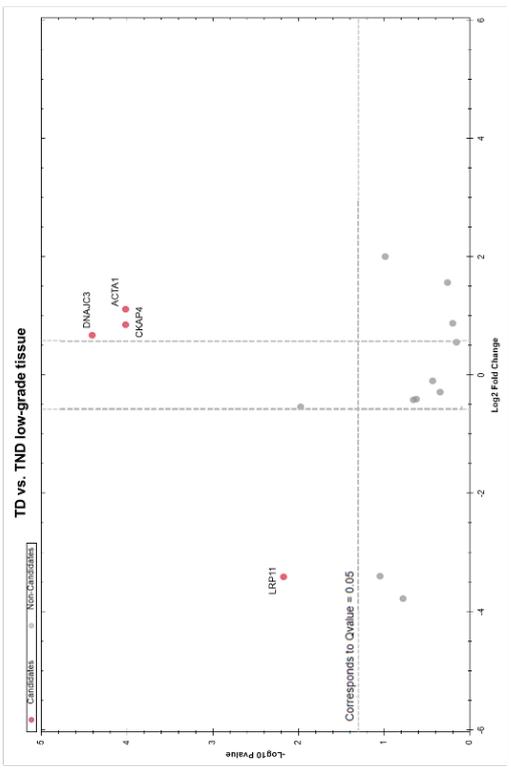
**B**

**PTM fold change vs. PG fold change**

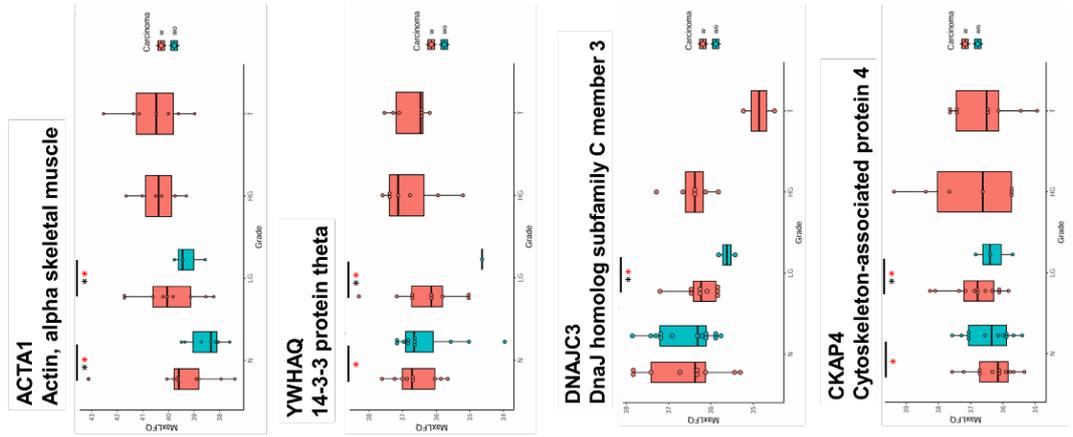
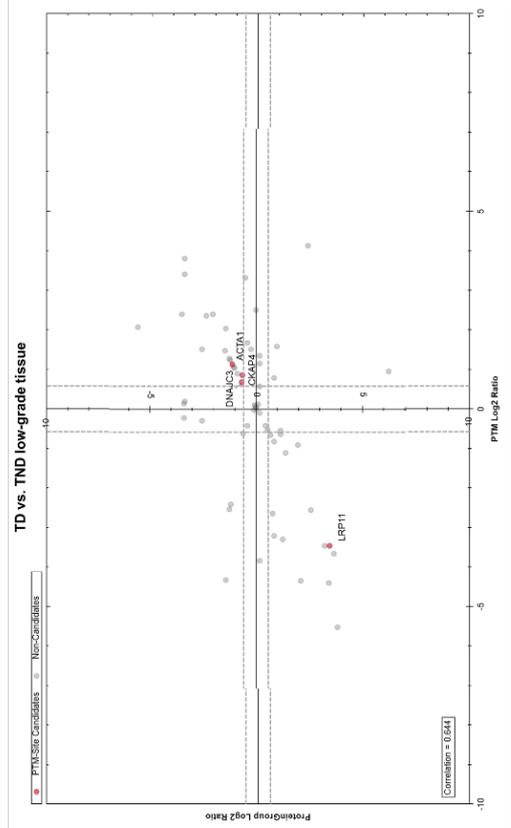


**C**

**TD vs. TND low-grade tissue**

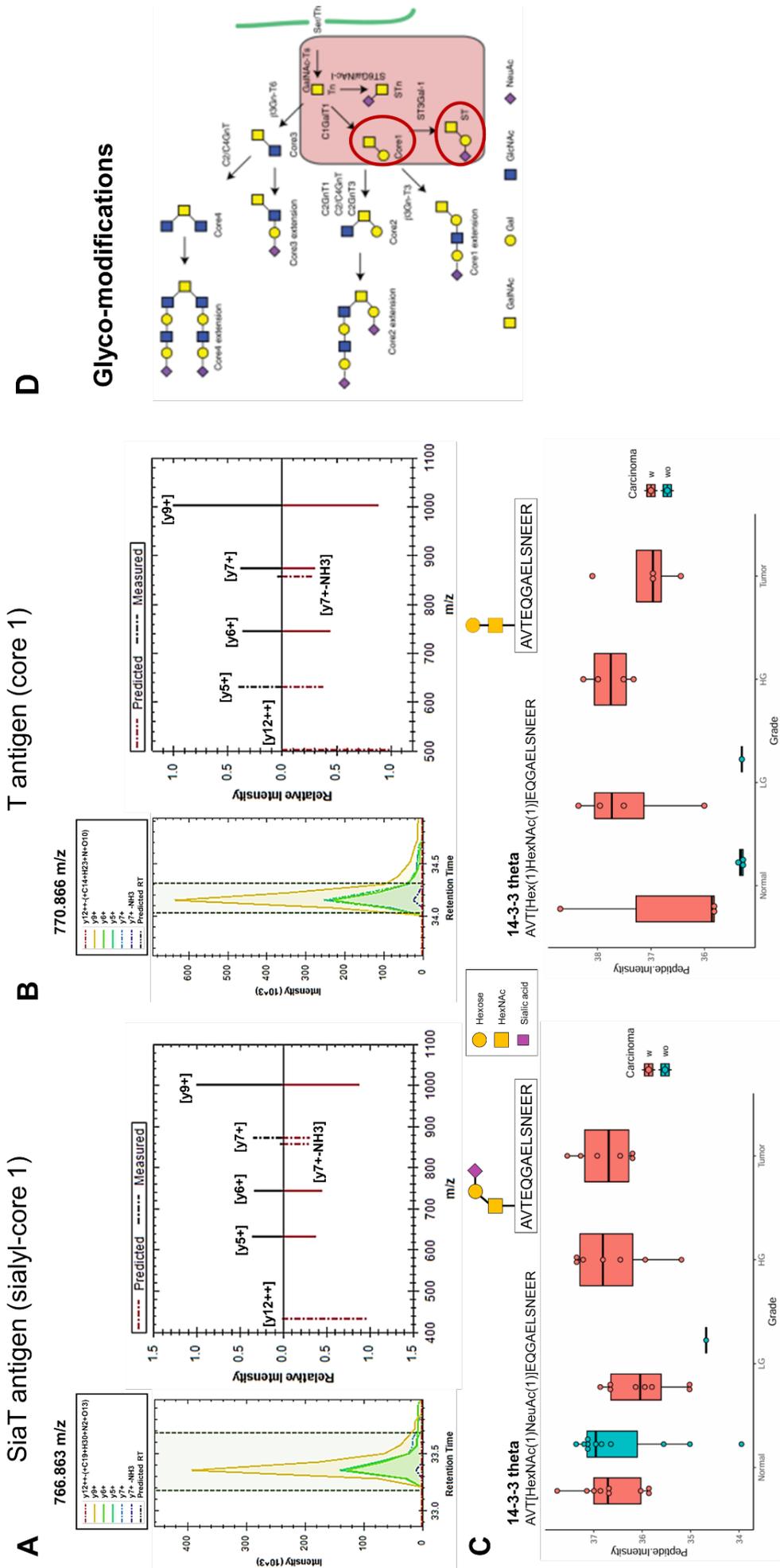


**TD vs. TND low-grade tissue**

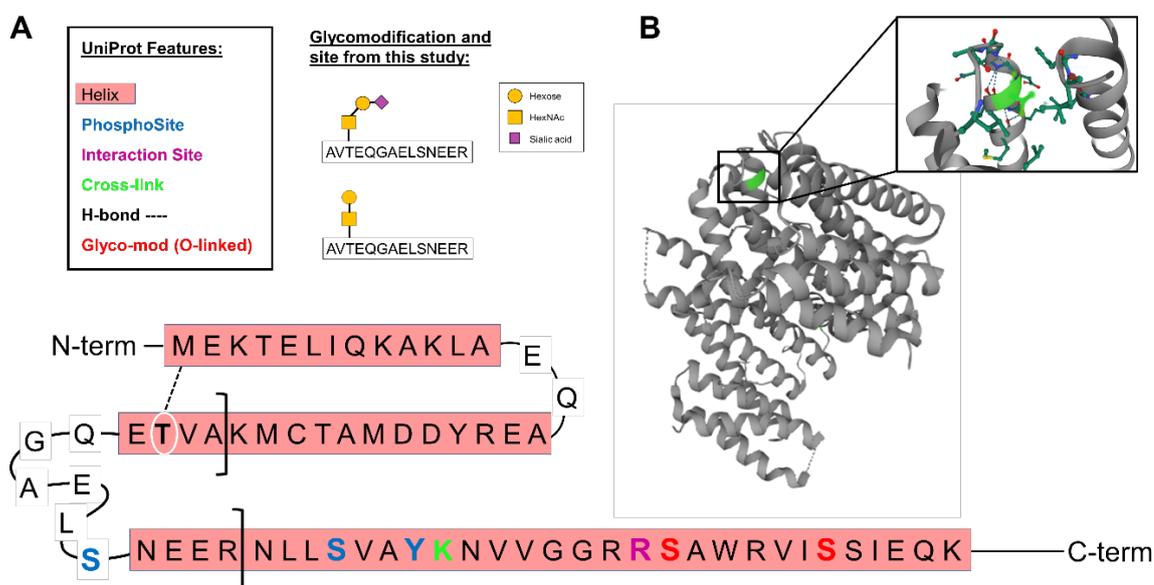


However, only at low-grade level, both overall protein abundance and the modification type differ for 14-3-3 theta. Therefore, a detailed analysis of the modified peptide sequence was conducted and reveals insights to modification type and quantity. It can be observed that the peptide AVTEQGAELSNEER, coming from 14-3-3 theta is carrying the sialyl-core 1 (HexNAc(1)NeuAc(1)) or the core-1 glycan (Hex(1)HexNAc(1)). The corresponding extracted ion chromatograms (XIC) and the MS2 intensity correlation plots are presented in Figure 22**Error! Reference source not found.** and confirm the identification of each modified peptide. Peptide intensities of each modified peptide are visualized in boxplots (Figure 22Figure 22C) to observe intensity differences between TD and TND grades and to figure out, which modification is uniquely present in TD progression. The results confirm higher accumulation of core 1 and sialyl-core 1 modification on 14-3-3 theta for all TD grades, except for sialyl-core 1 modification at normal grade. Especially for advanced grades, such as low-grade and high-grade lesions a strong difference in the identification of both modified peptides is observed. In the TD progression, the core 1 modification is consistently associated with high peptide intensities across all grades, whereas in TND progression, peptide intensities bearing the core 1 modification are low. The modification is found to be localized on Thr30 on a peptide region, that is involved in a helix structure motif, but also covers dynamic regions. Different modifications on structural regions within a protein can cause structural changes and may influence protein folding and the generation of protein-protein complexes, due to the modulation of different affinities.

Figure 23: A) Extracted ion chromatogram for a peptide mass of 766.863 m/z from 14-3-3 theta, which is identified as peptide carrying the sialyl-core 1 with the corresponding MS2 intensity correlation from predicted and measured fragment masses. B) Extracted ion chromatogram for a peptide mass of 770.866 m/z from 14-3-3 theta, which is identified as peptide carrying the core 1 with the corresponding MS2 intensity correlation from predicted and measured fragment masses. C) Boxplot for all Grades to visualize peptide intensities of each modified peptide in TD and TND progression. Illustration D for Glyco-modifications from Ye, et al.<sup>11</sup>



Furthermore, structural changes can block or enhance phosphorylation on a certain position, followed by different biological downstream processes. 14-3-3 theta is known to have binary interactions with 30 different proteins and to localize on many different subcellular locations (Golgi, mitochondria, ER, cytoplasm, extracellular exosomes, membrane). On BioGRID 1,200 interactors are reported for 14-3-3 theta ([YWHAQ Result Summary | BioGRID \(thebiogrid.org\)](#)). These observations show the manifold of different possible interactions with several different biological pathways involved. For 14-3-3 theta, 6 sites for O-linked glycosylation are reported on glygen.org (Ser57, Ser63, Thr69, Thr71, Thr72, Ser156). Literature evidence for O-glycosylation on Thr30 is not reported so far. Analysis of different features reported for the sequence segment nearby the glyco-modification could reveal possible structural influence of the modification site. In *Figure 23A* the sequence segment is shown with all structural features and the crystal structure of the whole protein is shown in B. It can be observed that Thr30 as modification site is involved in the formation of a  $\alpha$ -helix structure motif, parallel to the first  $\alpha$ -helix structure of 14-3-3, which is stabilized by Thr30 through hydrogen bond interactions to Met1. In this segment (1 – 68) three phospho-sites on S37, S45 and Y48 are reported on UniProt. An interaction site within the 14-3-3 protein at position R56 has been documented in the literature, where this site interacts with phosphoserine residues from proteins that are interaction partners from 14-3-3. Furthermore, one crosslink site is reported in this peptide segment of 14-3-3, which links with a glycy l lysine isopeptide of SUMO2 interchain. Possible changes in the modification on Thr30 may influence structural changes, which could influence binding affinities for interaction sites or induce enhancement or blockage of phosphorylation events on distinct phosphorylation sites and results in different types of downstream processes.



*Figure 24: A) Structural features involved in the sequence segment (1 – 68) in which the modified sequence (AVT(Glyco)EQGAELSNEER) is included and for which the glyco-modification difference is observed. B) Crystal structure of 14-3-3 with the highlighted site, where the glyco-modification is observed.*

## 4 Discussion

Proteomic analysis of pancreatobiliary precursor lesions is challenging due to several different cell types involved in these tissue sections. In the past, proteomic results for FFPE tissue sections arising from these precursor lesions are always biased by signals from different cell types such as acinar, endocrine, stromal and inflammatory cells. Within new sensitive methods, including distinct morphomolecular subtyping combined with laser microdissection to isolate distinct cell types, followed by sensitive tissue disruption and sample preparation methods for LC-MS/MS analysis involving the use of data independent acquisition enables precise analysis of biological pathways involved in lesion progression. Translational processes are highly regulated in the progression of precursor lesions with several pathways indicated as downregulated but on the other hand there are indices that some translational processes, often also associated with viral translation are upregulated. Biological pathways that play an important role in tumor progression and development are protein (de)neddylation, N-glycosylation and vesicle-mediated protein targeting from/to ER and/or Golgi. The primary biological distinctions between the progression of TD and TND lesions are associated with pathways such as oxidative phosphorylation and ATP synthesis, alternative RNA splicing, and glycolytic processes. Additionally, indications such as increased vesicle-mediated transport, Notch signaling, and protein targeting to the endoplasmic reticulum may be linked to increased cell-cell communication activity in TD progression. The downregulation of apoptotic signaling pathways in response to DNA damage could elucidate the transition to proliferative behavior in TD progression. Several protein candidates are observed as marker for TD progression, even though there is already literature evidence on [The Human Protein Atlas](#), including RNA expression data and mAb staining on tumor tissue for all observed candidates. In [proteinatlas.org](#), all of the validated protein candidates are excluded as prognostic markers due to different criteria, such as low consistency of RNA expression data and antibody staining results or the elevation of these markers in TND conditions. For EMC6 low consistency of RNA expression data and antibody staining results is the criterion to exclude it as prognostic marker. We provide proteomic data, that confirm RNA expression data, because EMC6 levels in TD progression is higher than in TND condition for all grades and the difference becomes significant within advanced grades. The reasons for low antibody staining could lay in the binding efficiency of the currently available antibodies used for EMC6 validation, why these results seem contradictory to RNA expression data and our provided proteomic data. Other candidates arising from differential analysis are excluded due to inconsistency of abundance differences concerning all grades, because the majority of the protein candidates are not significant upregulated at all grades. Literature evidence for biomarker validation via RNA expression data and antibody staining is available for all protein candidates, except for PSME2, for which only RNA expression data are available on [proteinatlas](#). Two protein candidates arising from differential analysis and associated with glycosylation activity are POGLUT1 and ST6GAL1. These protein candidates are only observed in TD progression. POGLUT1 is found in all grades with increasing protein levels within lesion progression but is not found at any TND grade. Nonetheless, antibody staining for POGLUT1 did not give sufficient results, due to the high number of cancer tissue samples where

POGLUT1 antibody staining failed. In the proteomic data we confirm that POGLUT1 is not detectable in all cancer samples and is only found in IPMN samples and not in PanIN precursor lesions, which is the most common precursor lesion for aggressive pancreaticobiliary cancers. Another protein candidate uniquely expressed in TD precursor lesion is ST6GAL1, which is a glycoprotein with sialylation activity. ST6GAL1 appears in precursor grades, like low-grade or high-grade precursor lesions but is not present at normal stage and rarely detected in tumor samples, why it can be suggested that ST6GAL1 is a driving force for tumor progression. Increased O-glycosylation and sialylation can influence tumor microenvironment and the interaction to immune cells.<sup>191</sup> To uncover these processes, the Glyco-DIA approach (Ye, Mao, Clausen and Vakhrushev<sup>10</sup>) is used for further glyco-modification analysis. Provided Glyco-DIA spectral libraries mainly consist peptides carrying the core-1 or the sialylated (capped) core-1 O-glycans (T and Sia-T antigen), or peptides only carrying GalNAc-glycomodification like for the Tn antigen, and the sialylated form Sia-Tn. As mentioned before, a characteristic of many adenocarcinomas is the blockage of the core-1 extension, which results in the accumulation of these core-1 O-glycans, while in healthy cells core extension takes place.<sup>191</sup> Therefore, higher accumulation of core-1 O-glycans is expected for the TD tissue grades and can be supported by the Glyco-DIA data, globally. The analysis of the whole proteome dataset unveiled that the expression of protein O-glycosyltransferases 1 (POGLUT1) is exclusively detected in TD tissue samples across all grades, thus reinforcing the indication of increased O-glycosylation activity. O-Glycosylation is involved in the generation of antigens, such as mucin-bound carbohydrate antigens and can influence immune modulation, which is supported by the pathway analysis of lesion progression, because pathways related to the immune system and immune modulation are upregulated in TD progression. Another important fact is, that O-Glycosylation mainly occurs in the ER with the potential of following protein targeting to Golgi and finally to membrane, as secreted vesicle/exosome or to incorporate into the cell membrane, which increases the potential to find a liquid biomarker. These transport processes and the activity of ER protein targeting can be supported by the comparison of early events in tissue transformation and results in increased Notch signaling. Glyco-DIA modification analysis revealed insights to differential modified glycoproteins, in which 4 modification candidates (ACTA1, YWHAQ, DNAJC3, CKAP4) are observed. For all of the modification candidates, except for 14-3-3 theta (YWHAQ), a combination of modification type and abundance differences are observed, why 14-3-3 theta is analyzed in more detail concerning the modification type. In samples of normal grade, the abundance levels of the identified peptides for 14-3-3 theta exhibit similarity, albeit with differing modifications observed on the AVTEQGAELSNEER peptide. Therefore, higher levels of the peptide carrying the core-1 glycan (Hex(1)HexNAc(1)) are observed for TD progression on 14-3-3 theta. For this modification, literature evidence is not yet given and the results need to get further validated by external methods. Furthermore, structural influence of other sites nearby the glyco-modification site within the 14-3-3 theta protein needs to get further analyzed to unveil the consequences of different modification type on 14-3-3 theta in the context of TD lesion progression.

## 5 References

- (1) *Pancreatic cancer statistics*. World Cancer Research Fund International, 2020. <https://www.wcrf.org/cancer-trends/pancreatic-cancer-statistics/> (accessed 11.01.2024).
- (2) Siegel, R. L.; Miller, K. D.; Wagle, N. S.; Jemal, A. Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians* **2023**, *73* (1), 17-48.
- (3) UK, P. C. Pancreatic Cancer UK, 2022. <https://www.pancreaticcancer.org.uk/information/just-diagnosed-with-pancreatic-cancer/pancreatic-ductal-adenocarcinoma-and-other-exocrine-tumours/> (accessed 12.01.2024).
- (4) Nakanuma, Y.; Sato, Y. Hilar cholangiocarcinoma is pathologically similar to pancreatic duct adenocarcinoma: suggestions of similar background and development. *Journal of Hepato-Biliary-Pancreatic Sciences* **2014**, *21* (7), 441-447.
- (5) Nakanuma, Y.; Kakuda, Y.; Sugino, T.; Sato, Y.; Fukumura, Y. Pathologies of Precursor Lesions of Biliary Tract Carcinoma. *Cancers* **2022**, *14* (21), 5358.
- (6) Nakanuma, Y. A novel approach to biliary tract pathology based on similarities to pancreatic counterparts: Is the biliary tract an incomplete pancreas? *Pathology International* **2010**, *60* (6), 419-429.
- (7) Ikeda, K.; Monden, T.; Kanoh, T.; Tsujie, M.; Izawa, H.; Haba, A.; Ohnishi, T.; Sekimoto, M.; Tomita, N.; Shiozaki, H. Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections. *Journal of Histochemistry & Cytochemistry* **1998**, *46* (3), 397-403.
- (8) Hughes, C. S.; Moggridge, S.; Müller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols* **2019**, *14* (1), 68-85.
- (9) Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods* **2020**, *17* (1), 41-44.
- (10) Kumar, L.; Futschik, M. E. Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* **2007**, *2* (1), 5.
- (11) Ye, Z.; Mao, Y.; Clausen, H.; Vakhrushev, S. Y. Glyco-DIA: a method for quantitative O-glycoproteomics with in silico-boosted glycopeptide libraries. *Nature Methods* **2019**, *16* (9), 902-910.
- (12) Rodrigues, E.; Macauley, M. S. Hypersialylation in Cancer: Modulation of Inflammation and Therapeutic Opportunities. *Cancers* **2018**, *10* (6), 207.
- (13) Springer, G. F.; Desai, P. R.; Ghazizadeh, M.; Tegtmeier, H. T/Tn pancarcinoma autoantigens: fundamental, diagnostic, and prognostic aspects. *Cancer Detect Prev* **1995**, *19* (2), 173-182.
- (14) Springer, G. F. Immunoreactive T and Tn epitopes in cancer diagnosis, prognosis, and immunotherapy. *Journal of Molecular Medicine* **1997**, *75* (8), 594-602.
- (15) Springer, G. F. T and Tn Pancarcinoma Markers: Autoantigenic Adhesion Molecules in Pathogenesis, Prebiopsy Carcinoma-Detection, and Long-Term Breast Carcinoma Immunotherapy. **1995**, *6* (1), 57-85.
- (16) Jensen, P. H.; Kolarich, D.; Packer, N. H. Mucin-type O-glycosylation – putting the pieces together. *The FEBS Journal* **2010**, *277* (1), 81-94.
- (17) Wells, L.; Vosseller, K.; Hart, G. W. Glycosylation of Nucleocytoplasmic Proteins: Signal Transduction and O-GlcNAc. *Science* **2001**, *291* (5512), 2376-2378.
- (18) Chou, T.-Y.; Hart, G. W.; Dang, C. V. c-Myc Is Glycosylated at Threonine 58, a Known Phosphorylation Site and a Mutational Hot Spot in Lymphomas \*. *Journal of Biological Chemistry* **1995**, *270* (32), 18961-18965.
- (19) Kelly, W. G.; Dahmus, M. E.; Hart, G. W. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *Journal of Biological Chemistry* **1993**, *268* (14), 10416-10424.
- (20) Comer, F. I.; Hart, G. W. O-Glycosylation of Nuclear and Cytosolic Proteins: DYNAMIC INTERPLAY BETWEEN O-GlcNAc AND O-PHOSPHATE. *Journal of Biological Chemistry* **2000**, *275* (38), 29179-29182.
- (21) Karsten, U.; Goletz, S. What controls the expression of the core-1 (Thomsen—Friedenreich) glycotope on tumor cells? *Biochemistry (Moscow)* **2015**, *80*, 801-807.



## Author contributions to Manuscript 3

### **Proteome analysis with combined DIA and Glyco-DIA approach of laser microdissected precursor lesions from pancreatobiliary cancer to improve early cancer diagnostics**

#### Stella Pauls

Planned and performed method optimization of HIAR protocol. Designed, performed and analyzed FFPE samples from normal pancreas tissue. Performed data analysis and post processing. Prepared figures and tables and wrote and edited the manuscript.

#### Anja Stefanski

Conceived and supervised the study.

#### Friederike Opitz

Coordinated and supervised the isolation of FFPE tissue samples and delivery.

#### Sandra Biskup

Performed isolation of FFPE tissue samples. Coordinated sample collection and delivery.

#### Gereon Poschmann

Helped performing data analysis and post processing.

#### Thomas Lenz

Offered CaproMag Device and helped performing data analysis and post processing.

#### Marc Daniel Driessen

Gave expertise in Glyco-proteomic data analysis

#### Irene Esposito

Conceived and supervised the study, acquired funding.

#### Kai Stühler

Conceived and supervised the study, acquired funding, reviewed/edited the manuscript.



## 8 Manuscript 4:

Targeting CLDN6 in germ cell tumors by an antibody-drug-conjugate and studying therapy resistance of yolk-sac tumors to identify and screen specific therapeutic options



RESEARCH ARTICLE

Open Access



# Targeting CLDN6 in germ cell tumors by an antibody-drug-conjugate and studying therapy resistance of yolk-sac tumors to identify and screen specific therapeutic options

Margaretha A. Skowron<sup>1†</sup>, Mara Kotthoff<sup>1†</sup>, Felix Bremmer<sup>2†</sup>, Katja Ruhnke<sup>1</sup>, Fatma Parmaksiz<sup>1</sup>, Annika Richter<sup>2</sup>, Stefan Küffer<sup>2</sup>, Kirsten Reuter-Jessen<sup>2</sup>, Stella Pauls<sup>3</sup>, Anja Stefanski<sup>3</sup>, Philipp Ströbel<sup>2</sup>, Kai Stühler<sup>3</sup> and Daniel Nettersheim<sup>1\*</sup>

## Abstract

**Background** Being the standard-of-care for four decades, cisplatin-based chemotherapy is highly efficient in treating germ cell tumors (GCT). However, often refractory patients present with a remaining (resistant) yolk-sac tumor (YST(-R)) component, resulting in poor prognosis due to lack of novel treatment options besides chemotherapy and surgery. The aim of this study was to identify novel targets for the treatment of YST by deciphering the molecular mechanisms of therapy resistance. Additionally, we screened the cytotoxic efficacy of a novel antibody-drug-conjugate targeting CLDN6 (CLDN6-ADC), as well as pharmacological inhibitors to target specifically YST.

**Methods** Protein and mRNA levels of putative targets were measured by flow cytometry, immunohistochemical stainings, mass spectrometry of formalin-fixed paraffin-embedded tissues, phospho-kinase arrays, or qRT-PCR. Cell viability, apoptosis and cell cycle assays of GCT and non-cancerous cells were performed using XTT cell viability assays or Annexin V / propidium iodide flow cytometry, respectively. Druggable genomic alterations of YST(-R) tissues were identified by the TrueSight Oncology 500 assay.

**Results** We demonstrated that treatment with a CLDN6-ADC enhanced apoptosis induction specifically in CLDN6<sup>+</sup> GCT cells in comparison with non-cancerous controls. In a cell line-dependent manner, either an accumulation in the G2 / M cell cycle phase or a mitotic catastrophe was observed. Based on mutational and proteome profiling, this study identified drugs targeting the FGF, VGF, PDGF, mTOR, CHEK1, AURKA, or PARP signaling pathways as promising approaches to target YST. Further, we identified factors relevant for MAPK signaling, translational initiation and RNA binding, extracellular matrix-related processes as well as oxidative stress and immune response to be involved in therapy resistance.

<sup>†</sup>Margaretha A. Skowron, Mara Kotthoff and Felix Bremmer contributed equally to this work

\*Correspondence:  
Daniel Nettersheim  
Daniel.Nettersheim@med.uni-duesseldorf.de  
Full list of author information is available at the end of the article



**Conclusions** In summary, this study offers a novel CLDN6-ADC to target GCT. Additionally, this study presents novel pharmacological inhibitors blocking FGF, VEGF, PDGF, mTOR, CHEK1, AURKA, or PARP signaling for the treatment of (refractory) YST patients. Finally, this study shed light on the mechanisms of therapy resistance in YST.

**Keywords** Germ cell tumors, Yolk-sac tumor, Therapy, Resistance, CLDN6, Antibody-drug-conjugate

## Background

Germ cell tumors (GCT) are the most common solid tumors among young men aged between 17 and 45 years (Cheng et al. 2018; Park et al. 2018). GCT can be stratified into seminomas (SEM) and non-seminomas (NS), both arising from a misguided or defective primordial germ cell (PGC), resulting in a germ cell neoplasia in situ (GCNIS) (Cheng et al. 2018). The non-seminomatous stem cell-like embryonal carcinomas (EC) have the potential to differentiate into all three germ layers (teratoma) or extra-embryonic tissues, i. e. choriocarcinomas (CC) and yolk-sac tumors (YST) (Cheng et al. 2018). For four decades, cisplatin-based chemotherapy has remained the standard treatment approach for metastatic GCT. Overall, most of the GCT cases can successfully be cured by this strategy, but especially patients with YST receive a poor prognosis due to a high risk of developing therapy resistance, suggesting that the formation of YST represents a therapy escape mechanism of GCT cells (Che et al. 2021).

In the last years, mechanisms and targets were identified that could lead to an altered cisplatin response and subsequent resistance due to decreased uptake, increased efflux or detoxification, and / or modified DNA repair or apoptosis induction (Galluzzi et al. 2012; Skowron et al. 2021a). Cisplatin resistant GCT mainly show deficits in DNA damage repair mechanisms (Skowron et al. 2021a), such as microsatellite instability (MSI), the downregulation of *OCT4*, absent expression of the pro-apoptotic factors *PUMA* and *NOXA*, changes in the microRNA (miR) expression profiles (*miR-17 / -106b*, *miR-302*, *miR371-373*), high MDM2 levels, and phosphorylation-dependent translocation of p21 from the nucleus to the cytoplasm (Skowron et al. 2021a; Jacobsen and Honecker 2015; Lobo et al. 2020; Koster et al. 2010; Mayer et al. 2002; Kitayama et al. 2022).

Generally, while pre-clinical trials showed promising results when combining cisplatin-based chemotherapy with drugs specifically targeting tumor cells, there is a need to further investigate on druggable targets linked especially to (resistant) YST to further increase therapy efficacy.

The tetraspanin membrane protein CLDN6 has been identified as a cancer-associated cell surface biomarker, which is rarely expressed in healthy adult tissues (Reinhard et al. 2020; Zhang et al. 2021; Günzel and

Fromm 2012; Micke et al. 2014). Also in GCT, based on immunohistochemical stainings, CLDN6 has been described as a potential novel diagnostic marker for SEM, EC, and YST (Ushiku et al. 2012). Specifically, all of the tested SEM (n = 14), EC (n = 10) and YST tissues (n = 12) demonstrated moderate to strong CLDN6 levels (Ushiku et al. 2012). Recently, Mackensen et al. published their first results from a phase I / IIa study using CLDN6 CAR-T-cells (BNT211, BioNTech SE) with or without a CLDN6-encoding CAR-T cell amplifying RNA vaccine (CARVac) for the treatment of seven pre-treated patients suffering from testicular-, ovarian-, or endometrial cancer, as well as soft-tissue sarcoma (Mackensen et al. 2021). Updated data from this trial (NCT04503278) indicated tumor shrinkage as well as CAR T-cell persistence in five GCT patients after six weeks of treatment (Mackensen et al. 2022).

An approach targeting CLDN6 to treat GCT is using related antibodies coupled with a potent cytotoxin, e. g. monomethyl auristatin E (MMAE) (Johansson et al. 2017). Antibody-drug-conjugates (ADC) are notably specific, effective, and well-tolerated therapeutics, since the protein-of-interest-associated antibody domain allows the direction of the conjugated cytotoxin specifically to the tumor cells (Johansson et al. 2017).

In this study, we decided to design a novel ADC (coupled to MMAE) against CLDN6 to examine the cytotoxicity and specificity to target GCT.

Beyond that, since specifically YST cells represent the remaining non-responsive component in refractory GCT, we further aimed at deciphering the molecular mechanisms resulting in the formation of cisplatin-resistant YST. This approach was chosen for the identification of putative multikinase inhibitor-based therapeutic options, which are already in clinical trials and / or approved for the treatment of other tumor entities.

With regard to molecular profiling of YST, we previously observed high molecular similarities between YST and hepatocytes as well as hepatocellular carcinomas (HCC) regarding expression of endodermal factors like *FOXA2*, *SOX17*, *APOA1 / A2 / B*, *ALB*, *FGA / B / G*, and *GATA3 / 4 / 6* (Ang et al. 2018; D'Amour et al. 2005; Wruck et al. 2021). Additionally, we found similarities in signaling pathway activities when comparing YST and HCC (enhanced WNT and BMP signalling)

(Ang et al. 2018; D'Amour et al. 2005; Wruck et al. 2021). Based on these observations, we hypothesized that drugs used to treat HCC might also be suitable for the treatment of YST and, thus, screened drugs already in use or in clinical trials in the field of HCC therapy. Additionally, this study deciphered mechanisms of cisplatin therapy resistance in primary and resistant (-R) YST on mutational and proteome level, thereby offering novel therapeutic strategies for the treatment of YST(-R).

## Methods

### Cell culture and standard laboratory techniques

The used (tumor) cell lines were cultured as described previously and summarized in Additional file 5: Table S1A (Skowron et al. 2021b; Burmeister et al. 2022). Polarization of THP-1 cells was described earlier and was based on Genin et al. (2015), Skowron et al. (2022). XTT cell viability assays upon treatment with inhibitors (Additional file 5: Table S1B) were performed as described previously (Skowron et al. 2022). The human phospho-kinase arrays (R&D Systems via Bio-Techne, Wiesbaden, Germany) were carried out according to the manufacturer's protocol and evaluated using the 'Protein Array Analyzer'-Plugin for 'Image J' (<https://imagej.nih.gov/ij/>) (Schneider et al. 2012; Carpentier 2022). Further standard laboratory techniques, such as cDNA synthesis, qRT-PCR, flow cytometry-based measurement of apoptosis rates and the cell cycle phase distribution, as well as immunohistochemistry have been described elsewhere (Wruck et al. 2021; Skowron et al. 2021b; Burmeister et al. 2022). See Additional file 5: Table S1B–D for detailed information on the utilized drugs, oligos, and antibodies, respectively.

### Development of antibody-drug-conjugates

Anti-human CLDN6 monoclonal mouse IgG<sub>2B</sub> antibody (Clone #342927, R&D Systems via Bio-Techne) has been conjugated to MMAE via the drug-linker OSu-Glu-vc-PAB (CLDN6-Glu-vc-PAB-MMAE) by Levena Biopharma (San Diego, CA, USA), resulting in an antibody-drug-ratio of at least 1 : 3.

### Nucleic acid extraction and quality assessment

DNA and RNA were extracted from tumor enriched 2 × 5 µm formalin-fixed, paraffin-embedded (FFPE) slices using the 'InnuPREP FFPE DNA Kit' on the 'InnuPure C16 System' (Jena Analytika, Jena, Germany) or the 'Maxwell RNA extraction kit' (Promega, Walldorf, Germany) according to manufacturer's recommendations, respectively.

### Library preparation, sequencing and analysis

DNA libraries were prepared using the hybrid capture-based 'TruSight Oncology 500 Library Preparation Kit' (Illumina, San Diego, CA, USA) following 'Illumina's TruSight Oncology 500 Reference Guide' (document #1000000067621 v00, Illumina Cambridge, UK) and sequenced on an 'Illumina NextSeq 500' instrument. FastQ files were analyzed using 'CLC Genomics Workbench' (Qiagen, Hilden, Germany). The reads were mapped on hg19 followed by an initial variant calling.

### Liquid chromatography coupled to mass spectrometry (LC-MS)

For sample preparation, a modified FFPE tissue lysis protocol of Ikeda et al. was applied (Ikeda et al. 1998). Briefly, after deparaffinization by shaking in Xylene for 5 minutes (min), tissues were resuspended in lysis buffer (300 mM TRIS / HCl, 2 % SDS, pH 8.0), shock-frozen in liquid nitrogen and heated at 99 °C for 25 min. Subsequently, tissues were ultrasonicated twice on ice for 20 min with 30 seconds (s) on / off cycles and then shaken for 2 hours (h) at 80 °C and 500 rounds per minute (rpm). After centrifugation, protein concentration of supernatants was determined by the 'Pierce 660 nm Protein Assay' (Thermo Fisher Scientific, Idstein, Germany). The LC-MS analysis was performed using a modified magnetic bead-based sample preparation protocol as described previously (Hughes et al. 2014). Here, a total of 20 µg protein was reduced using 300 mM DTT and shaking at 56 °C and 1000 rpm for 20 min, followed by alkylation and the addition of 200 µg beads (Sera-Mag SpeedBeads Sigma Aldrich, Taufkirchen, Germany) per sample. Subsequently, 80 % ethanol was added for protein aggregation capture, followed by thrice rinsing steps using 80 % ethanol and once using 100 % ACN. Beads were resuspended in 50 mM TEAB buffer and trypsinized at 37 °C at 1000 rpm. For the LC-MS on the 'Orbitrap Fusion Lumos Tribrid Mass Spectrometer' equipped with an 'Acclaim PepMap 100 C18' column (75 µm inner diameter, 25 cm length, 2 mm particle size) as a separation column and an 'Acclaim PepMap 100 C18' column (75 µm inner diameter, 2 cm length, 2 mm particle size) as a trap column (all equipment from Thermo Fisher Scientific), 500 ng of each sample were used. Data analysis was performed using the 'Proteome Discoverer' (version 2.4.1.15, Thermo Fisher Scientific), while the RAW files were matched against the human 'Swissprot' database (Download: 23.01.2020) and the 'Maxquant Contaminant' database (Download: 20.02.2021), using 'SequestHT' integrated in the 'LFQ Tribrid' processing workflow (Thermo Fisher Scientific).

### Online analysis tools

The TCGA ('The Cancer Genome Atlas') GCT cohort was analyzed using cBioportal (<https://www.cbioportal.org/>) (Gao et al. 2013; Cerami et al. 2012). LC-MS data were analyzed by 'PCAGO' (<https://pcago.bioinf.uni-jena.de/>) for principal component analyses (PCA) (Gerst and Hölzer 2019), while the 'pandas', 'seaborn', and 'matplotlib' libraries were used in 'Python' for Pearson's correlation analyses and visualization via volcano plots (Hunter 2007; Waskom 2012; McKinney 2010; Reback, et al. 2021; Flyamer 2017). The 'DAVID Functional Annotation Tool' using 'GOTERM\_BP\_DIRECT' and 'GOTERM\_MF\_DIRECT' (<https://david.ncifcrf.gov>) (Dennis et al. 2003) and 'STRING' analyses (<https://string-db.org/>) (Szklarczyk et al. 2019) predicted the molecular functions and protein interactions of deregulated proteins, respectively. The 'SIGNAL' web-based analysis platform was used for the identification of signaling cascades (<https://signal.niaid.nih.gov/>) (Katz et al. 2021). The online platform 'ImageGP' (<https://www.bic.ac.cn/ImageGP>) was used to generate dot plots (Chen et al. 2022). The 'ADMETlab 2.0' web platform (<https://admet.scbdd.com>) has been used to screen for the absorption, distribution, metabolism, excretion and toxicity features of tested drugs (Dong et al. 2018; Xiong et al. 2021).

## Results

### CLDN6 as a therapeutic option to target YST

In this study, CLDN6 was chosen to be evaluated as therapeutic targets by using ADC to treat GCT cells.

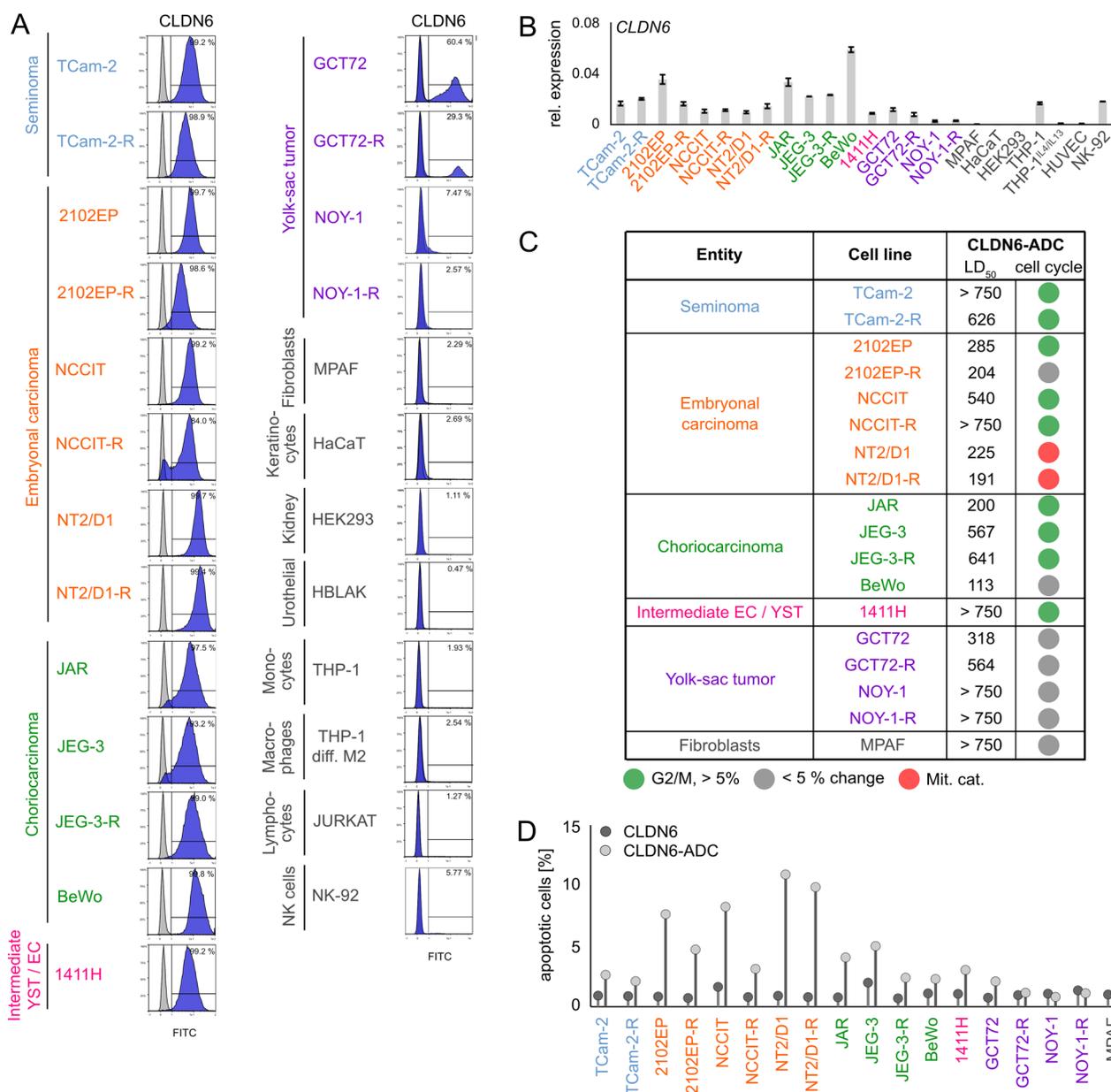
While genomic alterations in *CLDN6* were merely observed in the TCGA GCT cohort, we could further show *CLDN6* / CLDN6 being detectable on mRNA and protein level in GCT cell lines including cisplatin-resistant subclones (-R) derived from SEM (TCam-2), EC (2102EP, NCCIT, NT2/D1), CC (JAR, JEG-3, BeWo), and an EC-YST-intermediate (1411H) (Additional file 1: Fig. S1A; Fig. 1A, B). In male YST cells (GCT72(-R)), a predominant CLDN6<sup>-</sup> and a small CLDN6<sup>+</sup> population was found, while only low levels of *CLDN6* / CLDN6 were observed in the female YST cell line NOY-1(-R), which were comparable to those of non-cancerous control cells (i. e. fibroblasts, immune cells, keratinocytes) (Fig. 1A, B). Next, the effects on cell viability, apoptosis rates, and the cell cycle distribution of the novel CLDN6-ADC were evaluated by XTT assays and flow cytometry, respectively (Fig. 1C, D). Treatment with the CLDN6-ADC reduced cell viability (LD<sub>50</sub> 72 h 191 - 641 ng / ml) and induced apoptosis in most CLDN6<sup>+</sup> GCT(-R) cells (i. e. SEM, EC, CC, and YST cell lines) in comparison to the monoclonal antibody alone (Fig. 1D; Additional file 1: Fig. S1B). After 48 h, EC(-R) cell lines showed the strongest increase in apoptosis rates, while male YST cell lines

(GCT(-R)) showed only a mild increase in apoptosis and no alterations in the cell cycle phase distribution (Fig. 1C, D). Female CLDN6<sup>-</sup> NOY-1 as well as non-cancerous control cells did not respond to CLDN6-ADC treatment (Fig. 1D), while MMAE alone expectedly reduced cell viability in GCT cells at low concentrations (LD<sub>50</sub> 72 h 0.19 - 10.7 nM) (Additional file 1: Fig. S1B). Moreover, the CLDN6-ADC caused mainly accumulation in the G2 / M cell cycle phase in CLDN6<sup>+</sup> cells (TCam-2(-R), 2102EP, NCCIT(-R), JAR, JEG-3(-R), 1411H), but not in CLDN6<sup>-</sup> cells (Fig. 1C, Additional file 1: Fig. S1C). Similar to the treatment with MMAE alone, also mitotic catastrophes were observed upon treatment with the CLDN6-ADC (NT2/D1(-R)) (Fig. 1C, Additional file 1: Fig. S1C). Therefore, the CLDN6-ADC is suitable for the treatment of the GCT subtypes SEM, EC, and CC. In male YST cells, the CLDN6-ADC is less efficient compared to the other GCT entities, while the ADC is not suitable to target female YST cells. In fact, this observation was further validated via immunohistochemical stainings of CLDN6 in YST-R tissues (n = 10), where only 40 % of the investigated cases presented as CLDN6<sup>+</sup> (Additional file 1: Fig. S1D).

### Identification of novel therapeutic options for YST

Since CLDN6 levels were rather low in YST cells, representing the most aggressive and persistent GCT subtype, eventually, the CLDN6-ADC showed only a moderate efficiency in YST cells. Hence, we characterized therapy-resistant YST to identify putative therapeutic targets, which can be attacked by multikinase inhibitors.

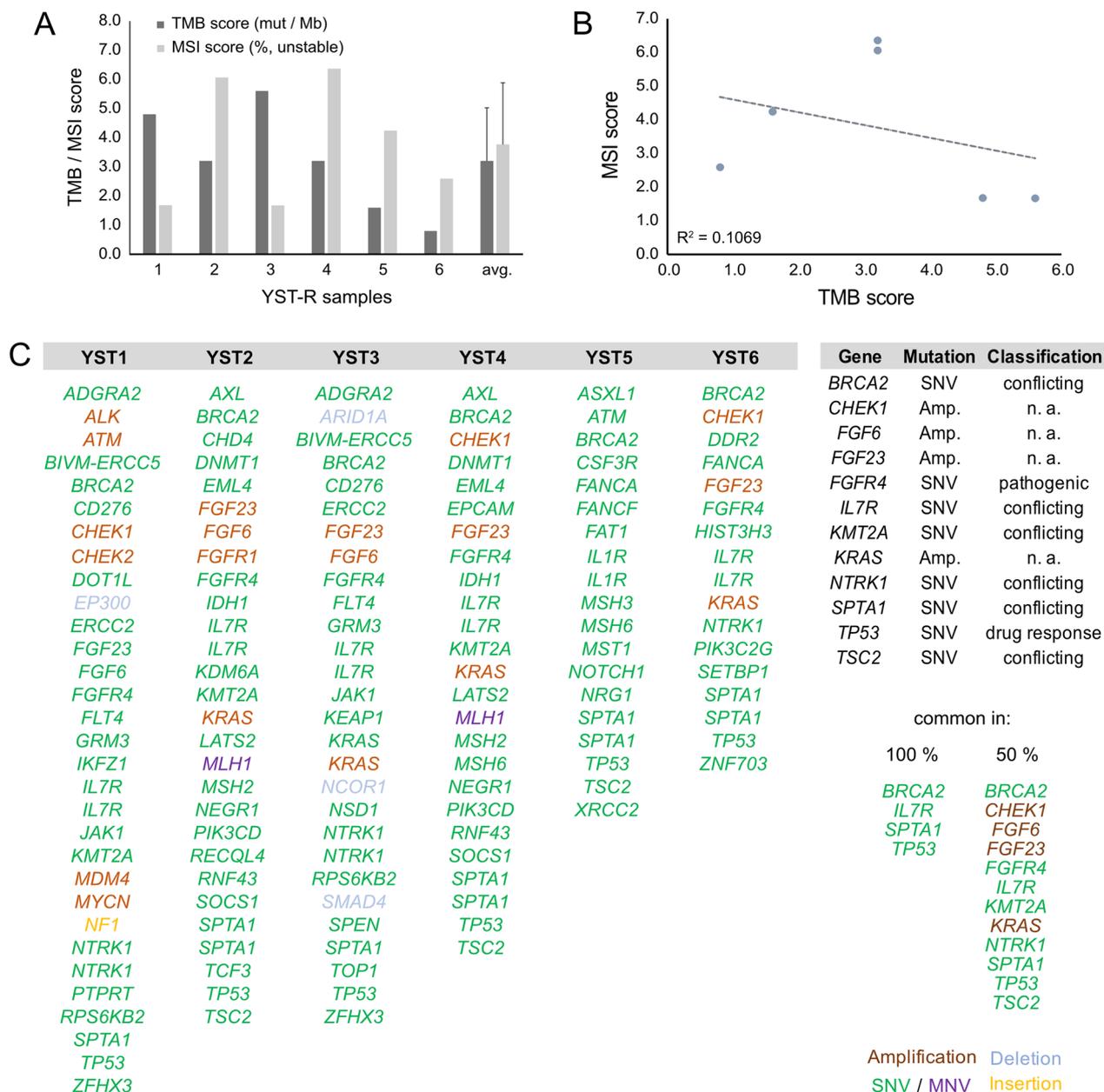
TSO analyses of refractory YST (YST-R) tissues (n = 6) were performed to identify druggable genomic alterations. We detected a mean of 3.2 mut / Mb (0.8 - 5.6 mut / Mb) in the YST-R samples, though, the tumor mutational burden (TMB) did not correlate to the microsatellite instability score (MSI; 3.7 % (1.67 - 6.36 %)) (Fig. 2A, B). Single nucleotide variants (SNV) in *TP53* (c.215C>G), *BRCA2* (c.7397T>C), *IL7R* (c.197T>C, c.412G>A), and *SPTA1* (c.5077A>C) were observed in all YST-R samples. Furthermore, *CHEK1*, *FGF6*, *FGF23*, and *KRAS* were amplified, but with a low fold change (max 2.2), and SNVs were detected in *FGFR4* (c.1162G>A), *KMT2A* (c.10841T>C), *NTRK1* (c.53G>A, c.1810C>T, c.1838G>T), and *TSC2* (1747G>A, c.4285G>T) in at least 50 % of the evaluated samples (Fig. 2C; Additional file 6: Data S1A). Additionally, besides further SNV, amplifications of *ALK*, *ATM*, *CDK4*, *CHEK2*, *FGFR1*, *MDM4*, and *MYCN* were observed in individual samples (Fig. 2C; Additional file 6: Data S1B). Hence, tumor suppressors and DNA repair key players, as well as factors related to the cell cycle, actin skeleton, or the MAPK and FGF signaling pathways were frequently altered in YST-R. It has to be noted that most found mutations were



**Fig. 1** CLDN6-ADC as a novel therapeutic option to target GCTs. **A** Raw flow cytometry data of CLDN6-FITC stained (blue) GCT cell lines, including their cisplatin-resistant sublines, and non-cancerous control cells compared with unstained controls (grey). **B** Relative *CLDN6* expression in GCT cell lines and non-cancerous control cells. *ACTB* and *GAPDH* were used as housekeeping genes. **C** LD<sub>50</sub> values (ng / ml) acquired by XTT cell viability assays 72 h after treatment with CLDN6-ADC and color-coded changes in cell cycle distribution (G2 / M = green, mitotic catastrophe = red, changes < 5 % = grey) upon treatment with CLDN6-ADC as compared to treatment with the CLDN6 antibody alone in GCT cell lines, including their cisplatin-resistant sublines, as well as fibroblast control cells (MPAF). **D** Lollipop graph summarizing relative number of apoptotic cells in GCT cell lines and fibroblast control cells after treatment with either CLDN6-ADC or CLDN6 antibody alone

SNV classified as ‘conflicting interpretations of pathogenicity’, suggesting that further work is necessary to narrow down the consequences of these mutations, with the exception of *TP53* and *FGFR4*, whose SNV were classified as affecting ‘drug response’ and ‘pathogenic’, respectively (Additional file 6: Data S1A).

A phospho-kinase array of YST-like cells (GCT72, NOY-1(-R), 1411H) has been performed to identify signaling molecules and putative YST-specific targets, which were not present in non-cancerous control cells (MPAF) (Additional file 3: Fig. S3A). Compared to fibroblasts, in YST cell lines high levels of AKT1 - 3 (T308,



**Fig. 2** Mutational profiling of YST-R. **A** Tumor mutational burden score (TMB) and microsatellite instability score (MSI) found in six cisplatin-resistant YST samples. **B** Pearson's correlation plot of TMB and MSI. **C** List of identified individual and common genomic alterations found in six cisplatin-resistant YST tissues

S473), ERK1 / 2 (T202 / Y204, T185 / Y187), GSK3 $\alpha$  /  $\beta$  (S21 / S9) and p53 (S15, S46, S392) phosphorylation were detected (Fig. 3D, Additional file 2: Fig. S2 A).

Thus, based on the TSO and phospho-kinase array, we included AZD4547 (FGFR1-4), Nintedanib (FGFR1 - 3), AZD7762, MK-8776 (both CHEK1), and Rapamycin (mTOR) as potential inhibitory drugs to target YST(-R) (Fig. 3B). Additionally, based on the previously described resemblances between YST and HCC,

we included drugs to treat HCC, i. e. Sorafenib, Lenvatinib, Regorafenib, and Cabozantinib (Fig. 3B) (Fonseca et al. 2020). The mRNA expression levels of the putative targets of these (multikinase) inhibitors were evaluated in GCT72, 1411H, NOY-1, and MPAF cells. Here, *AURKA* / *B*, *CSK*, *FGFR1* / 2, *KIT*, *PARP1* / 2, *PDGFRA*, *RAF1*, and *YES1* were specifically expressed in the YST cells in comparison to fibroblasts (Additional file 2: Fig. S2B). The mutational status as well

as mRNA expression level of respective targets of the putative (multikinase) inhibitors were also evaluated in the TCGA GCT cohort (Additional file 3: Fig. S3). No aberrations were observed in *ABL2*, *AURKA*, *AURKB*, *AXL*, *BRAF*, *BTK*, *FGFR2 / 3*, *MET*, *PDGFRB*, *REG1A*, and *SRC*, and only few deep deletions (0.7 %) were noted in *ABL1*, *FGFR1 / 4*, *FLT4*, *MTOR*, *PARP*, *PDK1*, *RET*, while *CSK*, *EGFR*, *ERBB2*, and *RAF1* harbored missense mutations (Additional file 3: Fig. S3 A). *PDGFRA* amplifications and *CHEK1* deletions were noted in 2.1 % and 8.0 % of GCT cases, respectively. Missense mutations as well as amplifications in *KIT* were observed in 15 % of the GCT patients, though mostly in SEM (Additional file 3: Fig. S3A). Regarding the mRNA levels of these putative targets, specific expression profiles / clusters were noted. As such, most SEM tissues showed specifically high levels of *BTK*, *CSK*, *FGFR3*, *KIT*, *ABL2*, *RET*, *PARP2*, *RAF1*, and *PDK1*, while non-seminomatous GCT were positive for *AURKA / B*, *FGFR1 / 4*, *YES1*, *MET*, *ERBB2*, *EGFR*, *FLT4*, *PDGFRA / B*, *AXL*, and *SRC* (Additional file 3: Fig. S3B). High expression levels of *ABL1*, *FGFR2*, *MTOR*, *BRAF*, and *PARP1* were seen in both tumor subtypes (Additional file 3: Fig. S3B).

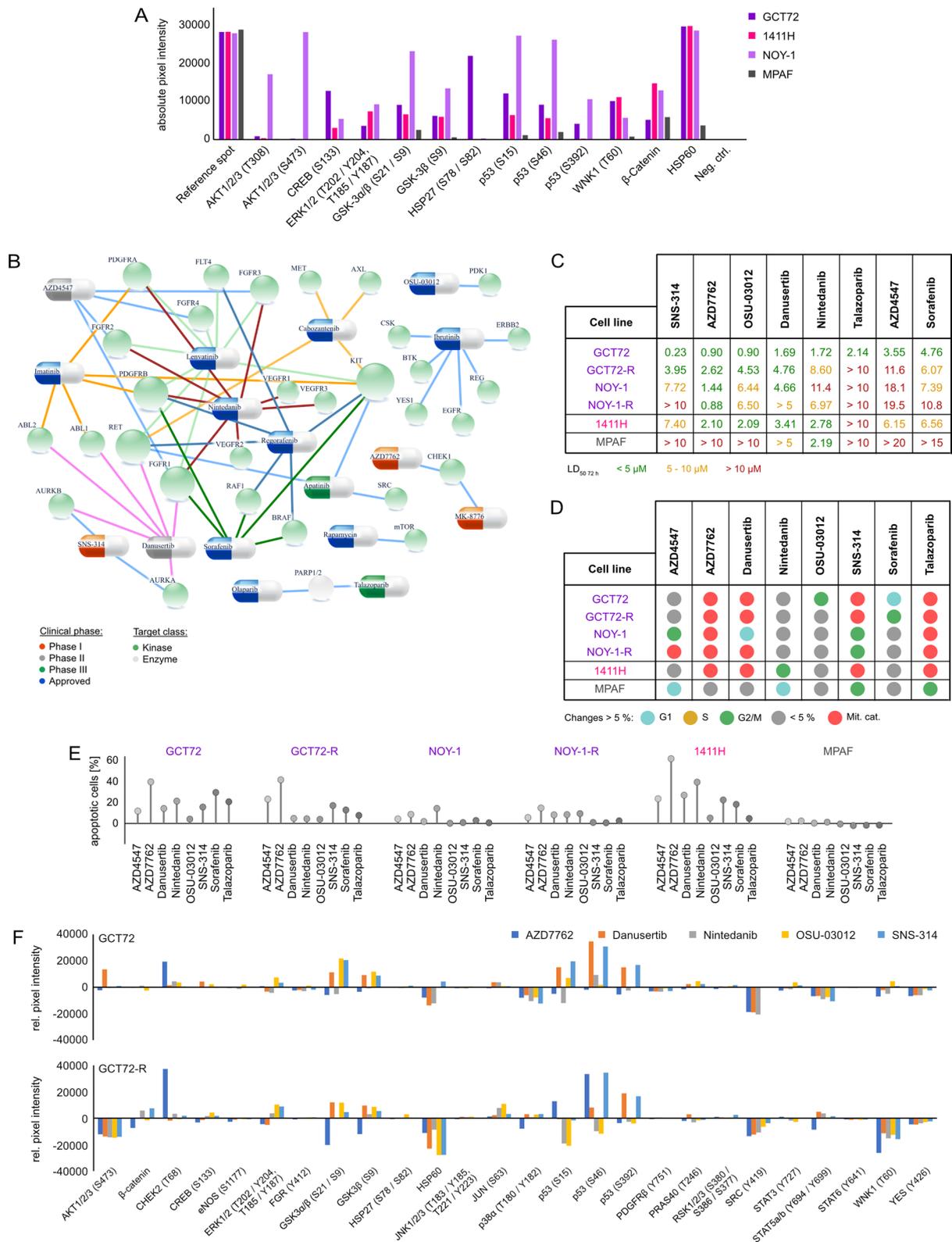
Out of the 17 tested multikinase inhibitors, treatment of GCT72 cells with Danusertib, SNS-314 (both AURKA - C), Nintedanib (VEGFR1 - 3, FGFR1 - 3, PDGFRA / B), Sorafenib (RAF1, BRAF, VEGFR2 / 3, PDGFRB, FLT3, KIT), Talazoparib (PARP1 / 2), OSU-03012 (PDK1), AZD4547 (FGFR1 - 3), AZD7762 (CHEK1), or Rapamycin (mTOR) resulted in LD<sub>50</sub> values of below 5 μM (Fig. 3C, Additional file 4: Fig. S4A, B). All other drugs showing higher LD<sub>50</sub> values were excluded from further analyses (Fig. 3C, Additional file 4: Fig. S4 A, B). Effects on cell viability upon treatment with the most potent inhibitors were further evaluated in NOY-1(-R) (YST), 1411H (EC-YST-intermediate), and fibroblasts (MPAF). With the exception of Nintedanib, LD<sub>50</sub> values in fibroblasts upon treatment with these drugs were above 5 μM, thereby offering a therapeutic window (Fig. 3C, Additional file 4: Fig. S4B). Next, the cell

cycle distribution as well as apoptosis induction upon treatment with AZD4547, AZD7762, Danusertib, Nintedanib, OSU-03012, Rapamycin, SNS-314, Sorafenib, and Talazoparib were evaluated in the four YST-like cell lines (GCT72, NOY-1(-R), 1411H) and fibroblast controls (MPAF) (Fig. 3D, Additional file 4: Fig. S4C). In comparison to the solvent control (DMSO), treatment with AZD7762, Danusertib, OSU-03012, SNS-314, Sorafenib, and Talazoparib affected the cell cycle in most GCT cells in a cell line-dependent manner. Prominently, treatment with Danusertib, SNS-314, or Talazoparib resulted in a mitotic catastrophe in YST-like cells after 24 h, while fibroblasts were only affected slightly, showing a small accumulation in the G<sub>0</sub> / G<sub>1</sub> or G<sub>2</sub> / M phase upon treatment with AZD4547 or SNS-314, respectively (Fig. 3D, Additional file 4: Fig. S4C).

Of the four tested YST-like cell lines, the GCT72 and 1411H showed the highest apoptosis induction under most conditions, while the female NOY-1(-R) were the least sensitive YST-like cells (Fig. 3E). Induction of apoptosis remained rather low (<5 %) in fibroblasts (Fig. 3E). Taking together, treatment with AZD4547 and Nintedanib resulted in apoptosis induction without altering the cell cycle distribution, while treatment with AZD7762, Danusertib, SNS-314, Sorafenib, and Talazoparib not only disrupted the cell cycle, but also induced apoptosis specifically in GCT72 YST cells (Fig. 3D, E; Additional file 4: Fig. S4C). Subsequently, the molecular effects upon treatment with the most sensitive multikinase inhibitors AZD7762, Danusertib, Nintedanib, OSU-03012, and SNS-314 have been evaluated in GCT72(-R) cells (Fig. 3F, Additional file 2: Fig. S2C, D). As such, treatment with the CHEK1 inhibitor AZD7762 enhanced phosphorylation of CHEK2 (T68), while it decreased activity of GSK3α / β (S21 / S9), SRC (Y419), STAT5a / b (Y694 / Y699), and WNK1 (T60) in GCT72(-R) cells. Danusertib treated cells presented elevated phosphorylation of GSK3α / β (S21 / S9) and p53 (S46, S392), while HSP60 and phosphorylation of ERK1 / 2 (T202 / Y204, T185 / Y187), SRC (Y419), and WNK1 (T60) were diminished in both cell lines. Nintedanib treatment resulted

(See figure on next page.)

**Fig. 3** Identification of novel targets for the treatment of YST. **A** Densitometric evaluation of absolute pixel intensities of the 13 most prominent phosphorylation sites in cell lysates from GCT72, 1411H, NOY-1, and MPAF, as measured by the human phospho-kinase array. **B** Graphical illustration of potential pharmacological inhibitors based on genomic alterations found in at least 50 % of YST-R samples and changes on protein level. **C** LD<sub>50</sub> values (72 h) of GCT72(-R), 1411H, NOY-1(-R), and MPAF upon treatment with the inhibitors selected in (B). Inhibitors showing LD<sub>50</sub> values below 5 μM (green) in GCT72 and higher LD<sub>50</sub> values in MPAF (5 - 10 μM = yellow, > 10 μM = red) were further evaluated. **D** Color-coded changes in cell cycle distribution (G<sub>1</sub> = light blue, S = yellow, G<sub>2</sub> / M = green, mitotic catastrophe = red, changes < 5 % = grey) upon treatment with LD<sub>50</sub> (72 h) concentrations for 24 h with indicated drugs, as compared to the solvent control (DMSO) in GCT72(-R), 1411H, NOY-1(-R), and MPAF. **E** Lollipop graph summarizing relative number of apoptotic cells in GCT cell lines and fibroblast control cells after treatment with LD<sub>50</sub> (72 h) concentrations for 48 h with the indicated drugs in comparison to the solvent control. Of note, due to high autofluorescence of Nintedanib, all cell types were treated with LD<sub>50</sub> values (72 h) of GCT72. **F** Densitometric evaluation of relative pixel intensities of the most prominent phosphorylation sites in cell lysates from GCT72 and GCT72-R treated with AZD7762, Danusertib, Nintedanib, OSU-03012, or SNS-314 (24 h, LD<sub>50</sub> 72 h) in comparison to the solvent control (DMSO), as evaluated by the human phospho-kinase array



**Fig. 3** (See legend on previous page.)

commonly in both cell lines in decreased levels of HSP60 and phosphorylation of p53 (S15, S392), SRC (Y419), and WNK1 (T60). Treatment with the PDK1 inhibitor OSU-03012 resulted in several shared abundances on phospho-proteome level in both cell lines (increase in CREB (S133), ERK1 / 2 (T202 / Y204, T185 / Y187), GSK3 $\alpha$  /  $\beta$  (S21 / S9) activity), however, phosphorylation of p38 $\alpha$  (T180 / Y182), p53 (S15, S46), PRAS40 (T246), STAT3 (Y727), and WNK1 (T60) were oppositional in GCT72(-R) cells. As an AURKA / B inhibitor, treatment with SNS-314 led to increased activity of GSK3 $\alpha$  /  $\beta$  (S21 / S9) and p53 (S46, S392) in both cell lines. Additionally, decrease in YES (Y426) phosphorylation was found in both cell lines upon treatment with all here tested inhibitors. Remarkably, all five inhibitors resulted in diminished phosphorylation of AKT1 - 3 (S473) and WNK1 (T60) specifically in the resistant cell line, thereby indicating that the PI3K / PDK1 signaling cascade might be putatively targetable in YST-R (Fig. 3F, Additional file 2: Fig. S2C, D).

Consequently, based on a molecular-guided approach, the authors could identify suitable YST-specific candidates that could be targeted using already tested or even approved multikinase inhibitors. Nevertheless, one of the major obstacles during cisplatin-based chemotherapy of YST is the development of resistance mechanisms. At present, this fundamental process is poorly understood. Thus, we subsequently aimed at the molecular characterization of potential mechanisms driving towards a resistant phenotype in YST. Analyzing the differences between YST cells and their resistant sublines, elevated phosphorylation of AKT1 - 3 (S473) and p53 (S15, S46) were seen in GCT72-R cells, while activity of ERK1 / 2 (T202 / Y204, T185 / Y187), FGR (Y412), GSK3 $\alpha$  /  $\beta$  (S21 / S9), p38 $\alpha$  (T180 / Y182), p53 (S392), PDGFR $\beta$  (Y751), SRC (Y419), STAT5a / b (Y694 / Y699), WNK1 (T60), and YES (Y426) was reduced in the resistant subline (Fig. 4A, Additional file 2: Fig. S2C). To further decipher the underlying molecular mechanisms driving YST to a resistant phenotype and identify further targets for the treatment of YST, mass spectrometry-based proteome analyses were performed. A PCA revealed that YST-R samples (n=5) clustered distinguishably apart from primary YST tissues (n=9) (Fig. 4B). Even though both tissue types had a high correlation ( $r^2=0.95$ ) (Fig. 4C), 84 proteins were highly enriched and 67 were significantly depleted (abundance ratio <0.5 or >2, p-value <0.05) in YST-R compared to therapy-naïve YST (Fig. 4D, Additional file 6: Data S1C). A DAVID-based gene ontology and STRING interaction analysis of the proteins enriched in YST-R (abundance ratio >2) revealed that these factors are involved in translational initiation and RNA binding (e. g. RPL34, RPS8, EIF3G), extracellular matrix (ECM)-related processes (e.

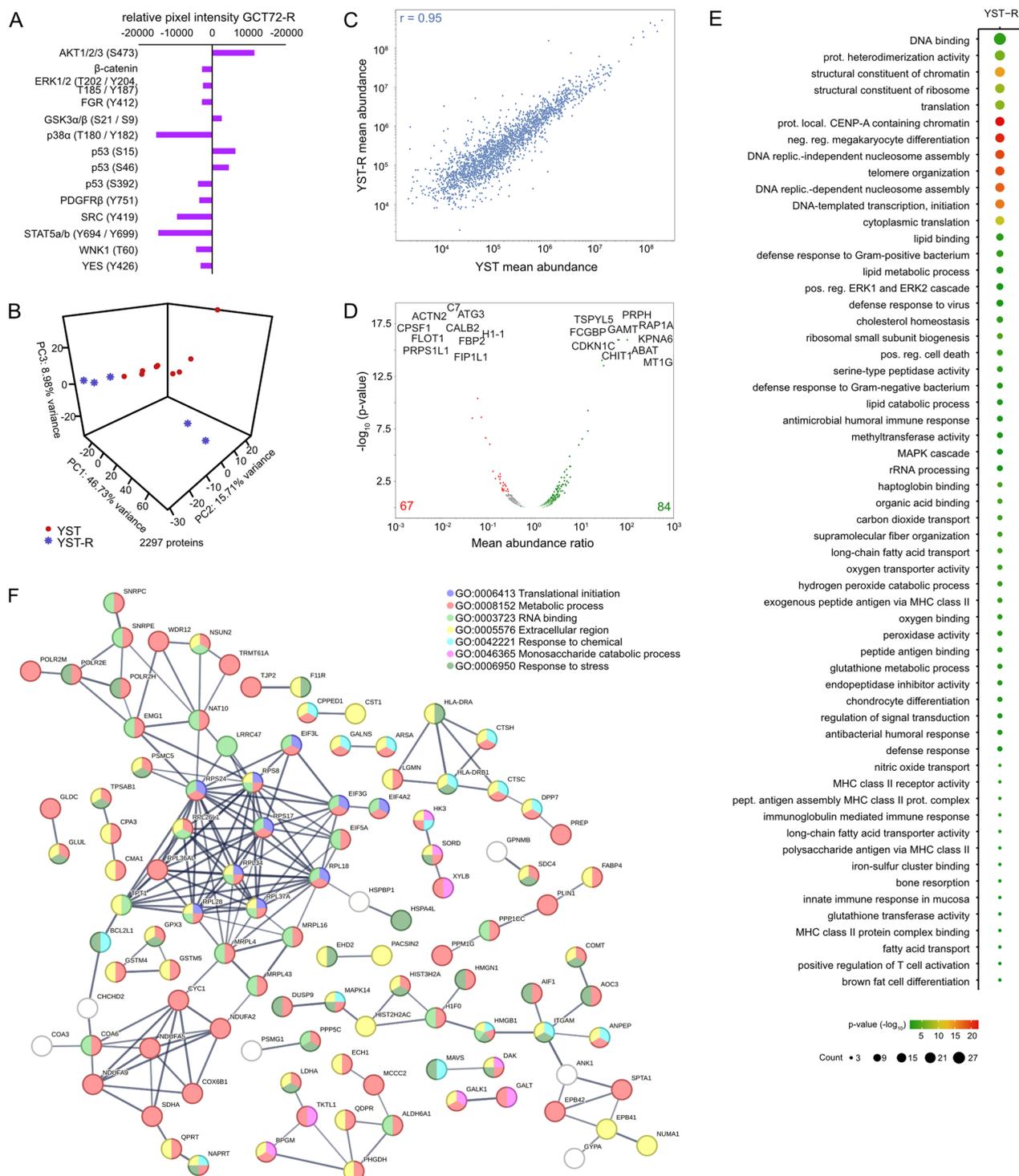
g. COL3A1, COL2A1, ITGAX, MFAP5), as well as innate / humoral (pathogen-dependent) immune response (e. g. CD36, HLA-DPB1, HLA-DRB1, LILRB5). Additionally, factors relevant during oxidative stress response and MAPK / Ras / Rap1 signaling (RALB, RAP1A, GNG12, DUSP9, PPP5C) were identified as putative supporting processes in the acquisition of resistance in YST (Fig. 4E, F).

## Discussion

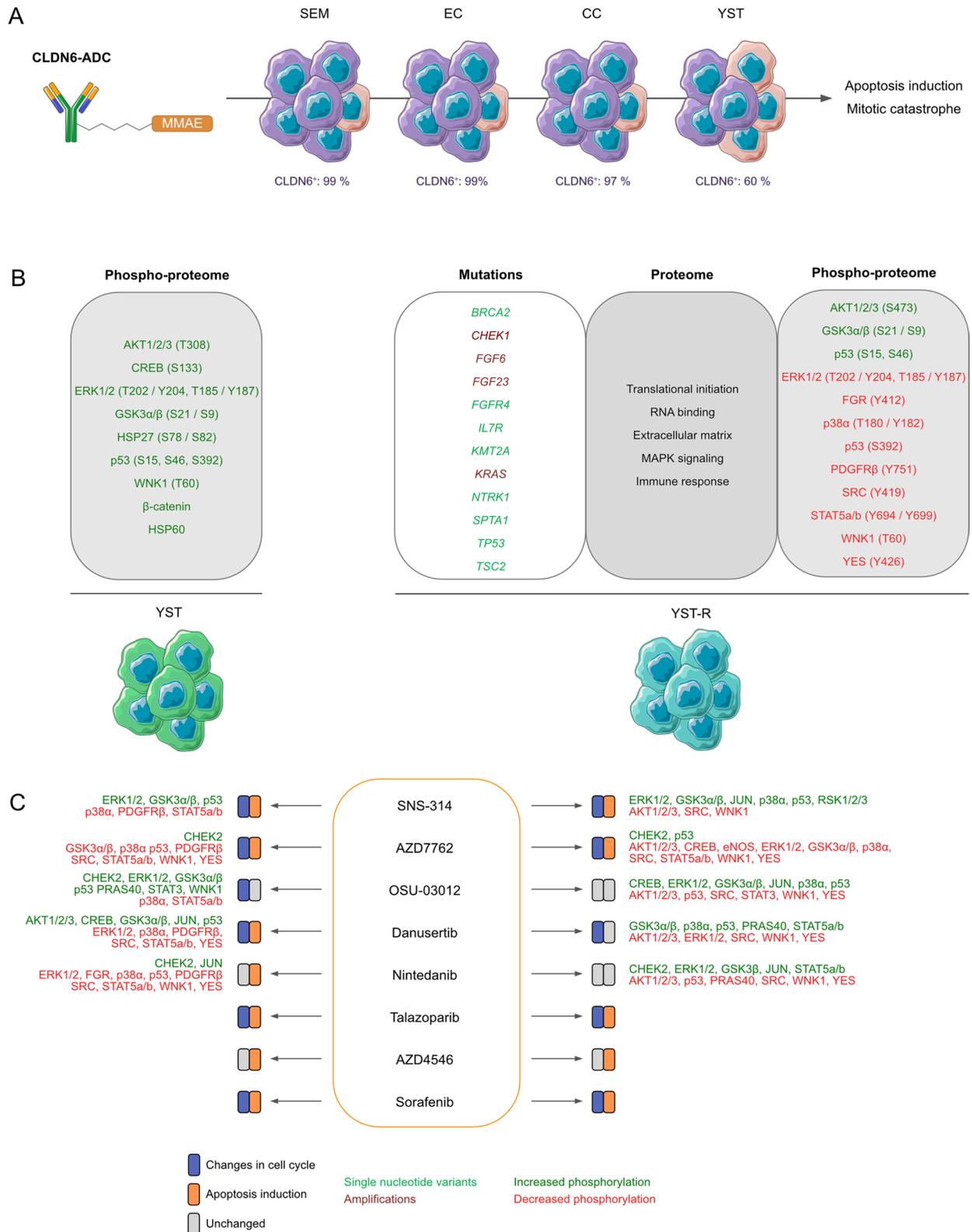
Cisplatin-based chemotherapy is the standard treatment for metastatic GCT. Especially the YST component remains in refractory cases, thereby resulting in a resistant phenotype with limited treatment options. Even though immune checkpoint inhibition has been a promising novel therapeutic approach in various solid tumor entities (Jacob et al. 2021; Dall'Olio et al. 2022) and albeit high levels of PD-L1 and CTLA-4 have been noted in GCT tissues (Lobo et al. 2019; Fankhauser et al. 2015; Cierna et al. 2016), up to now, treatment with immune checkpoint inhibitors, such as pembrolizumab (NCT02499952), durvalumab and tremelimumab (NCT03081923), avelumab (NCT03403777), nivolumab and / or ipilimumab (NCT03333616) showed rather limited efficacy in related clinical trials (Mego et al. 2019; Kawahara et al. 2022; Zschäbitz et al. 2016; Zschäbitz et al. 2017; Tsimberidou et al. 2021; McGregor et al. 2021; Necchi et al. 2019a; Adra et al. 2018). Hence, there is a high necessity of finding novel therapeutic approaches for the treatment of (refractory) GCT.

This study validated the cytotoxic efficacy of a novel ADC targeting CLDN6 as a therapeutic option for GCT. As most ADC, the here presented payload is based on the microtubule inhibitor MMAE (Fu et al. 2022). Beside targeting the antigen-positive cell, also off-target effects for MMAE-based ADC have been noted e. g. for the CD30-ADC brentuximab vedotin (Romano et al. 2019). As such, treatment with brentuximab vedotin not only resulted in apoptosis induction resembling immunogenic cell death, but also led to a pro-inflammatory immune response against lymphoma cells, thereby offering the putative combination with an anti-PD-1 therapy (Cao et al. 2017).

Treatment with CLDN6-ADC resulted in mitotic catastrophes and induction of apoptosis in CLDN6<sup>+</sup> GCT(-R) cells (Fig. 5A). Compared to SEM, EC and CC cells, the CLDN6-ADC was less efficient in GCT72 YST cells (Fig. 5A). By flow cytometry, we measured a CLDN6<sup>+</sup> and a predominant CLDN6<sup>-</sup> population in GCT72 cells. Thus, we conclude that the CLDN6-ADC targets the smaller CLDN6<sup>+</sup> population only, explaining the weaker responses of GCT72 cells in performed analyses on cell cycle phase distribution and apoptosis rates. Due to the low levels of CLDN6 on protein level, female



**Fig. 4** The molecular resistance mechanisms in YST. **A** Densitometric evaluation of relative pixel intensities of the most prominent phosphorylation sites in cell lysates from GCT72-R cells in comparison to the parental cell line, as evaluated by the human phospho-kinase array. **B** PCA plot, **C** Pearson's correlation plot and **D** Volcano plot of mass-spectrometry data of FFPE-embedded YST(-R) tissues. **E** Enrichment plots showing gene ontology terms found exclusively in YST-R tissues. **F** STRING interaction analysis of YST-R-specific proteins



**Fig. 5** Graphical summary of the observed findings of this study. **A** Targetability of GCT cells by the CLDN6-ADC and provoked effects. **B** Molecular features of YST(-R) cells on DNA and protein level. **C** Molecular effects of the drugs screened as therapeutic option for YST(-R). Created using bioicons (<https://bioicons.com/>)

YST cells are not suitable for an attack by the CLDN6-ADC. Beyond GCT, CLDN6 would be also a suitable target for other CLDN6<sup>+</sup> tumor entities, such as myeloid leukemia, ovarian, endometrial, or urothelial carcinoma (Zhang et al. 2021). Several clinical trials are evaluating the targetability of CLDN6 using immune therapeutic approaches. Currently, a phase 1 / 2 trial using CLDN6-targeting CAR-NK cells is recruiting patients suffering from advanced ovarian, endometrial cancer, or GCT (NCT05410717), while other phase 1 / 2 trials using CLDN6 CAR-T cells (BNT211; NCT04503278) or mRNA encoded bispecific T cell engaging antibody targeting CD3 and CLDN6 (BNT142; NCT05262530) are also currently recruiting for the treatment of solid tumors. However, a phase 2 trial using the monoclonal antibody targeting CLDN6 (ASP1650) for the treatment of refractory GCT patients had to be terminated due to the lack of efficacy (NCT03760081) (Adra et al. 2022). Though, a current study describing the generation and preclinical characterization of a CLDN6-ADC for the treatment of ovarian and endometrial cancer using a patient-derived xenograft model showed reduced tumor volume specifically in CLDN6<sup>+</sup> tumors (McDermott et al. 2022). Here, a phase 1 trial is currently recruiting to test this CLDN6-ADC (DS-9606a) for the treatment of advanced ovarian cancer and GCT (NCT05394675). As such, one major advantage of using CLDN6-ADC instead of CAR-T-cell-based therapy is its rapid therapeutic accessibility, which, in case of the manufacturing process of CLDN6-CAR-T-cells, might otherwise require a longer time (Reinhard et al. 2020; Rasche et al. 2021). Additionally, in case of autologous CAR-T-cells, harvesting T-cells from heavily pretreated patients might be challenging due to the quality and quantity of T-cells (Rasche et al. 2021). Alternatively, allogeneic T-cells from healthy donors might circumvent these manufacturing issues, however, putative graft versus host disease might be a risk factor (Rasche et al. 2021; Rafiq et al. 2020).

To identify factors that allow for attacking YST components specifically, we further aimed at the identification of novel therapeutic targets based on genomic alterations and changes on the proteome level. Similar to previous observations in primary and resistant GCT tissues (Necchi et al. 2020; Necchi et al. 2019b; Cheng et al. 2020; González-Barrios et al. 2022), we observed an overall low TMB of averaged 3.2 mut / Mb in the YST-R samples. Nevertheless, factors, such as *CHEK1*, *FGF6*, *FGF23*, and *TP53* were commonly amplified in at least 50 % of the evaluated samples (Fig. 5 B). Additionally, we observed AKT1 - 3 (T308 and S473), ERK1 / 2 (T202 / Y204, T185 / Y187), GSK3 $\alpha$  /  $\beta$  (S21 / S9) and p53 (S15, S46, S392) phosphorylation in YST samples, while resistant YST cells further showed elevated

activity of AKT1 - 3 (S473), GSK3 $\alpha$  /  $\beta$  (S21 / S9), and p53 (S15, S46) (Fig. 5B). Together with our previous observations of certain molecular analogies between YST and HCC, we identified novel therapeutic targets whose inhibition resulted in apoptosis induction and / or cell cycle arrest in a drug-dependent manner in YST cells, while fibroblasts remained mostly unaffected, thereby opening a therapeutic window for the treatment of (cisplatin-resistant) YST patients. Overall, several phase 2 studies showed only limited clinical benefit from treatment with either the PARP1 / 2 inhibitor Olaparib (Giorgi et al. 2020) and Veliparib (Mego et al. 2021), c-Met inhibitor Tivantinib (Feldman et al. 2013), ABL1 / 2 / KIT / PDGFRA / B inhibitor Imatinib (Einhorn et al. 2006; Piulats et al. 2007), BRAF / FGFR1 / KIT / PDGFRB / RAF1 inhibitor Sorafenib (Skoneczna et al. 2014), VEGFR1-3 / PDGFR / FGFR / KIT / c-Fms inhibitor Pazopanib (Necchi et al. 2017), mTOR inhibitor Everolimus (Fenner et al. 2018), or VEGFR2 / PDGFRB inhibitor Sunitinib (Feldman et al. 2010; Oechsle et al. 2011; Reckova et al. 2012) in heavily pretreated refractory GCT patients. Nevertheless, most of these studies did not stratify between GCT subtypes, so that no conclusion can be taken with regard to the efficacy for YST patients. Previously described success stories of treatment with tyrosine kinase inhibitors, as it has been shown for human epidermal growth factor receptor 2 (HER2) positive breast cancer (Schlam and Swain 2021) or chronic myeloid leukemia (Kim et al. 2017), are often based on specific mutational patterns, which are rarely found in GCT (Shen et al. 2018). Using mutational profiling as well as proteome-wide analyses, the here presented (pre-clinical) investigation identified (multikinase) inhibitors targeting CHEK1, AURKA - C, VEGFR1 - 3, FGFR1 / 2, PDGFRA / B, mTOR / AKT, RAF1, BRAF, PDGFRB, KIT, or PARP1 / 2 to be considered promising therapeutic options for (refractory) YST patients. As such, an in silico evaluation regarding the absorption, distribution, metabolism, excretion and toxicity (ADMET) of AZD4647, AZD7762, Danusertib, Nintedanib, OSU-03012, SNS-314, Sorafenib and Talazoparib using the 'ADMETlab 2.0' web platform revealed that most of these multikinase inhibitors had a good / moderate intestinal absorption and volume distribution (Additional file 7: Data S2). Since clearance was 2.067 - 7.25 mL / min / kg, hepatotoxicity should be considered upon treatment with these inhibitors (Additional file 7: Data S2). Moreover, treatment with the most sensitive inhibitors AZD7762, Danusertib, Nintedanib, OSU-03012, and SNS-314 specifically decreased AKT1 - 3 (S473), SRC (Y419), WNK1 (T60), and YES (Y426) phosphorylation especially in YST-R cells (Fig. 5C). Further, treatment with AZD7762, Danusertib, or SNS-314 enhanced JUN (S63) and p53 activity (in at least two of

the three evaluated phosphorylation sites), thereby offering novel combined treatment approaches (Fig. 5C).

It is generally believed that cisplatin resistance might occur due to a diminished import, enhanced export, increased detoxification, elevated DNA repair mechanisms, decreased apoptosis induction, and / or augmented alternating signaling pathways eventually resulting in the circumvention of drug-induced cytotoxicity (Galluzzi et al. 2012; Skowron et al. 2021a). Resistance mechanisms in GCT have been often reported to be involved in altered DNA repair mechanisms or apoptosis induction (Skowron et al. 2021a; Jacobsen and Honecker 2015; Vries et al. 2020). Here, we present the first description of putative resistance mechanisms specifically in YST, i. e., besides the previously mentioned role of p53, YST-R were prominently enriched in factors relevant for the translational initiation, RNA binding, immune response, ECM, and oxidative stress response. We identified factors relevant during MAPK signaling to be exclusively enriched in YST-R tissues (Fig. 5B). Additionally, resistant YST cells showed elevated activity of the AKT and p53 pathway. Hence, this study identified signaling cascades that could be targeted using multikinase inhibitors as an alternative treatment approach for (resistant) YST.

## Conclusion

This study offers a decisive groundwork for the understanding of molecular pathways resulting in cisplatin resistance of YST, proposes pertinent therapeutic strategies and offers alternative therapeutic options using (multikinase) inhibitors or an ADC targeting CLDN6.

## Abbreviations

ABTS	2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)
ADC	Antibody-drug-conjugate
ADMET	Absorption, distribution, metabolism, excretion and toxicity
CAF	Cancer-associated fibroblast
CAR	Chimeric antigen receptor
CC	Choriocarcinoma
EC	Embryonal carcinoma
ECM	Extracellular matrix
ELISA	Enzyme-linked immunosorbent assay
FDR	False discovery rate
FFPE	Formalin-fixed, paraffin-embedded
g	Grams
GCNIS	Germ cell neoplasia in situ
GCT	Germ cell tumor
h	Hour
HCC	Hepatocellular carcinoma
L	Liter
LD <sub>50</sub>	Lethal dose, 50%
LC-MS	Liquid chromatography coupled to mass spectrometry
M	Molar
m	Meter
min	Minute
miR	MicroRNA
MMAE	Monomethyl auristatin E
MNV	Multi-nucleotide variants

MSI	Microsatellite instability
Mut / Mb	Mutations per megabase
NS	Non-seminoma
p-	Phosphorylation
PCA	Principal component analysis
PGC	Primordial germ cell
R	Resistant
rpm	Rounds per minute
s	Second
SEM	Seminoma
SNV	Single nucleotide variants
TCGA	The Cancer Genome Atlas
TMB	Tumor mutational burden
TSO	TruSight Oncology 500 Sequencing
YST	Yolk-sac tumor
XTT	2,3-Bis-(2-methoxy-4-nitro-5-sulphophenyl)-2H-tetrazolium-5-carboxanilide

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s10020-023-00636-3>.

**Additional file 1: Figure S1.** A) Mutational status and mRNA expression profile of *CLDN6* in the GCT TCGA (The Cancer Genome Atlas) cohort as analyzed through <https://www.cbiportal.org/>. B) XTT cell viability assays of *CLDN6* antibody, *CLDN6*-ADC, or MMAE treated GCT cell lines and non-cancerous fibroblasts (MPAF) (24 - 96 h). C) Raw flow cytometry data indicating cell cycle distributions of GCT cell lines and MPAF treated with either *CLDN6* antibody alone (red), *CLDN6*-ADC (green), or MMAE (blue) in comparison with untreated controls (grey). D) Immunohistochemical evaluation of *CLDN6* in YST-R tissues (n = 10).

**Additional file 2: Figure S2.** A) Raw human phospho-kinase array of various cell lysates (GCT72, 1411H, NOY-1, MPAF). Corresponding membrane layout with the most prominent dots (marked in red) was used for quantification using ImageJ as shown in Fig. 3 A. B) Relative mRNA expression of potential YST target genes in GCT72, 1411H, NOY-1, and MPAF. *ACTB* and *GAPDH* were used as housekeeping genes. C) Raw data and D) densitometric evaluation of GCT72(-R) cells treated with AZD7762, Danusertib, Nintedanib, OSU-03012, or SNS-314 (24 h, LD<sub>50 72 h</sub>) in comparison to the solvent control (DMSO).

**Additional file 3: Figure S3.** A, B) Mutational status (A) and mRNA expression profiles (B) of potential targets to treat YST as found in the TCGA GCT cohort.

**Additional file 4: Figure S4.** A) XTT cell viability assays of GCT72 cells treated with indicated drugs for 24 - 96 h. B) XTT cell viability assays of GCT cell lines GCT72(-R), NOY-1(-R), 1411H, and non-cancerous fibroblasts (MPAF) treated with indicated drugs for 24 - 96 h. C) Raw flow cytometry data of cell cycle distributions of GCT cell lines and MPAF treated with indicated drugs in comparison with solvent controls.

**Additional file 5: Table S1.** List of used A) cell lines, B) inhibitors, C), oligonucleotides, and D) antibodies.

**Additional file 6: Data S1.** Raw and processed mass spectrometry data of YST(-R) tissues and TSO assay data.

**Additional file 7: Data S2.** Structures and ADMET (absorption, distribution, metabolism, excretion, toxicity) prediction of AZD4647, AZD7762, Danusertib, Nintedanib, OSU-03012, SNS-314, Sorafenib and Talazoparib as analyzed by the "ADMETlab 2.0" web platform (<https://admet.cbdd.com/>).

## Acknowledgements

We kindly thank Anna Pehlke and Olga Dschun for excellent technical assistance. We kindly thank PD Dr. Michèle Hoffmann, Dr. Christoph Oing, Prof. Dr. Michal Mego, Dr. Janet Shipley, and PD Dr. Michael Peitz for providing cell lines and healthy control cells (see Additional file 5; Table S1A for affiliations and provided cell lines). We kindly thank Prof. Dr. Tilman Rau and Mrs. Mara Walther (both Institute of Pathology, UK Düsseldorf) for providing YST tissues from the local "Universitätsklinikum" biobank (UK Düsseldorf). We thank Dr. Ernest

Kovacs (Levena Biopharma, San Diego, CA, USA) and Dr. Simon A. Sumer (Bio-Techne / R&D Systems, Wiesbaden, Germany) for the assistance regarding technical inquiries during the development of the CLDN6-ADC.

#### Author contributions

Conception and design: DN. Acquisition of data: MAS, FB, MK, KR, FP, AR, SK, KRJ, FP, SP, AS, DN. Analysis and interpretation of data: MAS, FB, MK, KR, FP, AR, SK, KRJ, FP, SP, AS, DN. Visualization: MAS, DN. Drafting of the manuscript: MAS, MK, DN. Critical revision of the manuscript: FB, KS, PS. Statistical analysis: SP, MAS. Obtaining funding: DN, FB. Administrative, technical, or material support: FB, KS, PS. Supervision: DN. All authors read and approved the final manuscript.

#### Funding

Open Access funding enabled and organized by Projekt DEAL. D. Nettersheim was supported by the Deutsche Forschungsgemeinschaft (NE 1861/8-1). F. Bremmer was supported by the ‘Wilhelm Sander-Stiftung’ (2016.041.1 / .2 / .3).

#### Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files. LC-MS raw data generated in this study can be accessed via ProteomeXchange (<http://www.proteomexchange.org>) (PXD039063).

#### Declarations

##### Ethics approval and consent to participate

The ethics committee of the Heinrich-Heine-University Düsseldorf raised no concerns on utilizing cell lines for in vitro experiments (vote 2018-178 to D. N.). The ethics committees of the Heinrich Heine University Düsseldorf and the University Medical Center Göttingen raised no concerns on performing the described experiments on GCT tissues from local biobanks (vote 2020.1247\_L1) to D. N.; vote 20/09/20 to F. B., vote 24/4/20 to P.S.).

##### Consent for publication

All authors are aware of this article and agreed on publication.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Department of Urology, Urological Research Laboratory, Translational UroOncology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Moorenstraße 5, 40225 Düsseldorf, Germany. <sup>2</sup>Institute of Pathology, University Medical Center Göttingen, Göttingen, Germany. <sup>3</sup>Molecular Proteomics Laboratory, Heinrich Heine University Düsseldorf, Düsseldorf, Germany.

Received: 25 November 2022 Accepted: 13 March 2023

Published online: 29 March 2023

#### References

- Adra N, et al. Phase II trial of pembrolizumab in patients with platinum refractory germ-cell tumors: a Hoosier Cancer Research Network Study GU14-206. *Ann Oncol*. 2018;29:209–14.
- Adra N, et al. A phase II study assessing the safety and efficacy of ASP1650 in male patients with relapsed refractory germ cell tumors. *Invest New Drugs*. 2022;40:1087–94.
- Ang LT, et al. A roadmap for human liver differentiation from pluripotent stem cells. *Cell Rep*. 2018;22:2190–205.
- Burmeister A, et al. Establishment and evaluation of a dual HDAC / BET inhibitor as a therapeutic option for germ cell tumors and other urological malignancies. *Mol Cancer Ther*. 2022;21:1674–88.
- Cao AT, Law C-L, Gardai SJ, Heiser RA. Abstract 5588: Brentuximab vedotin-driven immunogenic cell death enhances antitumor immune responses, and is potentiated by PD1 inhibition in vivo. *Cancer Res*. 2017;77:5588.
- Carpentier G. Protein Array Analyzer. Accessed 03 Feb 2022. <https://imagej.nih.gov/ij/macros/toolsets/ProteinArrayAnalyzer.txt>.
- Cerami E, et al. The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2:401–4.
- Che Y, et al. Late relapsing germ cell tumors with elevated tumor markers. *World J Urol*. 2021;40:363–71.
- Chen T, Liu Y-X, Huang L. ImageGP: an easy-to-use data visualization web server for scientific researchers. *iMeta*. 2022;9:e12415.
- Cheng L, et al. Testicular cancer. *Nat Rev Dis Prim*. 2018;4:29.
- Cheng ML, et al. Germ cell tumor molecular heterogeneity revealed through analysis of primary and metastasis pairs. *JCO Precision Oncol*. 2020;4:PO.20.00166.
- Cierna Z, et al. Prognostic value of programmed-death-1 receptor (PD-1) and its ligand 1 (PD-L1) in testicular germ cell tumors. *Ann Oncol*. 2016;27:300–5.
- D’Amour KA, et al. Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotechnol*. 2005;12:1534–41.
- da Fonseca LG, Reig M, Bruix J. Tyrosine kinase inhibitors and hepatocellular carcinoma. *Clin Liver Dis*. 2020;24:719–37.
- Dall’Olio FG, et al. Tumour burden and efficacy of immune-checkpoint inhibitors. *Nat Rev Clin Oncol*. 2022;19:75–90.
- De Giorgi U, et al. Olaparib as salvage treatment for advanced germ cell tumors after chemotherapy failure: results of the open-label, single-arm, IGG-02 phase II trial. *J Clin Oncol*. 2020;38:5058.
- de Vries G, Rosas-Plaza X, van Vugt MATM, Gietema JA, de Jong S. Testicular cancer: determinants of cisplatin sensitivity and novel therapeutic opportunities. *Cancer Treat Rev*. 2020;88:102054.
- Dennis G, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol*. 2003;4:P3.
- Dong J, et al. ADMETlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database. *J Cheminform*. 2018;10:29.
- Einhorn LH, Brames MJ, Heinrich MC, Corless CL, Madani A. Phase II study of imatinib mesylate in chemotherapy refractory germ cell tumors expressing KIT. *Am J Clin Oncol*. 2006;29:12–3.
- Fankhauser CD, et al. Frequent PD-L1 expression in testicular germ cell tumors. *Br J Cancer*. 2015;113:411–3.
- Feldman DR, et al. Phase II trial of sunitinib in patients with relapsed or refractory germ cell tumors. *Invest New Drugs*. 2010;28:523–8.
- Feldman DR, et al. A phase 2 multicenter study of tivantinib (ARQ 197) monotherapy in patients with relapsed or refractory germ cell tumors. *Invest New Drugs*. 2013;31:1016–22.
- Fenner M, et al. Everolimus in patients with multiply relapsed or cisplatin refractory germ cell tumors: results of a phase II, single-arm, open-label multicenter trial (RADIT) of the German Testicular Cancer Study Group. *J Cancer Res Clin Oncol*. 2018;145:717–23.
- Flyamer I. adjustText—automatic label placement for matplotlib. <https://github.com/Phlya/adjustText>. 2017.
- Fu Z, Li S, Han S, Shi C, Zhang Y. Antibody drug conjugate: the “biological missile” for targeted cancer therapy. *Signal Transduct Target Ther*. 2022;7:93.
- Galluzzi L, et al. Molecular mechanisms of cisplatin resistance. *Oncogene*. 2012;31:1869–83.
- Gao J, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6:11.
- Genin M, Clement F, Fattaccioli A, Raes M, Michiels C. M1 and M2 macrophages derived from THP-1 cells differentially modulate the response of cancer cells to etoposide. *BMC Cancer*. 2015;15:577.
- Gerst R, Hölzer M. PCAGO: an interactive tool to analyze RNA-Seq data with principal component analysis. *bioRxiv*. 2019. <https://doi.org/10.1101/433078>.
- González-Barrios R, et al. Genomic profile in a non-seminoma testicular germ-cell tumor cohort reveals a potential biomarker of sensitivity to platinum-based therapy. *Cancers (basel)*. 2022;14:2065.
- Günzel D, Fromm M. Claudins and other tight junction proteins. *Compr Physiol*. 2012;2:1819–52.
- Hughes CS, et al. Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol Syst Biol*. 2014;10:757.
- Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9:90–5.
- Ikeda K, et al. Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections. *J Histochem Cytochem*. 1998;46:397–403.

- Jacob JB, Jacob MK, Parajuli P. Review of immune checkpoint inhibitors in immuno-oncology. *Adv Pharmacol*. 2021;91:111–39.
- Jacobsen C, Honecker F. Cisplatin resistance in germ cell tumours: models and mechanisms. *Andrology*. 2015;3:111–21.
- Johansson MP, Maaheimo H, Ekholm FS. New insight on the structural features of the cytotoxic auristatins MMAE and MMAF revealed by combined NMR spectroscopy and quantum chemical modelling. *Sci Rep*. 2017;7:15920.
- Katz S, et al. SIGNAL: a web-based iterative analysis platform integrating pathway and network approaches optimizes hit selection from genome-scale assays. *Cell Syst*. 2021;12:338–52.
- Kawahara T, et al. Phase II trial of nivolumab monotherapy and biomarker screening in patients with chemo-refractory germ cell tumors. *Int J Urol*. 2022;29:741–7.
- Kim TH, et al. Spectrum of somatic mutation dynamics in chronic myeloid leukemia following tyrosine kinase inhibitor therapy. *Blood*. 2017;129:38–47.
- Kitayama S, et al. Testis-expressed gene 11 inhibits cisplatin-induced DNA damage and contributes to chemoresistance in testicular germ cell tumor. *Sci Reports*. 2022;12:18423.
- Koster R, et al. Cytoplasmic p21 expression levels determine cisplatin resistance in human testicular cancer. *J Clin Invest*. 2010;120:3594–605.
- Lobo J, et al. Detailed characterization of immune cell infiltrate and expression of immune checkpoint molecules PD-L1/CTLA-4 and MMR proteins in testicular germ cell tumors disclose novel disease biomarkers. *Cancers (Basel)*. 2019;11:1535.
- Lobo J, et al. p53 and MDM2 expression in primary and metastatic testicular germ cell tumors: association with clinical outcome. *Andrology*. 2020;8:1233–42.
- Mackensen A, et al. 958 BNT211: a phase I/II trial to evaluate safety and efficacy of CLDN6 CAR-T cells and vaccine-mediated in vivo expansion in patients with CLDN6-positive advanced solid tumors. *J Immunother Cancer*. 2021;9.
- Mackensen A, et al. BNT211-01: A phase I trial to evaluate safety and efficacy of CLDN6 CART cells and CLDN6-encoding mRNA vaccine-mediated in vivo expansion in patients with CLDN6-positive advanced solid tumors. *Ann Oncol*. 2022;33:808–869.
- Mayer F, et al. Microsatellite instability of germ cell tumors is associated with resistance to systemic treatment. *Cancer Res*. 2002;62:2758–60.
- McDermott MS, et al. Abstract 342: development and characterization of a novel anti-CLDN6 antibody drug conjugate for the treatment of CLDN6 positive cancers. *Cancer Res*. 2022;82:342.
- McGregor BA, et al. Results of a multicenter, phase 2 study of nivolumab and ipilimumab for patients with advanced rare genitourinary malignancies. *Cancer*. 2021;127:840–9.
- Mckinney W. Data structures for statistical computing in python. In: van der Walt S, Millman J, editors. *Proceedings of the 9th Python in Science Conference*, vol 445. 2010. p. 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>.
- Mego M, et al. Phase II study of avelumab in multiple relapsed/refractory germ cell cancer. *Invest New Drugs*. 2019;37:748–54.
- Mego M, et al. Phase II study of gemcitabine, carboplatin and veliparib in multiple relapsed/refractory germ cell tumors (GCTs). *J Clin Oncol*. 2021;39:e17009–e17009.
- Micke P, et al. Aberrantly activated claudin 6 and 18.2 as potential therapy targets in non-small-cell lung cancer. *Int J Cancer*. 2014;135:2206–14.
- Necchi A, et al. Pazopanib in advanced germ cell tumors after chemotherapy failure: results of the open-label, single-arm, phase 2 Pazotest trial. *Ann Oncol*. 2017;28:1346–51.
- Necchi A, et al. An open-label randomized phase 2 study of Durvalumab Alone or in combination with tremelimumab in patients with advanced germ cell tumors (APACHE): results from the first planned interim analysis. *Eur Urol*. 2019a;75:201–3.
- Necchi A, et al. Genomic features for therapeutic insights of chemotherapy-resistant, primary mediastinal nonseminomatous germ cell tumors and comparison with gonadal counterpart. *Oncologist*. 2019b;24:e142.
- Necchi A, et al. Genomic characterization of testicular germ cell tumors relapsing after chemotherapy. *Eur Urol Focus*. 2020;6:122–30.
- Oechsle K, et al. Preclinical and clinical activity of sunitinib in patients with cisplatin-refractory or multiply relapsed germ cell tumors: a Canadian Urologic Oncology Group/German Testicular Cancer Study Group cooperative study. *Ann Oncol*. 2011;22:2654–60.
- Park JS, Kim J, Elghiyat A, Ham WS. Recent global trends in testicular cancer incidence and mortality. *Medicine (Baltimore)*. 2018;97:e12390.
- Piulats JM, et al. Phase II multicenter study of imatinib in patients with chemorefractory germ cell tumors that express c-kit. *Cancer Res*. 2007;67:2648.
- Rafiq S, Hackett CS, Brentjens RJ. Engineering strategies to overcome the current roadblocks in CAR T cell therapy. *Nat Rev Clin Oncol*. 2020;17:147–67.
- Rasche L, Wäsch R, Munder M, Goldschmidt H, Raab MS. Novel immunotherapies in multiple myeloma - chances and challenges. *Haematologica*. 2021;106:255–2565.
- Reback J et al. pandas-dev/pandas: Pandas 1.2.3. Zenodo. 2021. <https://doi.org/10.5281/zenodo.3509134>.
- Reckova M, et al. Sunitinib in patients with cisplatin-refractory germ cell tumors. *Onkologie*. 2012;35:455–6.
- Reinhard K, et al. An RNA vaccine drives expansion and efficacy of claudin-CAR-T cells against solid tumors. *Science*. 2020;367:446–53.
- Romano A, et al. Immune off-target effects of Brentuximab Vedotin in relapsed/refractory Hodgkin Lymphoma. *Br J Haematol*. 2019;185:468–79.
- Schlam I, Swain SM. HER2-positive breast cancer and tyrosine kinase inhibitors: the time is now. *NPJ Breast Cancer*. 2021;7:56.
- Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;2012(9):671–5.
- Shen H, et al. Integrated molecular characterization of testicular germ cell tumors. *Cell Rep*. 2018;23:3392–406.
- Skoneczna IA, et al. Sorafenib monotherapy in patients with inoperable/recurrent germ cell tumors (GCT) refractory to chemotherapy: phase II study. *J Clin Oncol*. 2014;32:367–367.
- Skowron MA, et al. The developmental origin of cancers defines basic principles of cisplatin resistance. *Cancer Lett*. 2021a;519:199–210.
- Skowron MA, et al. The signal transducer CD24 suppresses the germ cell program and promotes an ectodermal rather than mesodermal cell fate in embryonal carcinomas. *Mol Oncol*. 2021b;16:982–1008.
- Skowron MA, et al. Profiling the 3D interaction between germ cell tumors and microenvironmental cells at the transcriptome and secretome level. *Mol Oncol*. 2022;16:3107–27.
- Szklarczyk D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47:D607–13.
- Tsimberidou AM, et al. Pembrolizumab in patients with advanced metastatic germ cell tumors. *Oncologist*. 2021;26:558–e1098.
- Ushiku T, Shinozaki-Ushiku A, Maeda D, Morita S, Fukayama M. Distinct expression pattern of claudin-6, a primitive phenotypic tight junction molecule, in germ cell tumours and visceral carcinomas. *Histopathology*. 2012;61:1043–56.
- van Horssen R, ten Hagen TLM, Eggermont AMM. TNF-alpha in cancer treatment: molecular insights, antitumor effects, and clinical utility. *Oncologist*. 2006;11:397–408.
- Waas M, et al. SurfaceGenie: a web-based application for prioritizing cell-type-specific marker candidates. *Bioinformatics*. 2020;36:3447–56.
- Waskom M. Seaborn: statistical data visualization. *J Open Source Softw*. 2012;6:3021.
- Wruck W, et al. The pioneer and differentiation factor FOXA2 is a key driver of yolk-sac tumour formation and a new biomarker for paediatric and adult yolk-sac tumours. *J Cell Mol Med*. 2021;25:1394–405.
- Xiong G, et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res*. 2021;49:W5–14.
- Yang S, Xiao H, Cao L. Recent advances in heat shock proteins in cancer diagnosis, prognosis, metabolism and treatment. *Biomed Pharmacother*. 2021;142: 112074.
- Zhang C, et al. Identification of claudin-6 as a molecular biomarker in pancreatic cancer through multiple omics integrative analysis. *Front Cell Dev Biol*. 2021;9: 726656.
- Zschäbitz S, Lasitschka F, Jäger D, Grüllich C. Activity of immune checkpoint inhibition in platinum refractory germ-cell tumors. *Ann Oncol*. 2016;27:1356–60.
- Zschäbitz S, et al. Response to anti-programmed cell death protein-1 antibodies in men treated for platinum refractory germ cell cancer relapsed after high-dose chemotherapy and stem cell transplantation. *Eur J Cancer*. 2017;76:1–7.

## Publisher's Note

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

## Author contributions to Manuscript 4

### **Targeting CLDN6 in germ cell tumors by an antibody-drug-conjugate and studying therapy resistance of yolk-sac tumors to identify and screen specific therapeutic options**

Margaretha A. Skowron

Acquisition of data, analysis and interpretation of data, visualization, drafting of the manuscript and statistical analysis.

Mara Kotthoff

Acquisition of data, analysis and interpretation of data and drafting of the manuscript.

Felix Bremmer

Acquisition of data, analysis and interpretation of data, critical revision of the manuscript, Statistical analysis, obtaining funding and administrative, technical, or material support.

Katja Ruhnke

Acquisition of data and analysis and interpretation of data.

Fatma Parmaksiz

Acquisition of data and analysis and interpretation of data.

Annika Richter

Acquisition of data and analysis and interpretation of data.

Stefan Küffer

Acquisition of data and analysis and interpretation of data.

Kirsten Reuter-Jessen

Acquisition of data and analysis and interpretation of data.

Stella Pauls

Acquisition of LC-MS/MS data for FFPE tissues, proteome data analysis, interpretation of data and statistical analysis.

Anja Stefanski

Acquisition of data and analysis and interpretation of data.

Philipp Ströbel

Critical revision of the manuscript and administrative, technical, or material support.

Kai Stühler

Critical revision of the manuscript and administrative, technical, or material support.

Daniel Nettersheim

Conception and design, acquisition of data, analysis and interpretation of data, visualization, drafting of the manuscript, obtaining funding and supervision.

## 9 Manuscript 5:

Characterizing the mutational burden, DNA methylation landscape, and proteome of germ cell tumor-related somatictype malignancies to identify the tissue-of-origin, mechanisms of therapy resistance, and druggable targets



## ARTICLE OPEN



# Characterizing the mutational burden, DNA methylation landscape, and proteome of germ cell tumor-related somatic-type malignancies to identify the tissue-of-origin, mechanisms of therapy resistance, and druggable targets

Felix Bremmer<sup>1</sup>, Pailin Pongratanakul<sup>2</sup>, Margaretha Skowron<sup>2</sup>, Yue Che<sup>3</sup>, Annika Richter<sup>1</sup>, Stefan Küffer<sup>1</sup>, Kirsten Reuter-Jessen<sup>1</sup>, Hanibal Bohnenberger<sup>1</sup>, Stella Pauls<sup>4</sup>, Catena Kresbach<sup>5</sup>, Ulrich Schüller<sup>5</sup>, Kai Stühler<sup>4</sup>, Philipp Ströbel<sup>1</sup>, Peter Albers<sup>3</sup> and Daniel Nettersheim<sup>3</sup>

© The Author(s) 2023

**BACKGROUND:** Germ cell tumors (GCT) might undergo transformation into a somatic-type malignancy (STM), resulting in a cell fate switch to tumors usually found in somatic tissues, such as rhabdomyosarcomas or adenocarcinomas. STM is associated with a poor prognosis, but the molecular and epigenetic mechanisms triggering STM are still enigmatic, the tissue-of-origin is under debate and biomarkers are lacking.

**METHODS:** To address these questions, we characterized a unique cohort of STM tissues on mutational, epigenetic and protein level using modern and high-throughput methods like TSO assays, 850k DNA methylation arrays and mass spectrometry.

**RESULTS AND CONCLUSIONS:** For the first time, we show that based on DNA methylation and proteome data carcinoma-related STM more closely resemble yolk-sac tumors, while sarcoma-related STM resemble teratoma. STM harbor mutations in FGF signaling factors (*FGF6/23*, *FGFR1/4*) highlighting the corresponding pathway as a therapeutic target. Furthermore, STM utilize signaling pathways, like AKT, FGF, MAPK, and WNT to mediate molecular functions coping with oxidative stress, toxin transport, DNA helicase activity, apoptosis and the cell cycle. Collectively, these data might explain the high therapy resistance of STM. Finally, we identified putative novel biomarkers secreted by STM, like EFEMP1, MIF, and DNA methylation at specific CpG dinucleotides.

*British Journal of Cancer* (2023) 129:1580–1589; <https://doi.org/10.1038/s41416-023-02425-5>

## INTRODUCTION

Testicular germ cell tumors (GCT) represent a heterogeneous group with different histological subtypes stratified into seminomas and non-seminomas [1, 2]. Based on histology, gene expression profiles and epigenetics, seminomas are considered to be the default developmental pathway of the precursor lesion germ cell neoplasia in situ (GCNIS), which itself is the result of a defective primordial germ cell development. In contrast, non-seminomas arise by reprogramming of GCNIS cells to a pluripotent embryonal carcinoma (EC) [1, 2]. EC are able to differentiate into cells of all three germ layers (teratoma) or into extra-embryonic tissues (yolk-sac tumor (YST), choriocarcinoma [1–3]).

A rare but deadly subtype of GCT is the somatic-type malignancy (STM), a secondary tumor component of non-seminomas that resembles cancers seen in other organs and tissues [4]. A STM is defined in the current WHO classification (5th edition) as an area of  $\geq 5$  mm diameter with a population of

atypical mesenchymal or epithelial cells [5]. These STM span a wide variety of tumors, including rhabdomyosarcomas, adenocarcinomas, and embryonic-type neuroectodermal tumors (ENET). STM occur with an incidence of 2–6% at any point of GCT development, but are mainly diagnosed at a metastatic stage in a post-chemotherapeutic setting [6]. Patients with STM face a poor prognosis with a 5-year survival rate of 50–60% due to resistance towards cisplatin-based chemotherapy [7, 8]. Unfortunately, treatment guidelines are still missing due to a lack of knowledge about this special group of cancers and their biology.

Most GCT-related STM are found in association with TER, leading to the assumption that TER is the tissue-of-origin [9–12]. Nevertheless, there are also STM occurring in GCT without TER and in association with YST, indicating that YST cells (in particular their mesenchymal component) might transform into STM as well [13–15].

So far, the developmental origin and the underlying molecular and (epi)genetic mechanisms of STM formation remain elusive.

<sup>1</sup>Institute of Pathology, University Medical Center Goettingen, Goettingen, Germany. <sup>2</sup>Department of Urology, Urological Research Laboratory, Translational UroOncology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. <sup>3</sup>Department of Urology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. <sup>4</sup>Molecular Proteomics Laboratory (MPL), Biological and Medical Research Centre (BMFZ), Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. <sup>5</sup>Institute of Neuropathology, University Hospital Hamburg-Eppendorf, Hamburg, Germany. ✉email: Daniel.Nettersheim@med.uni-duesseldorf.de

Since specific treatments are still lacking, further research on the origin and pathogenesis of STM and the identification of potential therapeutic targets, is warranted. Thus, this study characterized the molecular and (epi)genetic features of STM on mutational, DNA methylation, and proteome level to identify the key processes driving STM formation and related therapy resistance, the tissue-of-origin as well as new therapeutic options and novel biomarkers.

## MATERIAL AND METHODS

### GCT/STM tissues

All GCT/STM tissues included in this study were collected from local biobanks (Institutes of Pathology at University Hospital Düsseldorf and University Medical Center Göttingen). All samples were re-evaluated by a reference pathologist for type II GCT (F.B.). In this study, we analyzed a GCT-related STM cohort consisting of 13 adenocarcinomas, 7 rhabdomyosarcomas, 4 carcinomas not otherwise specified (NOS), 2 angiosarcomas, 2 sarcomas without lineage-specific differentiation, and 2 ENET ( $n = 30$  in total) (Data S1A). We included 10 TER and 5 YST without STM as controls (Data S1A). The diagnosis had been made according to the WHO criteria of STM [5]. The STM accompanying histology is also given in "Data S1A".

### Immunohistochemistry

Immunohistochemistry (IHC) has been performed as described earlier [15]. Briefly, antigen retrieval was carried out in citrate-buffer. The primary antibodies were incubated for 30 min (min) at room temperature (RT). Sections were incubated with a ready-to-use-HRP-labeled secondary antibody at RT for 25 min. The substrate DAB+ Chromogen system was used to visualize the antigen. Tissues were counterstained with Meyer's hematoxylin. An overview of all IHC results is given in Data S1A (Data S1A). In total, 26 samples were analyzed (10 adenocarcinoma, 3 carcinoma NOS, 8 rhabdomyosarcoma, 2 angiosarcoma, 1 sarcoma, 2 ENET).

### Nucleic acid isolation

The STM area was highlighted on H&E-stained slides prior to the analysis by a reference pathologist for GCT. Only the marked areas were isolated from the FFPE-slides. DNA was extracted from  $2 \times 5 \mu\text{m}$  FFPE slices using the InnuPREP FFPE DNA Kit on the InnuPure C16 System (Jena Analytika, Jena, Germany) according to manufacturer instructions. RNA was isolated from  $2 \times 5 \mu\text{m}$  slices using the Maxwell RNA extraction kit (Promega, Walldorf, Germany) according to manufacturer's recommendations. DNA and RNA concentrations were measured on the Qubit 3 Fluorometer (Thermo Scientific, Paisley, UK).

### 12p gain PCR analysis

A PCR analysis measuring the 12p gain status of STM tissues was performed exactly as published [16]. A fold change normalized to controls of  $> 2$  was set as a cut-off value for samples considered to harbor a 12p gain. In total, ten samples were analyzed (two adenocarcinoma, two carcinoma NOS, two rhabdomyosarcoma, one angiosarcoma, one sarcoma).

### Illumina TruSight Oncology 500 (TSO) analyses

DNA libraries were prepared using the hybrid capture-based TSO Library Preparation Kit (Illumina, San Diego, CA, USA) following the manufacturer's instructions (#1000000067621 v00). Library concentrations and peak heights were evaluated on a Tape Station (Agilent, Santa Clara, USA). Equal amounts of up to eight library samples were pooled and diluted to 4 nM.  $10 \mu\text{l}$  of the library pool was mixed in 0.1 M NaOH and incubated for 5 min at RT. The library was neutralized and diluted to 20 pM with  $990 \mu\text{l}$  HT1, mixed and kept on ice. To generate 200,000 clusters/ $\text{mm}^2$  the pool was diluted to 0.6 pM by the addition of  $1261 \mu\text{l}$  HT1,  $39 \mu\text{l}$  library (20 pM) and  $1 \mu\text{l}$  PhiX (20 pM). Libraries were sequenced on an Illumina NextSeq 500 instrument. The FastQ files were analyzed in CLC Biomedical Workbench (Qiagen). Reads were mapped to hg19 followed by initial variant calling. Then local realignments, primer clipping, and low-frequency variant calling were performed. False-positives were removed based on read quality and forward/reverse balance. All variants were checked manually for sequencing artefacts. The average coverage was  $> 500$  in all samples; the mutations had at least 50 variant reads. In total, 11 samples were analyzed (2 adenocarcinoma, 2 carcinoma NOS, 3 rhabdomyosarcoma, 1 angiosarcoma, 2 sarcoma).

### Illumina 850k DNA methylation array (850k array)

DNA was isolated from FFPE tissue using the ReliaPrep™ FFPE gDNA Miniprep System (Promega, Walldorf, Germany) according to manufacturer's instructions. 100–500 ng DNA were used for bisulfite conversion with the EZ DNA Methylation Kit (Zymo Research, Freiburg, i. B., Germany). Afterwards, the DNA Clean & Concentrator-5 (Zymo Research) and the Infinium HD FFPE DNA Restore Kit (Illumina) were used to clean and restore the converted DNA. Finally, the Infinium MethylationEPIC BeadChip (Illumina) was used to evaluate the methylation status of 850,000 CpG sites on an iScan device (Illumina). In total, 11 samples were analyzed (5 adenocarcinoma, 6 rhabdomyosarcoma, 5 TER, 4 YST).

### Liquid chromatography coupled to mass spectrometry (LC-MS)

For sample preparation, a modified FFPE tissue lysis protocol of Ikeda et al. was applied [17]. FFPE tissues were deparaffinized by shaking in  $500 \mu\text{l}$  Xylene for 5 min, followed by removal of the solvent and air-dry the residual solvent. Tissues were resuspended in  $200 \mu\text{l}$  lysis buffer (300 mM TRIS/HCl, 2% SDS, pH 8.0), shock-frozen in liquid nitrogen and immediately heated for 25 min at  $99^\circ\text{C}$  and 350 rounds per minute (rpm). Samples were ultrasonicated on ice for 20 min with 30 seconds (s) on/off cycles and then shook for 2 hours (h) at  $80^\circ\text{C}$  and 500 rpm followed by a second ultrasonication step. After centrifugation for 5 min at 3500 rpm, the pellet was resuspended in  $100 \mu\text{l}$  lysis buffer for a second extraction round. Supernatants were combined and protein concentration was determined using the Pierce 660 nm Protein Assay (Thermo Fisher Scientific, Idstein, Germany).

For LC-MS analysis a modified magnetic bead-based sample preparation protocol according to Hughes and colleagues were applied [18]. Briefly,  $20 \mu\text{g}$  total protein were reduced by adding  $10 \mu\text{l}$  300 mM DTT and shaking for 20 min at  $56^\circ\text{C}$  and 1000 rpm, followed by alkylation with the addition of  $13 \mu\text{l}$  100 mM IAA and incubation for 15 min in the dark.  $10 \mu\text{l}$  of a  $20 \mu\text{g}/\mu\text{l}$  bead stock (1:1 Sera-Mag SpeedBeads) were added to each sample. For protein aggregation capture, ethanol (EtOH) was added to a final concentration of 80% and incubated for 15 min at  $20^\circ\text{C}$ . After three rinsing steps with 80% EtOH and one rinsing step with 100% ACN, beads were resuspended in 50 mM TEAB buffer and digested with final 1:50 trypsin at  $37^\circ\text{C}$  and 1000 rpm overnight. Extra-digestion was carried out by adding trypsin (final 1:50) and shaking at  $37^\circ\text{C}$  and 1000 rpm for 4 h.  $500 \text{ ng}$  of each sample were subjected to LC-MS.

For the LC-MS acquisition an Orbitrap Fusion Lumos Tribrid Mass Spectrometer coupled to an Ultimate 3000 Rapid Separation liquid chromatography system equipped with an Acclaim PepMap 100 C18 column ( $75 \mu\text{m}$  inner diameter, 25 cm length, 2 mm particle size) as separation column and an Acclaim PepMap 100 C18 column ( $75 \mu\text{m}$  inner diameter, 2 cm length, 2 mm particle size) as trap column (all equipment from Thermo Fisher Scientific). A LC-gradient of 180 min was applied and the MS operated in positive mode with a scan range of 200–2000  $m/z$  at a resolution of 120,000. The capillary temperature was set to  $275^\circ\text{C}$ , the source voltage (V) to 1.5 kV, the normalized AGC target was set to 62.5% and the maximum injection time was 60 ms. HCD fragmentations were carried out within a cycle time of 2 s.

Data were analyzed by Proteome Discoverer (version 2.4.1.15, Thermo Fisher Scientific). RAW files were matched against the human Swissprot database (Download: 23.01.2020) and the Maxquant Contaminant database (Download: 20.02.2021), using SequestHT integrated in the LFO Tribid processing workflow (Thermo Fisher Scientific). The maximum number of missed cleavages was set to 2 and the peptide length was 6–144 amino acids. Precursor mass tolerance was set to 10 ppm and the fragment mass tolerance was 0.6 Dalton. All samples were analyzed in a match between run search. Post processing, peptides were ungrouped and filtered to 1% FDR on protein and peptide level and to all proteins identified with  $\geq 2$  peptides. Contaminants were filtered out.

In total, 46 samples were analyzed (7 adenocarcinoma, 5 carcinoma NOS, 10 rhabdomyosarcoma, 5 angiosarcoma, 2 sarcoma, 2 ENET, 10 TER, 5 YST).

### Online analyses tools and software

"The Cancer Genome Atlas" (TCGA) datasets were analyzed using "cBioPortal for Cancer Genomics" [19]. STRING was used to predict protein–protein-interaction by confidence [20]. DAVID was used to predict molecular and biological functions of proteins based on "Gene Ontology" (GO) [21]. "Phyton" was used to generate volcano and violin plots [22, 23]. Venn diagrams were generated by "Venny 2.1.0" [24]. Pearson's correlation matrices and heatmaps were generated by "Morpheus" (<https://software.broadinstitute.org/Morpheus>).

## RESULTS

In this study, we analyzed the molecular and (epi)genetic features of a cohort of GCT-related STM consisting of 13 adenocarcinomas, 7 rhabdomyosarcomas, 4 carcinomas not otherwise specified (NOS), 2 angiosarcomas, 2 sarcomas without lineage-specific differentiation, and 2 ENET (Fig. 1a) ( $n = 30$  in total). We included 10 TER and 5 YST without STM as controls (Data S1A).

Detailed clinical data were available from 21 STM patients. Ages ranged from 20 to 68 years (mean 34). The majority of patients (20/21; 95%) who developed a STM initially presented with metastasis with 5% in CS I (1), 28% in CS II (6), and 67% in CS III (14) (Fig. 1a). Prognosis of the IGCCCG risk classification was mainly favorable with good, intermediate and poor prognosis in 48% (10), 28% (6), and 24% (5) of patients, respectively (Fig. 1b). Prior to the diagnosis of STM, 90% of patients had received at least three cycles of platin-based chemotherapy (Fig. 1b). 2 (10%) patients presented with a STM at first diagnosis. At the time of STM diagnosis, 76% (16) and 14% (3) of patients had elevated AFP ( $> 7 \mu\text{g/l}$ ) or  $\beta$ -hCG ( $> 2 \text{ mU/ml}$ ), respectively (Fig. 1b).  $\beta$ -hCG was only elevated in combination with AFP (Fig. 1b). Normal tumor markers were found in 19% (4) (Fig. 1a; Data S1A). STM mainly manifested at retroperitoneal lymph nodes and retrocaval sites (Fig. 1b). Exemplary H&E stainings and IHC of typical markers for each STM entity are given in “Fig. 1c”, while an overview of all performed IHC stainings is given in “Data S1A”. Histologically, the adenocarcinomas were composed of neoplastic glands with nuclear atypia. In IHC, all cases were SALL4<sup>+</sup>, focally FOXA2<sup>+</sup> and CDX2<sup>+</sup> as well as AFP<sup>-</sup>, GPC3<sup>-</sup>, CK7<sup>-</sup> and TTF-1<sup>-</sup> in most cases. The proliferation rate (Ki67) was between 30 and 50%. The rhabdomyosarcomas were composed of spindled rhabdoid cells with pleomorphic nuclei. The IHC detected Myogenin<sup>+</sup> and Desmin<sup>+</sup> cells. The tumor cells were SALL4<sup>-</sup>, Caldesmon<sup>-</sup> and Actin<sup>-</sup>. The carcinomas NOS contained highly atypical cells with pleomorphic nuclei and without any noticeable pattern. The IHC detected pan-cytokeratin<sup>+</sup> and SALL4<sup>-</sup> cells. The proliferation rate was between 20 and 25%. The sarcoma NOS cells were completely pan-cytokeratin<sup>-</sup> with focal Actin<sup>+</sup> cells (without any noticeable pattern). The pleomorphic tumor cells showed a high proliferation rate ( $> 50\%$ ). The angiosarcomas showed anastomosing vascular spaces lined with atypical cells. Other parts showed a solid architecture with epithelioid or spindled cells. The IHC revealed CD31<sup>+</sup>, ERG<sup>+</sup>, CD34<sup>-</sup>, and SALL4<sup>-</sup> cells. The ENET samples contained small cells with minimal to modest pale eosinophilic cytoplasm and round to oval hyperchromatic nuclei. In IHC, the cells presented as pan-cytokeratin<sup>+</sup>, S100<sup>-</sup>, chromogranin<sup>-</sup>, and scattered synaptophysin<sup>+</sup>. The YST showed a variety of patterns composed of neoplastic glands with prismatic cells. The IHC detected SALL4<sup>+</sup>, FOXA2<sup>+</sup>, GPC3<sup>+</sup>, and AFP<sup>+</sup> cells. For TER, a typical arrangement of cells of all three germ layers (ectoderm, mesoderm, endoderm) was observed.

By a TSO analysis, we analyzed the mutational burden of the STM samples. We included 2 adenocarcinomas, 2 carcinomas NOS, 4 rhabdomyosarcomas, 1 angiosarcoma and 1 sarcoma (in total  $n = 10$ ). The rhabdomyosarcomas included a primary tumor (Rhabdo. 1.1) and a metastasis (Rhabdo. 1.2) of the same patient. All STM samples harbored the GCT-typical chromosome 12p gain, confirming their GCT origin (Fig. S1A) [16]. As expected for a GCT-derived malignancy, the tumor mutational burden (TMB) score (avg. 2.75 mutations/megabase) and microsatellite instability (MSI) score (avg. 2.45% unstable sites) were quite low in all STM (Fig. 2a). No correlation between TMB and MSI score was found (Fig. 2b). An overview of all detected genetic variants is given in “Data S1B, C”. All mutations classified as “(likely) benign” or known to be commonly distributed in the human population without any effect were excluded from further analyses. In adenocarcinomas, common alterations included *ASXL2* and *TP53*. In carcinomas NOS, frequent mutations included *FGF23*, *FGF6*, *GEN1*, *KRAS*, *MST1*, *PTPRD*, and *TP53* (Fig. 2c; Data S1B). Among the

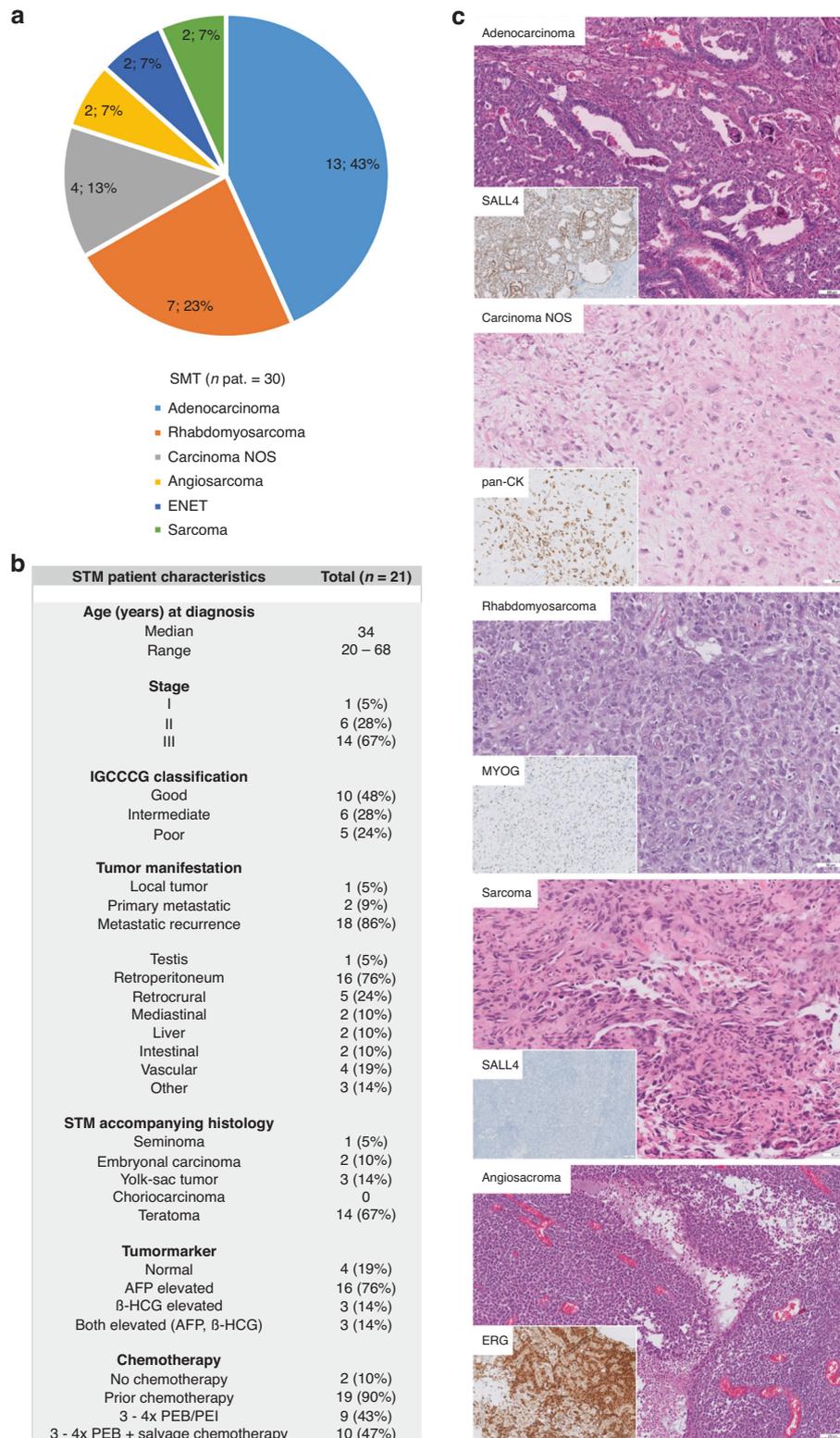
rhabdomyosarcomas, alterations in *FGFR1*, *KRAS*, and *MYC* were detected (Fig. 2c; Data S1B). Taking all STM samples together, *FGF6*, *KRAS*, and *TP53* were mutated in at least 70% of samples, while additionally *FGF23*, *FGFR1*, *FGFR4*, *MST1*, and *MYC* were mutated in at least 50% (Fig. 2c). SNP-mutations affecting *FGFR4* (c.1162G>A) and *TP53* (c.215C>G; c.380C>T) were classified as “pathogenic” or as affecting “drug response”, respectively, highlighting these mutated factors as putative therapeutic targets (Fig. 2c). We summarized drugs known to target the factors affected by mutations in “Fig. 2d”.

We next compared the mutational status of all genes found mutated by our TSO screen in any of the STM samples to the TCGA GCT cohort (149 samples) (Fig. S1B). *KRAS* (STM:  $> 70\%$ ; GCT: 16%) and *KIT* (STM: 30%; GCT: 14%) were the most amplified or missense mutated genes in the GCT cohort with the majority being mutated in seminomas (Fig. S1B). *FGF6* (STM:  $> 70\%$ ; GCT: 5%) and *FGF23* (STM:  $> 70\%$ ; GCT: 5%) were amplified in the same eight GCT samples (mainly non-seminomas) (Fig. S1B). Thus, *FGF6* and *FGF23* were also amplified in GCT, but with a considerably lower frequency than in STM. All other questioned genes were mutated with very low frequency (mainly  $< 1\%$ ) in GCT samples (Fig. S1B).

By using LC-MS, we analyzed the proteome of STM (adenocarcinomas ( $n = 7$ ), carcinomas NOS ( $n = 5$ ), angiosarcomas ( $n = 5$ ), ENET ( $n = 2$ ), rhabdomyosarcomas ( $n = 11$ ), sarcomas NOS ( $n = 3$ )). YST ( $n = 9$ ) and TER ( $n = 20$ ) served as controls. All samples showed comparable abundance levels of detected proteins (in total 3025) (Fig. S1C; Data S1D). As described initially, both TER and YST are believed to be the origin of STM [9–15]. To address this question, we compared STM to YST and TER samples by hierarchical clustering and in a Pearson’s correlation matrix (PCM) (Fig. 3a). Here, adenocarcinoma and carcinoma NOS (carcinoma-related entities) clustered to YST, while rhabdomyosarcomas, sarcomas and angiosarcomas (sarcoma-related entities) clustered to TER, (Fig. 3a). ENET cases were considerably different from the other entities, but were more similar to YST than TER (Fig. 3a).

By comparing the proteomes between STM and YST/TER using volcano plots, we highlighted the top 10 significantly enriched or depleted proteins (Fig. S2). We identified 363 common proteins (abundance  $> 10^6$ ) between all STM samples (Fig. 3b; Data S1E). We screened for putative interactions of these proteins and enriched GO categories using STRING (Fig. S3A, B). The proteins were predicted to be involved in translational initiation, regulation of cell death, the extracellular matrix (ECM), secretion, cell differentiation and the immune system (Fig. S3A). Functional clustering using DAVID demonstrated involvement of these proteins in toxin transport and telomere maintenance (cluster 4), positive regulation of signal transduction by p53 (6), antioxidant activity and removal of oxide radicals (clusters 8, 11), response to stress (8), DNA binding and regulation of DNA recombination (cluster 12), DNA helicase activity and duplex unwinding (cluster 35), chromatin remodeling (clusters 12, 20), regulation of apoptosis (clusters 19, 34), and the cell cycle (cluster 30) (Data S1E). Additionally, ECM-associated processes like ECM organization and collagen binding as well as adhesion and migration (clusters 19, 26, 29) were enriched (Data S1E). Furthermore, processes linked to activation of the (adaptive and innate) immune system, complement activation, Ig binding, neutrophil chemotaxis and B cell activation/signaling were enriched (clusters 21, 24, 25, 33) (Data S1E). To mediate these processes, interleukin, MAPK, WNT, FGFR2, PI3K-AKT, NOTCH, and HIF-1 signaling seem to be utilized via Rho GTPase effectors and ECM receptors (Fig. S3B). Furthermore, epigenetic processes like histone deacetylase binding, nucleosome positioning and chromatin remodeling (clusters 9, 21) were enriched, suggesting that development of STM is accompanied by epigenetic alterations (clusters 12, 17, 20) (Data S1E).

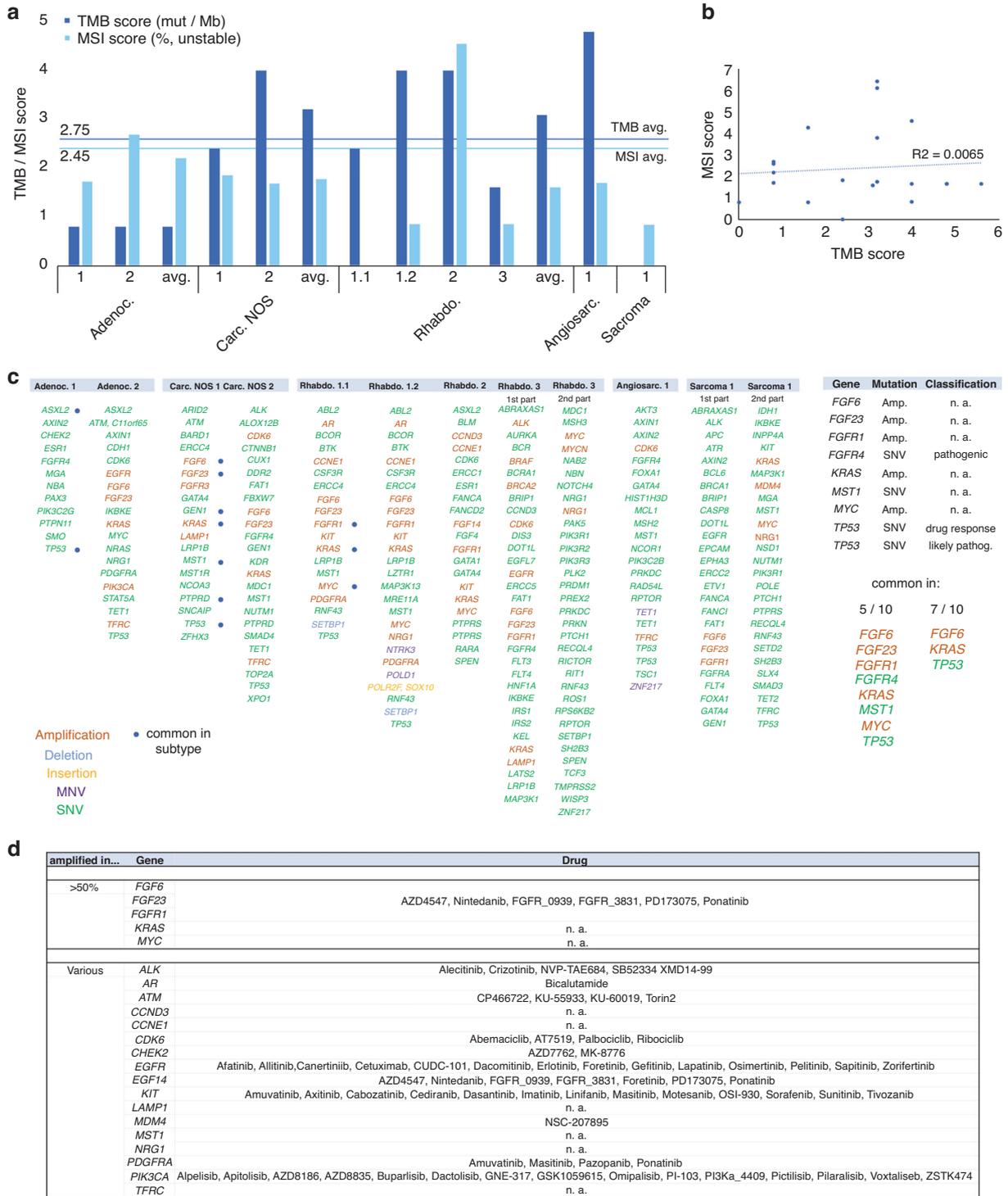
Comparing all proteins commonly found in STM entities to proteins found in TER and YST, an overlap of 54.5% (258 proteins)



**Fig. 1 Clinical and histological features of the STM cohort. a** Pie chart summarizing distribution of the various STM entities analyzed in this study. **b** Clinical parameters of the STM cohort (at diagnosis of STM) from the University Hospital Düsseldorf (Department of Urology) analyzed in this study. **c** Exemplary H&E stainings of each STM entity and IHC staining of typical marker proteins.

was found, while 29 proteins were exclusively found in STM (Fig. 3c). 22 of these 29 (75.8%) proteins were predicted to interact with each other, and were mainly related to focal adhesion, extracellular exosomes/vesicles, apoptosis, cellular response to

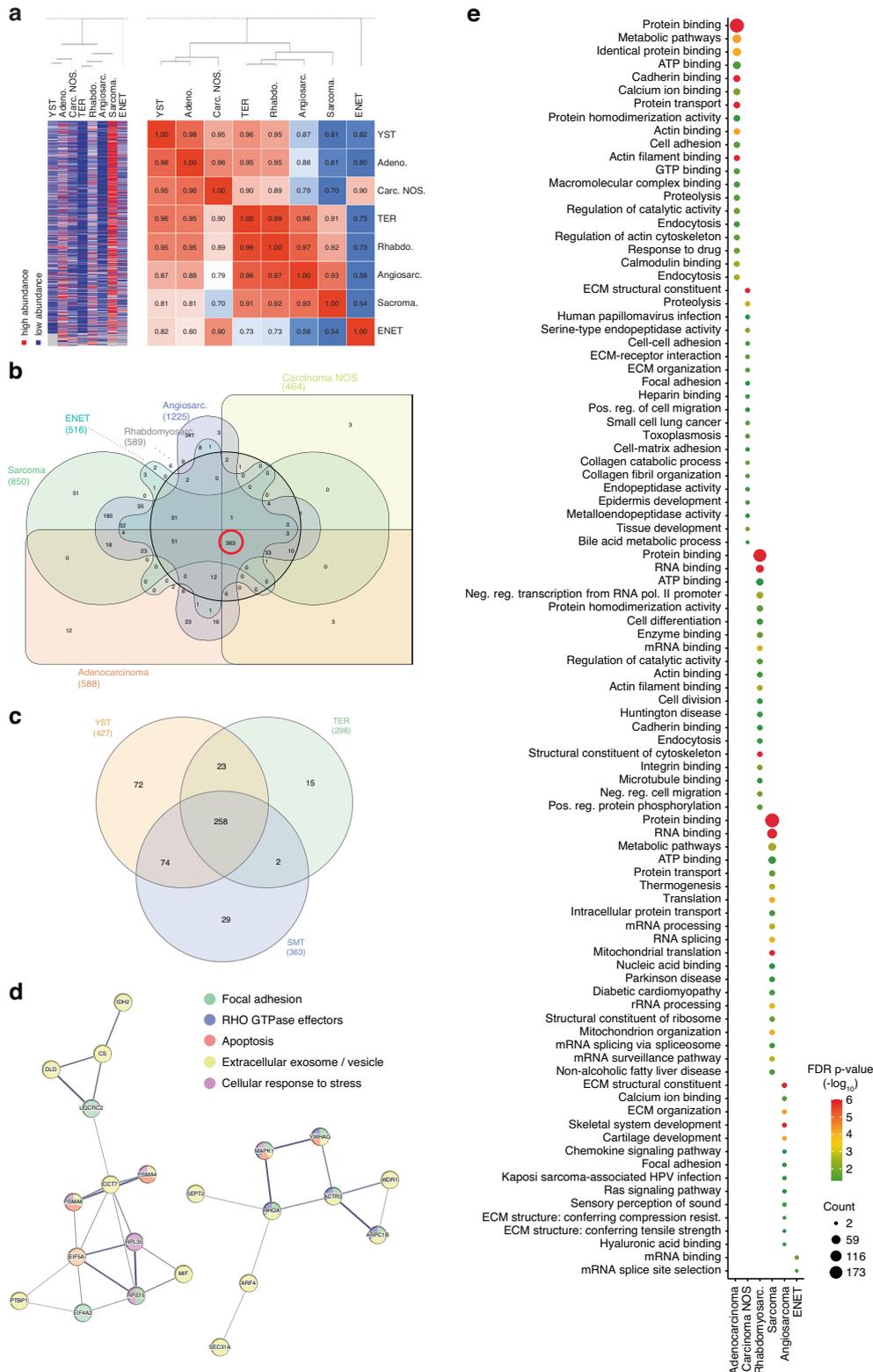
stress, RHO GTPase effectors, and MAPK signaling (Fig. 3d). Two proteins, EFEMP1 and MIF are extracellular factors or cytokines secreted exclusively by the STM, highlighting these proteins as putative biomarkers (Data S1E).



**Fig. 2 Analyzing druggable mutations of STM tissues. a, b** Illustration of the tumor mutational burden (TMB; mutations/megabase) and microsatellite instability score (MSI; % unstable) (a) and the ratio of both parameters (b) in STM samples analyzed by the TSO assay. **c** All detected mutations in indicated STM samples. Blue dots label mutations found in all samples of a STM subgroup. MNV: multiple nucleotide variants; SNV: single nucleotide variants. **d** Overview of drugs targeting found amplified genes/signaling factors.

Next, we searched for unique features of each STM entity by the DAVID algorithm (Fig. 3e; Data S1L). Proteins found exclusively in adenocarcinomas were linked to cell adhesion and migration, cadherin binding, endocytosis, response to drug and hypoxia, oxidoreductase activity, and regulation of angiogenesis (Fig. 3e). In carcinomas NOS, processes mainly related to the ECM (structure, organization, receptor-interaction, collagen catabolic process/fibrin

organization, metalloendopeptidase activity), adhesion, and migration were found (Fig. 3e). In rhabdomyosarcomas, unique proteins were associated with endocytosis, cadherin and integrin binding, cell differentiation (multicellular organism development, cardiac muscle contraction, response to TGF-β stimulus), and cell division (Fig. 3e). In sarcomas NOS, unique proteins were linked to regulation of NFκB signaling, RNA regulation (splicing, rRNA



**Fig. 3 Analyzing the proteome of STM tissues.** **a** A heatmap including hierarchical clustering and a Pearson's correlation matrix illustrate similarities and differences in the proteome (abundance > 10<sup>7</sup>) between the various STM groups as well as YST and TER. By Venn diagrams, shared and unique proteins (abundance > 10<sup>7</sup>) were identified between the STM entities (**b**) and compared to YST/TER (**c**). 363 proteins were found in all analyzed STM entities (**b**, red circle). **d** STRING-based protein-protein-interaction prediction of proteins commonly found in STM entities, but not in TER or YST. **e** DAVID-based GO and KEGG screen for biological processes and functions related to the proteins found exclusively in each STM entity.

processing, mRNA surveillance, spliceosome), and protein biosynthesis and trafficking (ribosome structure, protein transport, mitochondrial translation, ribosome biogenesis) (Fig. 3e). In angiosarcomas, processes related to the ECM (organization, structure, compression resistance, tensile strength, hyaluronic acid and heparin binding), adhesion, chemokine signaling pathways (RAS signaling), and mesodermal differentiation (skeletal system development, cartilage development) were found (Fig. 3e). Taken together, several key molecular functions are shared between STM (ECM interaction, molecule trafficking, adhesion, migration), although each entity engages different proteins to realize these functions.

To analyze differences in the DNA methylation (5mC) landscape, we performed Illumina 850k DNA methylation arrays. We included the two most common STM subtypes (i.e., adenocarcinomas ( $n = 5$ ) and rhabdomyosarcoma ( $n = 5$ )), while YST ( $n = 5$ ) and TER ( $n = 5$ ) served as controls (Data S1F). On a global level, compared to YST and adenocarcinomas, TER and rhabdomyosarcomas showed a higher amount of hypermethylated (>80%) sites, while YST and adenocarcinomas presented with a higher proportion of CpG dinucleotides with intermediate (20–80%) 5 mC levels (Fig. 4a, b). The average 5mC levels were similar between TER and rhabdomyosarcomas (49.4 and 48.4%), followed by YST and adenocarcinomas with slightly lower levels (44.7 and 44.4%) (Fig. 4a). By performing hierarchical clustering and a PCM, we demonstrated that YST and adenocarcinomas grouped to each other, while TER grouped with rhabdomyosarcomas (Fig. 4c). When sorting the 5mC data for regions showing only hypo- (< 20%) or hypermethylation (> 80%), followed by screening for distribution across genomic regions/CpG islands, we found that hypermethylated regions were strongly associated with gene bodies (i.e., coding regions) and open sea (i.e., not in CpG island context), while hypomethylated regions were mainly found at transcription start sites (TSS200, TSS1500) and in CpG island context (Fig. 4d). No considerable differences regarding 5mC distribution were observed between STM and YST/TER (Fig. 4d). We compared all CpG dinucleotides found hypo- (< 20%) or hypermethylated (> 80%) in adenocarcinomas or rhabdomyosarcomas to the CpG dinucleotides identified in YST or TER (Fig. 4e, f). Here, a considerable overlap of hypomethylated CpG dinucleotides was found between adenocarcinomas and YST, while in rhabdomyosarcomas a big proportion of hypermethylated CpG overlapped with TER, again reflecting the different 5mC distributions between adenocarcinomas/YST and rhabdomyosarcomas/TER (Figs. 4e, f; 1a, b).

By volcano plots, we identified differentially methylated CpG dinucleotides between adenocarcinomas and rhabdomyosarcomas compared to YST/TER (Fig. S4; Data S1G, H). We grouped the identified CpG dinucleotides (initial  $\Delta 5mC > 60\%$ ) for their ability to discriminate a given STM from the other types (Fig. 4g). These hypermethylated CpG dinucleotides might serve as epigenetic biomarkers to detect the occurrence of STM, e.g., by screening cell-free DNA.

Additionally, we correlated 5mC data (with hypo- (< 20%) or hypermethylation (> 80%) in at least 60% of all questioned CpG dinucleotides linked to an annotated gene) to the proteome data (abundance >  $10^6$ ) (Data S1I). In adenocarcinomas, we found 418 hypomethylated genes correlating to protein production and 384 genes/proteins in rhabdomyosarcomas, of which 326 were found shared between both STM entities (Fig. S4B; Data S1I). We found no hypermethylated genes correlating to protein production (Fig. S4B; Data S1I).

## DISCUSSION

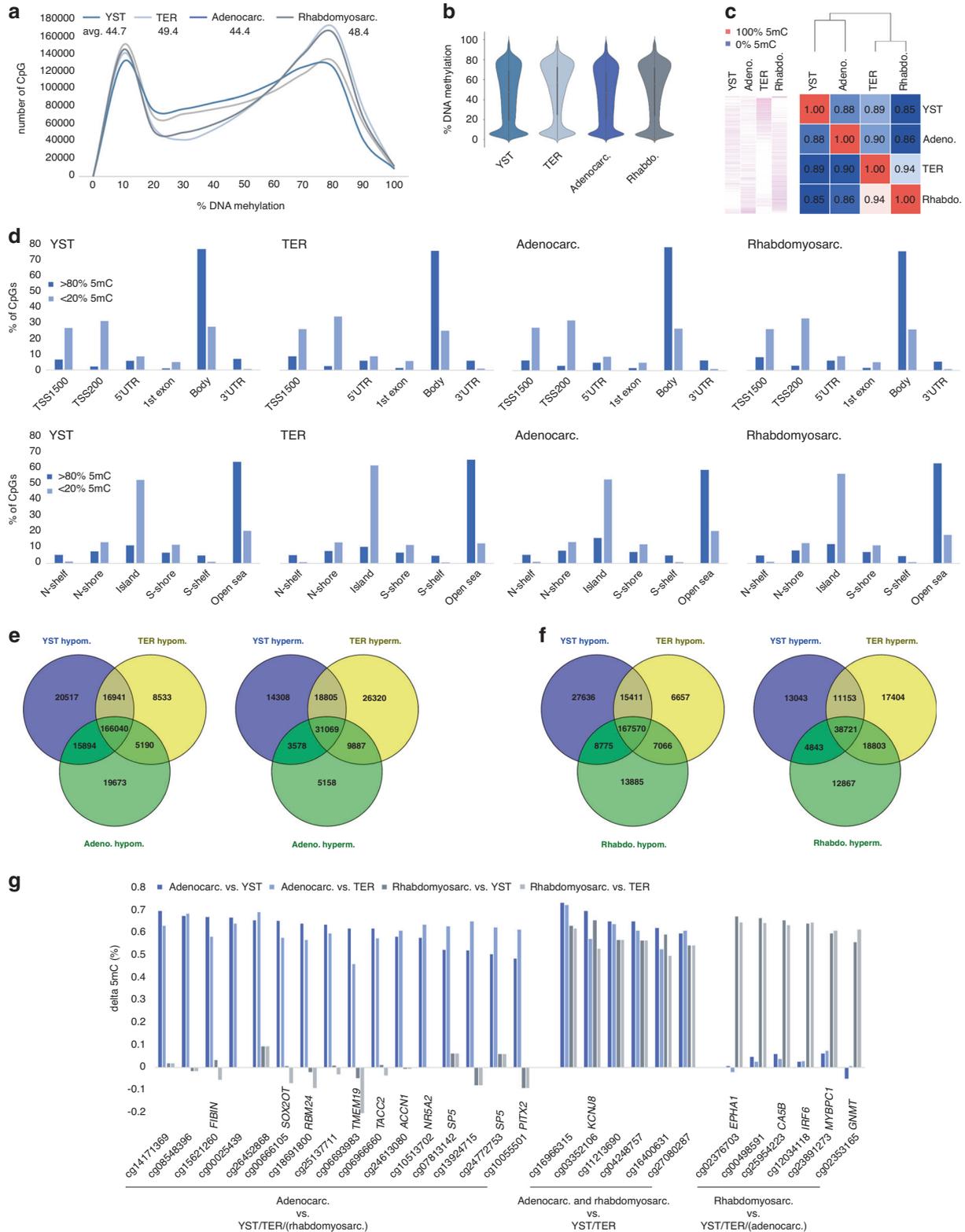
In this study, we characterized various GCT-related STM subtypes at the mutational, DNA methylation and proteome level and compared them to YST and TER.

The overall mutational burden including amplification fold changes were GCT typically low in STM, suggesting that mutations are not a crucial driver of STM formation. Nevertheless, our data and the correlation to the TCGA GCT cohort suggest that mutations detected in STM arose during formation of STM and are not generally detectable in GCT. We found amplifications in oncogenes, like *KRAS* or *MYC*, or mutations in *TP53*, which might affect drug response (c.215C>G), as well as FGF signaling factors might contribute to the aggressive character of STM by triggering proliferation, survival and anti-apoptotic signals. With mutations found in *FGF6*, *FGF23*, *FGFR1*, and *FGFR4* in at least 50% of the samples, FGF signaling seems to be a priority target of mutational events. There are some drugs available, mainly small molecule inhibitors and receptor-tyrosine-kinase inhibitors, targeting the FGF signaling cascade, i.e., AZD4547 (targeting FGFR1-3, but not FGFR4), Nintedanib, FGFR\_0939, FGFR\_3821, PD173075 and Ponatinib. Several completed or ongoing clinical trials screening some of these drugs were found (clinicaltrials.gov); AZD4547: 12, Nintedanib: 164, Ponatinib: 60. So, several FGF signaling related therapeutic options for treatment of STM are available and should be screened in follow-up studies and eventually clinical trials. Of note, although mutations in *KRAS*, *MYC*, and *TP53* were detected, to date no drugs targeting the specific mutations found in this study are available.

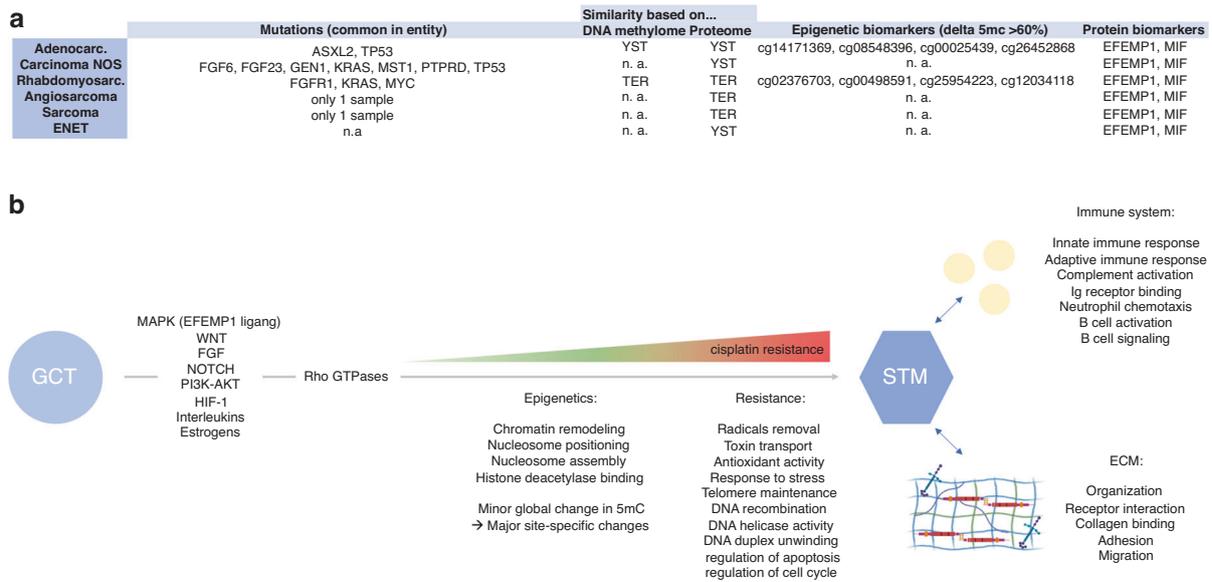
As found by LC-MS, the STM entities commonly utilize MAPK, WNT, FGF, NOTCH, PI3K-AKT, and HIF-1 signaling to mediate processes like response to oxidative stress, toxin transport, oxidant detoxification, DNA helicase activity, DNA duplex unwinding, the cell cycle and apoptosis (Fig. 5b). In combination with the frequently found SNV in *TP53* (c.215C>G), which might affect drug response, these processes might contribute to the insensitivity of the STM entities towards the cisplatin-based therapy by affecting key steps of cisplatin turnover, like influx/efflux, DNA repair, formation of radicals and (oxidative) stress caused by the therapy. Furthermore, ECM- and immune system-related processes were considerably enriched in all STM, pointing at a close interaction with the surrounding microenvironment including immune cells (Fig. 5b). Some proteins mediating the related biological functions were also found in YST and TER, suggesting that these GCT entities, which are also known for their high insensitivity towards cisplatin, might utilize similar mechanisms as the STM to increase the insensitivity towards cisplatin. Nevertheless, 29 proteins involved in regulation of apoptosis, stress response and adhesion as well as extracellular secretion were exclusively found in STM. Additionally, MAPK signaling related molecules (MAPK1, 14-3-3-gamma, RhoA) and the EGF ligand EFEMP1, which has been shown to activate MAPK signaling in pancreatic adenocarcinomas, were enriched in STM compared to YST/TER [25]. Thus, in STM these proteins and MAPK signaling triggering survival and growth might further contribute to cisplatin resistance.

Our study also highlighted putative biomarkers for SMT. With EFEMP1 and MIF, we identified two proteins exclusively secreted by the STM, which might serve as liquid biomarkers of STM, e.g., by blood sample screening in GCT patients (Fig. 5b). Additionally, we identified several hypermethylated CpG dinucleotides, which might serve as epigenetic biomarkers to detect the occurrence of STM, e.g., by screening cell-free DNA (Fig. 5b).

Regarding the tissue-of-origin of STM, based on the proteome adenocarcinoma, carcinoma NOS and ENET were more similar to YST, while rhabdomyosarcomas, angiosarcomas and sarcomas NOS were more closely related to TER (Fig. 5a). Taking the DNA methylation pattern into account, again a similarity between adenocarcinomas and YST as well as rhabdomyosarcomas and TER was demonstrated (Fig. 5a). The clinical data related to our cohort showed that YST (14%) and TER (67%) were the prevalent STM accompanying histology and in 76% of all cases elevated AFP levels were detected. These data support the hypothesis that both, YST and TER, are tissues-of-origin for the various STM



**Fig. 4 Analyzing the DNA methylation profile of STM tissues. a** Distribution of DNA methylation levels (%) across all analyzed CpG dinucleotides. **b** A violin plots illustrates genome-wide distribution of DNA methylation levels. **c** A heatmap and a Pearson's correlation matrix including hierarchical clustering illustrates and compares DNA methylation data, respectively. **d** Distribution of DNA hypo- (< 20%) and hypermethylated (> 80%) CpG dinucleotides across genomic regions/CpG islands. Venn diagrams comparing hyper- and hypomethylated CpG dinucleotides in adenocarcinomas (**e**) and rhabdomyosarcomas (**f**) with YST and TER. **g** Putative epigenetic biomarkers for adenocarcinomas and/or rhabdomyosarcomas based on the DNA methylation status of single CpG dinucleotides.



**Fig. 5 Key findings of this study. a** Summary of found mutations common in each STM entity as well as of similarities of STM entities to YST/TER on DNA methylation and proteome level. **b** Summary of molecular and epigenetic processes commonly found in STM entities putatively mediating therapy resistance and interaction with cells of the immune system and the ECM. Parts of this figure were generated by biorender.com.

entities. 90% of patients received at least three cycles of chemotherapy before diagnosis of a STM, suggesting that formation of a STM represents a therapy escape mechanisms for YST/TER cells. Importantly, formation of YST and TER seems to be an escape mechanism itself, since mostly YST and TER remain after chemotherapy regimen and are the leading cause of GCT-related death. Thus, the development of YST or TER from EC under therapy and eventually a STM represents an escalating cascade of escape mechanisms for GCT cells enabling survival.

During the submission/revision process of this article, Wyvekens et al. molecularly and epigenetically characterized a STM cohort of 36 male patients [26]. There, the authors found mutations in *KRAS* and *TP53* in 28% of cases each, which is in line with our findings in the mutational screen [26]. Similar to our TSO analysis, Wyvekens et al. found no oncogenic gene fusions in nine patient samples. Regarding DNA methylation, Wyvekens et al. detected distinct DNA methylation patterns for STM (ENET and rhabdomyosarcoma) and GCT samples, which is again in line with our 850k array analysis.

### Summary and outlook

Together with the article published by Wyvekens et al., both studies shed light on the molecular and (epi)genetic features of STM in a unique cohort of patient material providing comprehensive mutation, proteome and DNA methylation data as starting point for future studies. For the first time, we show that on a molecular level carcinoma-related STM more closely resemble YST, while sarcoma-related STM resemble TER. Additionally, we identified common mutations as well as molecular and epigenetic mechanisms contributing to the therapy resistance of STM. Finally, we identified new STM biomarkers and therapeutic options to treat STM patients, which should be translated into clinical testing.

### Limitations

Limitations of this study are the relatively small number of samples analyzed for epigenetic and genomic changes, which is due to the rarity of the STM. Nevertheless, in general our cohort represents one of the largest cohorts analyzed in the field, but studying more STM cases to confirm and verify our data would be of benefit. Additionally, a molecular and epigenetic similarity between tumor types does not necessarily indicate definitive

evolution from a precursor tumor subtype. Further, our cohort lacks the primary tumors of each STM patient, which would be an important control to recapitulate tumor evolution and STM formation with regard to mutations, epigenetics and changes on protein level. STM are not part of the TCGA GCT cohort, thus, comparing our findings to TCGA data is only possible for GCT in non-STM context. Furthermore, there is a lack of appropriate GCT-related STM model systems, i.e., cell lines are not available and setting up ex vivo cultures of these rarely occurring STM might be very time challenging and quite hard to organize. Thus, functional experiments or in vitro drug screenings are limited or not possible, respectively. Additionally, although we identified several drugs putatively suitable to target STM, setting up clinical trials is also very challenging due to the rarity of the STM phenomenon.

### DATA AVAILABILITY

The datasets and computer code produced in this study are available in the following databases: 850k DNA methylation data: Gene expression omnibus (GSE219033); LC-MS data: ProteomeXchange (PXD039546).

### REFERENCES

- Cheng L, Albers P, Berney DM, Feldman DR, Daugaard G, Gilligan T, et al. Testicular cancer. *Nat Rev Dis Prim.* 2018;4:1–24.
- Oosterhuis JW, Looijenga LHJ. Testicular germ-cell tumours in a broader perspective. *Nat Rev Cancer.* 2005;5:210–22.
- Müller MR, Skowron MA, Albers P, Nettersheim D. Molecular and epigenetic pathogenesis of germ cell tumors. *Asian J Urol.* 2020;8:144–54.
- Hwang MJ, Hamza A, Zhang M, Tu SM, Pisters LL, Czerniak B, et al. Somatic-type malignancies in testicular germ cell tumors: a clinicopathologic study of 63 cases. *Am J Surg Pathol.* 2022;46:11–17.
- Moch H, Amin MB, Berney DM, Compérat EM, Gill AJ, Hartmann A, et al. The 2022 World Health Organization classification of tumours of the urinary system and male genital organs—part a: renal, penile, and testicular tumours. *Eur Urol.* 2022;82:458–68.
- Little JS, Foster RS, Ulbright TM, Donohue JP. Unusual neoplasms detected in testis cancer patients undergoing post-chemotherapy retroperitoneal lymphadenectomy. *J Urol.* 1994;152:1144–51.
- Rice KR, Magers MJ, Beck SDW, Cary KC, Einhorn LH, Ulbright TM, et al. Management of germ cell tumors with somatic type malignancy: Pathological features, prognostic factors and survival outcomes. *J Urol.* 2014;192:1403–9.

8. Spiess PE, Pisters LL, Liu P, Pettaway CA, Kamat AM, Gomez JA, et al. Malignant transformation of testicular teratoma: a chemoresistant phenotype. *Urol Oncol Semin Orig Investig.* 2008;26:595–9.
9. Ahmed T, Bosl GJ, Hajdu SI. Teratoma with malignant transformation in germ cell tumors in men. *Cancer* 1985;56:860–3.
10. Ulbright TM, Loehrer PJ, Roth LM, Einhorn LH, Williams SD, Clark SA. The development of non-germ cell malignancies within germ cell tumors. A clinicopathologic study of 11 cases. *Cancer* 1984;54:1824–33.
11. Colecchia M, Necchi A, Paolini B, Nicolai N, Salvioni R. Teratoma with somatic-type malignant components in germ cell tumors of the testis: a clinicopathologic analysis of 40 cases with outcome correlation. *Int J Surg Pathol.* 2011;19:321–7.
12. Mikuz G, Colecchia M. Teratoma with somatic-type malignant components of the testis. A review and an update. *Virchows Arch.* 2012;461:27–32.
13. Magers MJ, Kao CS, Cole CD, Rice KR, Foster RS, Einhorn LH, et al. “Somatic-type” malignancies arising from testicular germ cell tumors: a Clinicopathologic study of 124 cases with emphasis on glandular tumors supporting frequent yolk sac tumor origin. *Am J Surg Pathol.* 2014;38:1396–409.
14. Michael H, Ulbright TM, Brodhecker CA. The pluripotential nature of the mesenchyme-like component of yolk sac tumor. *Arch Pathol Lab Med.* 1989;113:1115–9.
15. Ulbright TM, Roth LM. Recent developments in the pathology of germ cell tumors. *Semin Diagn Pathol.* 1987;4:304–19.
16. Fichtner A, Richter A, Filmar S, Gaisa NT, Schweyer S, Reis H, et al. The detection of isochromosome i(12p) in malignant germ cell tumours and tumours with somatic malignant transformation by the use of quantitative real-time polymerase chain reaction. *Histopathology.* 2021;78:593–606.
17. Ikeda K, Monden T, Kanoh T, Tsujie M, Izawa H, Haba A, et al. Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections. *J Histochem Cytochem.* 1998;46:397–403.
18. Hughes CS, Moggridge S, Müller T, Sorensen PH, Morin GB, Krijgsveld J. Singlepot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc.* 2019;14:68–85.
19. The Cancer Genome Atlas. TCGA. National Cancer Institute, National Human Genome Research Institute. 2013
20. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47:D607–13.
21. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 2003;4:418.
22. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng.* 2007;9:90–95.
23. Waskom M. Seaborn: Statistical Data Visualization. Seaborn. 2012.
24. Oliveros JC. Venny. An interactive tool for comparing lists with Venn’s diagrams. Accessed 8 Jan 2022.
25. Camaj P, Seeliger H, Ischenko I, Krebs S, Blum H, De Toni EN, et al. EFEMP1 binds the EGF receptor and activates MAPK and Akt pathways in pancreatic carcinoma cells. *Biol Chem.* 2009;390:1293–302.
26. Wyvekens N, Sholl LM, Yang Y, Tran I, Vasudevaraja V, Dickson BC, et al. Molecular correlates of male germ cell tumors with overgrowth of components resembling somatic malignancies. *Mod Pathol.* 2022;35:1966–73.

## ACKNOWLEDGEMENTS

We thank Anna Pehlke and Olga Dschun for excellent technical assistance.

## AUTHOR CONTRIBUTIONS

Conception and design: DN, FB. Acquisition of data: DN, FB, PP, MAS, YC, AR, SK, KRJ, HB, SP, CK. Analysis and interpretation of data: DN, FB, HB, US, KS, PS. visualization: DN, MAS, PP. Drafting of the manuscript: DN, FB. Critical revision of the manuscript: PS, PA. Statistical analysis: DN, FB, PP, MAS, YC, HB, SP, CK. Obtaining funding: DN, FB. Administrative, technical, or material support: DN, FB, KS, US, PP, PA. Supervision: DN, FB.

## FUNDING

DN is supported by the “Deutsche Forschungsgemeinschaft” (NE 1861/8-1). FB is supported by the “Wilhelm-Sander-Stiftung” (2016.041.1./2./3). Open Access funding enabled and organized by Projekt DEAL.

## COMPETING INTERESTS

The authors declare no competing interests.

## ETHICS APPROVAL

The ethics committees of the Heinrich Heine University Düsseldorf and the University Medical Center Göttingen raised no concerns on performing described experiments on used GCT/STM tissues from local biobanks (vote 2020.1247(\_1) to DN; vote 20/09/20 to FB, vote 24/4/20 to PS). Informed consent for use of samples in research was obtained from all subjects.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41416-023-02425-5>.

**Correspondence** and requests for materials should be addressed to Daniel Nettersheim.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

## Author contributions to Manuscript 5

### **Characterizing the mutational burden, DNA methylation landscape, and proteome of germ cell tumor-related somatic-type malignancies to identify the tissue-of-origin, mechanisms of therapy resistance, and druggable targets**

#### Felix Bremmer

Conception and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript, statistical analysis, obtaining funding, administrative, technical, or material support and supervision.

#### Pailin Pongratanakul

Acquisition of data, visualization, statistical analysis and administrative, technical, or material support.

#### Margaretha Skowron

Acquisition of data, visualization, statistical analysis and administrative, technical, or material support.

#### Yue Che

Acquisition of data and statistical analysis.

#### Annika Richter

Acquisition of data.

#### Stefan Küffer

Acquisition of data.

#### Kirsten Reuter-Jessen

Acquisition of data.

#### Hanibal Bohnenberger

Acquisition of data, analysis and interpretation of data and statistical analysis.

#### Stella Pauls

Acquisition of LC-MS/MS data of FFPE tissue material and statistical analysis of proteome data.

#### Catena Kresbach

Acquisition of data and statistical analysis.

#### Ulrich Schüller

Analysis and interpretation of data and administrative, technical, or material support.

Kai Stühler

Analysis and interpretation of data and administrative, technical, or material support.

Philipp Ströbel

Analysis and interpretation of data and critical revision of the manuscript.

Peter Albers

Critical revision of the manuscript and administrative, technical, or material support.

Daniel Nettersheim

Conception and design, acquisition of data, analysis and interpretation of data, visualization, drafting of the manuscript, statistical analysis, obtaining funding, administrative, technical, or material support and supervision.

10 Manuscript 6:

Assessing the risk to develop a growing teratoma syndrome based on molecular and epigenetic subtyping as well as novel secreted biomarkers





## Original Articles

## Assessing the risk to develop a growing teratoma syndrome based on molecular and epigenetic subtyping as well as novel secreted biomarkers

Pailin Pongratanakul<sup>a,g,1</sup>, Felix Bremmer<sup>b,1</sup>, Stella Pauls<sup>c</sup>, Gereon Poschmann<sup>c</sup>, Catena Kresbach<sup>d</sup>, Fatma Parmaksiz<sup>a</sup>, Margaretha A. Skowron<sup>a</sup>, Janina Fuß<sup>e</sup>, Alexa Stephan<sup>a</sup>, Pia Paffenholz<sup>f</sup>, Kai Stühler<sup>c</sup>, Ulrich Schüller<sup>d</sup>, Philipp Ströbel<sup>b</sup>, Axel Heidenreich<sup>f</sup>, Yue Che<sup>g</sup>, Peter Albers<sup>g</sup>, Daniel Nettersheim<sup>a,h,\*</sup>

<sup>a</sup> Department of Urology, Urological Research Laboratory, Translational UroOncology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

<sup>b</sup> Institute of Pathology, University Medical Center Goettingen, Goettingen, Germany

<sup>c</sup> Molecular Proteomics Laboratory (MPL), Biological and Medical Research Center (BMFZ), Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

<sup>d</sup> Institute of Neuropathology, University Hospital Hamburg-Eppendorf, Hamburg, Germany

<sup>e</sup> Competence Centre for Genomic Analysis, Kiel, Germany

<sup>f</sup> Department of Urology, University Hospital Cologne, University of Cologne, Cologne, Germany

<sup>g</sup> Department of Urology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

<sup>h</sup> Lighthouse Project Germ Cell Tumors, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf (CIO ABCD), Germany



## ARTICLE INFO

## Keywords:

Germ cell tumors  
Growing teratoma syndrome  
Teratoma  
DNA methylation  
Secretome  
Proteome  
Biomarker

## ABSTRACT

In germ cell tumors (GCT), a growing teratoma during chemotherapy with decreasing tumor markers was defined as 'growing teratoma syndrome' (GTS) by Logothetis et al. in 1982. So far, its pathogenesis and specific treatment options remain elusive.

We aimed at updating the GTS definition based on molecular and epigenetic features as well as identifying circulating biomarkers. We selected 50 GTS patients for clinical characterization and subsequently 12 samples were molecularly analyzed. We further included 7 longitudinal samples of 2 GTS patients. Teratomas (TER) showing no features of GTS served as controls.

GTS were stratified based on growth rates into a slow (<0.5 cm/month), medium (0.5–1.5) and rapid (>1.5) group. By analyzing DNA methylation, microRNA expression and the secretome, we identified putative epigenetic and secreted biomarkers for the GTS subgroups. We found that proteins enriched in the GTS groups compared to TER were involved in proliferation, DNA replication and the cell cycle, while proteins interacting with the immune system were depleted. Additionally, GTS<sup>rapid</sup> seem to interact more strongly with the surrounding microenvironment than GTS<sup>slow</sup>. Expression of pluripotency- and yolk-sac tumor-associated genes in GTS and formation of a yolk-sac tumor or somatic-type malignancy in the longitudinal GTS samples, pointed at an additional occult non-seminomatous component after chemotherapy. Thus, updating the Logothetis GTS definition is necessary, which we propose as follows:

**The GTS describes a continuously growing teratoma that might harbor occult non-seminomatous components considerably reduced during therapy but outgrowing over time again.**

## 1. Introduction

Testicular germ cell tumors (GCT) are the most common malignancy

among young men, appearing in different or mixed histological entities classified into seminomas and non-seminomas [1]. The latter has its own stem-cell-like population, the embryonal carcinoma (EC), which is able

\* Corresponding author. Department of Urology, Urological Research Laboratory, Translational UroOncology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University, Moorenstraße 5, 40225, Düsseldorf, Germany.

E-mail address: [Daniel.Nettersheim@med.uni-duesseldorf.de](mailto:Daniel.Nettersheim@med.uni-duesseldorf.de) (D. Nettersheim).

<sup>1</sup> Contributed equally.

<https://doi.org/10.1016/j.canlet.2024.216673>

Received 30 October 2023; Received in revised form 17 January 2024; Accepted 21 January 2024

Available online 29 January 2024

0304-3835/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

to differentiate into all three germ layers (teratoma (TER)), and into extra-embryonic tissue, i. e. yolk-sac tumors (YST) and choriocarcinomas (CC) [1].

During cisplatin-based standard chemotherapy, some patients present with a growing tumor mass on imaging, while serum tumor markers (alpha-fetoprotein (AFP), beta-human chorionic gonadotropin (beta-hCG), lactate dehydrogenase (LDH)) decreased or normalized. In these cases, complete surgical resection represents the only treatment option, revealing pure mature teratoma without evidence of other GCT entities in the final pathology. This phenomenon was first described by Logothetis et al. in 1982 based on six case reports and is called the 'growing teratoma syndrome' (GTS) [2]. To date, only few studies analyzing the GTS have been published [3–6]. Due to the small number of available cases worldwide, not much is known about GTS and its pathogenesis. Especially in cases with very space-demanding and surgical uncontrollable tumor mass, specific therapies and biomarkers early indicating presence of GTS are still lacking.

We aimed at updating and extending the current understanding of GTS, which was established 42 years ago. This is the first study subtyping GTS based on the growth rate over time and characterizing these subgroups on epigenetic (DNA methylation and microRNA), transcriptional (mRNA) and proteome/secretome level. By this, we not only identified the molecular and epigenetic features of the GTS subtypes, but also identified risk factors for rapidly growing GTS. Additionally, we deduced novel circulating biomarkers for the GTS subtypes.

## 2. Material and methods

### 2.1. GTS patient cohort

We retrospectively reviewed data of GCT patients undergoing a post-chemotherapy retroperitoneal lymph node dissection (RPLND) at the Departments of Urology of the University Hospitals Düsseldorf (UKD) and Cologne (UKK) from 2010 to 2023. Based on the definition by Logothetis et al., we identified 39 (UKD) and 11 (UKK) patients with a growing, histologically pure TER during or after chemotherapy associated with a decrease or normalization of serum tumor markers (AFP, beta-hCG, LDH). For the calculation of the tumor growth rate (cm/month), the transversal tumor diameter before and after chemotherapy on computed tomography (CT) were measured. The increase of tumor size (cm) was then divided by the time (month, 30 days = 1 month) from the start of chemotherapy to RPLND.

### 2.2. GCT tissues and cell lines

GTS tissues (formalin-fixed, paraffin-embedded (FFPE) and fresh frozen) were collected from the local biobank of the Department of Urology at the UKD (stored at the Institute of Pathology). All tissues were re-evaluated by a reference pathologist for type II GCT (F. B.). The utilized GCT cell lines were provided and cultivated as described in Table S1 A.

### 2.3. Immunohistochemistry

Immunohistochemistry has been performed as described [7]. Antigen retrieval was performed in citrate buffer. The Ki67 antibody (ready-to-use, Agilent Dako, Waldbronn, Germany) was incubated for 30 min at room temperature (RT). Samples were incubated with a ready-to-use-HRP-labelled secondary antibody at RT for 25 min. The substrate 'DAB + Chromogen System' (Agilent Technologies, Waldbronn, Germany) was used to visualize the antigen. Tissues were counterstained with Meyer's hematoxylin.

### 2.4. Nucleic acid isolation

According to manufacturer's recommendations, DNA was extracted

from 2×5 μm FFPE slides using the 'InnuPREP FFPE DNA Kit' on the 'InnuPure C16 System' (Jena Analytika, Jena, Germany). RNA was isolated by the TRIzol reagent according to the manual (Qiagen, Hilden, Germany). Nucleic acid concentrations and purities were measured by the 'Nanodrop 2000' photo-spectrometer (260/280 nm; 260/230 nm).

### 2.5. DNA methylation profiling

DNA methylation profiling was performed as described previously [8]. Briefly, 100–500 ng DNA were used for bisulfite conversion with the 'EZ DNA Methylation Kit' (Zymo Research, Freiburg, i. B., Germany). Afterwards, the 'DNA Clean & Concentrator-5' (Zymo Research) and the 'Infinium HD FFPE DNA Restore Kit' (Illumina, San Diego, CA, USA) were used to clean and restore the converted DNA. Finally, the 'Infinium 850k MethylationEPIC BeadChip' (850k array; Illumina) was used to evaluate the methylation status of 850,000 CpG sites on an 'iScan' device (Illumina).

### 2.6. Quantitative RT-PCR

cDNA synthesis and quantitative RT-PCR were performed as published previously [9]. 1 μg of total RNA was used for cDNA synthesis. Gene expression levels were determined on the 'C1000 cyclor' (BioRad, Feldkirchen, Germany) using 7.34 ng cDNA and in technical triplicates. *GAPDH* and *ACTB* were used as housekeeping genes and for data normalization. For oligonucleotide details, see Table S1 B.

### 2.7. microRNA sequencing

Library preparation was done with the 'NEXTFLEX Small RNA-Seq Kit v4' (Revvity, Hamburg, Germany) according to manufacturer's protocol with 200 ng input. MicroRNA sequencing (microRNAseq) of the library pool was done on a full 'NovaSeq6000 SP' flow cell with a 'NovaSeq 6000 SP Reagent Kit v1.5 (100 cycles)'. Paired-end sequencing has been performed twice for each sample. Demultiplexing was done with bcl2fastq v2.20.0.422. Analysis was performed using 'nf-core/smrnaseq v2.2.0', against the human reference genome GRCh38. The statistical QC was performed using the 'FastQC' tool (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), generating the mean quality scores of all sequences, sequence duplication levels and the count of unique/duplicate reads. Counts per million were calculated by the TMM normalization method in the 'edge' R package [10,11]. The variance between samples was assessed using the f-test and the significance (p-value <0.05) was checked by two-tailed Student's t-tests.

### 2.8. Proteome and secretome analysis

#### 2.8.1. Preparation of FFPE tissues for analysis

A modified FFPE tissue lysis protocol of Ikeda et al. was applied [12]. FFPE tissue slides were transferred with a scalpel into 1 mL protein-low binding tubes and deparaffinized by shaking in 500 μL xylene for 5 min, followed by removal of the solvent and air-drying. Next, tissues were resuspended in 200 μL lysis buffer (300 mM TRIS/HCl, 2 % SDS, pH 8.0), shock-frozen in liquid nitrogen and immediately heated for 25 min at 99 °C. For complete lysis, samples were ultrasonicated on ice for 20 min with 30 s on/off cycles and then shook or 2 h at 80 °C and 500 rpm, followed by a second ultrasonication step. After centrifugation, the supernatant was transferred into a new 0.5 mL protein-low binding tube and the pellet was resuspended in 100 μL buffer for a second extraction. The resulting supernatants were combined. Protein concentration was determined using the 'Pierce 660 nm Protein Assay' (Thermo Fisher Scientific, Bremen, Germany).

For liquid chromatography coupled to mass spectrometry (LC-MS) analysis, a modified magnetic bead-based sample preparation protocol was applied [13]. Briefly, 20 μg of total protein per sample was reduced by 10 μL 300 mM dithiothreitol (DTT) and shaking for 20 min at 56 °C

and 1000 rpm, followed by alkylation by 13  $\mu$ L 100 mM IAA and incubation for 15 min in the dark. A 20  $\mu$ g/ $\mu$ L bead stock of 1:1 'Sera-Mag SpeedBeads' was freshly prepared and 10  $\mu$ L were added to each sample. Afterwards, ethanol was added to a final concentration of 80 % for protein aggregation and the sample was incubated for 15 min at 20 °C. After three washing steps with 80 % ethanol and one washing step with 100 % ACN, beads were resuspended in 50 mM TEAB buffer and digested with final 0.4  $\mu$ g trypsin (1:50) at 37 °C and 1000 rpm overnight. Extra-digestion was carried out by adding trypsin (final 1:50) and shaking at 37 °C and 1000 rpm for 4 h. The supernatants were collected and 500 ng of each sample digest were diluted with 0.1 % TFA and subjected to LC-MS.

### 2.8.2. Secretome production and preparation for LC-MS analysis

0.5 $\times$ 0.2 cm slices of fresh frozen tissues were used for secretome production. For the removal of blood components, tumor slices were washed five times with 30 mL PBS. To avoid detecting secreted factors from a stress response, we carried out an initial incubation step in 400  $\mu$ L serum-free medium at 37 °C and 7.5 % CO<sub>2</sub> on a 24-well plate and discarded the medium after 2 h. Before incubation in new 400  $\mu$ L serum-free medium for 24 h, additional washing steps with 400  $\mu$ L PBS (5x) were performed. The secretomes were collected and centrifuged twice at 1000 g for 5 min and 3000 g for 10 min at 4 °C. According to manufacturer's recommendation (Roche, Mannheim, Germany), a protease inhibitor cocktail was prepared and 50  $\mu$ L were added to the secretome. For protein precipitation, trichloroacetic acid was added. Protein concentrations were determined by the 'Pierce 660 nm Protein Assay' (Thermo Fisher Scientific). Five  $\mu$ g of each supernatant was prepared by in-gel digestion as described previously [14]. Briefly, samples were loaded and separated on a polyacrylamide gel. After staining with Coomassie Brilliant blue, proteins were reduced with DTT, alkylated with iodoacetamide and digested with trypsin. Tryptic peptides (about 500 ng/sample) were prepared for LC-MS analysis in an aqueous solution of 0.1 % trifluoroacetic acid.

### 2.8.3. LC-MS analysis

For LC-MS, an 'Orbitrap Fusion Lumos Tribrid Mass Spectrometer with FAIMS' (Thermo Fisher Scientific) coupled to an 'Ultimate 3000 Rapid Separation' liquid chromatography system (Thermo Fisher Scientific, Idstein, Germany) equipped with an 'Acclaim PepMap 100C18 column' (75  $\mu$ m inner diameter, 25 cm length, 2 mm particle size from Thermo Fisher Scientific) as separation column and an 'Acclaim PepMap 100C18 column' (75  $\mu$ m inner diameter, 2 cm length, 3 mm particle size; Thermo Fisher Scientific) as trap column were used. A LC-gradient of 120 min separation duration was applied. The mass spectrometer was operated in positive data independent acquisition mode, the capillary temperature was set to 275 °C, the source voltage to 2.0 kV. It was additionally equipped with a FAIMS device (carrier gas flow: 4.5 L/min., compensation voltage (CV): 50 V). Precursor spectra were recorded in the orbitrap analyzer within a scan range of 380–985 m/z at a resolution of 60,000 (automatic gain control target value: 400,000, maximum injection time: 100 ms). Precursors were selected within isolation windows of 10 m/z (with an overlap of 1 m/z) within a precursor mass range of 380–980 m/z. After fragmentation by higher energy collisional dissociation (30 % collision energy, 5 % stepped collision energy), fragment spectra (scan range 145–1450 m/z) were recorded in the orbitrap analyzer at a resolution of 15,000 (automatic gain control target value: 100,000, maximum injection time: 40 ms). Cycle time was set to 3 s.

### 2.8.4. LC-MS raw data processing

Data analysis was carried out with 'DIA-NN' (version 1.8.1, <https://github.com/vdemichev/DiaNN>) [15]. All raw files were searched against the human proteome 'UniProt KB' dataset (UP000005640, downloaded on 12.01.2023) and the 'Maxquant Contaminant' database (downloaded on 03.05.2022), using the deep learning tool to generate

an *in silico* spectral library, which is implemented in 'DIA-NN'. The digestion enzyme was set to trypsin, the maximum number of missed cleavages was set to two (one for tissue secretome) and the peptide length was 7–30 amino acids. Mass accuracy was optimized by 'DIA-NN' using the first run in the experiment. As variable modifications were methionine oxidation, N-terminal methionine loss and methylation of lysines (only for FFPE tissues) defined. Fixed modification was carbamidomethylation of cysteines. All samples were analyzed in a match between run (MBR) search. During post processing, peptides were ungrouped and filtered to 1 % FDR on protein and peptide level and to all proteins identified with  $\geq$ 2 peptides. Contaminants were filtered out and the results were exported as excel sheet.

### 2.8.5. LC-MS data analysis of FFPE tissues

All samples were normalized by comparing each 'MaxLFQ' value of a protein to the corresponding value of a selected reference sample and the medians of the resulting log<sub>2</sub> fold changes (FC) were set to zero. The sample with the highest number of positive medians (of log<sub>2</sub> FC) was selected as reference. All other samples were normalized by multiplying the 'MaxLFQ' intensities with the calculated, delogarithmized medians of log<sub>2</sub> FC. All samples, except the samples of the longitudinal comparison undergo imputation of missing values from random numbers drawn from a defined width (0.3) and downshift (1.8) of the Gaussian distribution relative to the standard deviation of measured values. Protein abundance of different sample groups was compared by SAM analysis [16], carried out using R (version 4.2.3) with the 'sam()' function of the 'siggenes' package (version 1.72.0). Gene ontology clustering of samples that undergo longitudinal proteomic analysis were carried out using the 'Mfuzz' package (version 2.58.0) for R.

### 2.8.6. LC-MS data analysis of secretomes

Analysis of quantitative proteome data was carried out with 'Perseus 1.6.6.0' (Max-Planck Institute for Biochemistry, Planegg, Germany). Only proteins showing at least three valid values in at least one sample group were considered for further analysis (and statistical testing). Missing values of log<sub>2</sub> transformed normalized intensities (MaxLFQ) were filled in with values drawn from a downshifted normal distribution (width 0.3, downshift 1.8 standard deviations). Differences between groups were determined by ANOVA as well as 'significance analysis of microarrays' (SAM)<sup>4</sup> using an S<sup>0</sup> of 0.1 and permutation based false discovery rate set to 5 %. The secretion behavior or protein was predicted by 'OutCyte' [17].

### 2.8.7. Online analysis tools

Online analysis tools like 'STRING' (<https://string-db.org/>) and the 'DAVID Functional Annotation Tool' using 'GOTERM\_BP\_DIRECT', 'GOTERM\_MF\_DIRECT' and 'KEGG\_PATHWAY' (<https://david.ncifcrf.gov>) were used to predict protein interactions and their molecular functions. LC-MS data were evaluated by a principle component analysis using 'PCAGO' (<https://pcago.bioinf.uni-jena.de/>). The 'pandas', 'seaborn', and 'matplotlib' libraries were used in 'Python' for generation of Pearson's correlation matrices and volcano plots. The 'The Cancer Genome Atlas' (TCGA) GCT cohort was analyzed using 'cBioPortal' (<https://www.cbioportal.org/>) [18,19].

## 3. Results

In this study, we identified 50 GTS patients (Table 1). 43 patients presented with metastatic disease at first diagnosis, while 7 patients presented metastases during surveillance with clinical stage II and III in 33 and 17 cases, respectively. Prognosis based on the International Germ Cell Cancer Collaborative Group (IGCCCG) classification was mostly favorable, with good, intermediate and poor prognosis in 34, 7 and 9 cases, respectively. Initially, serum tumor markers AFP and beta-hCG were elevated (>7  $\mu$ g/L) in 44 and 39 patients, respectively. Every patient received platin-based chemotherapy with a majority of 90 %

**Table 1**

Clinical parameters of GTS patients treated at the Dept. of Urology of the University Hospital Düsseldorf and University Hospital Cologne included in this study.

GTS patient and tumor characteristics				
Tumor growth classification	all	slow	medium	rapid
Number	50 (46*)	19	21	6
Median age at diagnosis	27 (16 - 47)	31 (21 - 44)	26 (16 - 47)	28 (22 - 34)
Clinical stage				
II	33 (66 %)	15 (79 %)	14 (67 %)	4 (67 %)
III	17 (34 %)	4 (21 %)	7 (33 %)	2 (33 %)
IGCCCG classification				
good risk	34 (68 %)	16 (84 %)	14 (67 %)	4 (67 %)
intermediate risk	7 (14 %)	0 (0 %)	6 (28 %)	0 (0 %)
poor risk	9 (18 %)	3 (16 %)	1 (5 %)	2 (33 %)
Tumor marker initial*				
AFP >7.0 µg/l	44 (88 %)	18 (95 %)	17 (81 %)	6 (100 %)
β-HCG >2.0 mU/ml	39 (78 %)	14 (74 %)	17 (81 %)	4 (67 %)
both elevated	36 (72 %)	14 (74 %)	15 (71 %)	4 (67 %)
Tumor marker after chemotherapy				
negative	42 (84 %)	17 (89 %)	17 (81 %)	4 (67 %)
decreased	8 (16 %)	2 (11 %)	4 (19 %)	2 (33 %)
Initial clinical stage				
primary metastatic	43 (86 %)	16 (84 %)	18 (86 %)	5 (83 %)
metastatic recurrence	7 (14 %)	3 (16 %)	3 (14 %)	1 (17 %)
Tumor localization				
retroperitoneum	50 (100 %)	19 (100 %)	21 (100 %)	6 (100 %)
pulmonary	3 (6 %)	1 (5 %)	2 (10 %)	0 (0 %)
retrocrurol	7 (14 %)	2 (11 %)	3 (14 %)	1 (17 %)
liver	3 (6 %)	1 (5 %)	1 (5 %)	0 (0 %)
clavicular	2 (4 %)	1 (5 %)	1 (5 %)	0 (0 %)
Primary tumor histology				
pure teratoma	5 (10 %)	2 (11 %)	3 (14 %)	0 (0 %)
mix with teratoma	34 (68 %)	10 (53 %)	16 (76 %)	5 (83 %)
mix with embryonal carcinoma	39 (78 %)	14 (74 %)	16 (76 %)	5 (83 %)
mix with yolk-sac tumor	20 (40 %)	3 (16 %)	11 (52 %)	4 (67 %)
mix with choriocarcinoma	9 (18 %)	4 (21 %)	3 (14 %)	0 (0 %)
Chemotherapy				
3-4 x PEB	45 (90 %)	17 (89 %)	19 (90 %)	6 (100 %)
4 x PE/PEI	4 (8 %)	2 (11 %)	1 (5 %)	0
3 x TIP	1 (2 %)	0 (0 %)	1 (5 %)	0
* Median tumor marker levels [interquartile range]		AFP (µg/l)	β-HCG (mU/ml)	
all		181.5 [15.8-221129]	164 [2.7-31844]	
slow		147.5 [15.8-22129]	90 [6-11936]	
medium		153 [18-9937]	139.5 [6.2-5827]	
rapid		270 [27-40282]	6322 [66.5-31844]	

receiving 3-4 cycles of the regime cisplatin, etoposide and bleomycin (BEP). After chemotherapy, serum tumor markers were normalized in 42 and decreased to low but still slightly elevated (7.9-43 µg/l) levels in 8 patients. In terms of tumor manifestation, GTS formation was always found retroperitoneal (n = 50). Other additional localizations were retrocrurol (n = 7), lung (n = 3), liver (n = 3), and clavicular or cervical lymph nodes (n = 2).

Measurements of the transversal tumor diameter of the retroperitoneal mass before and after chemotherapy on CT scans for the calculation of the tumor growth rate were available for 46 of 50 GTS patients. Based on the tumor growth rate, patients were arbitrarily stratified into a slow (<0.5 cm/month), medium (0.5-1.5) and rapid (>1.5) group (Fig. 1 A). The median tumor growth rate was 0.8 cm/month (0.06-3.6 cm/month) (Fig. 1 B). Exemplary CT scans of the GTS subgroups before and after chemotherapy are shown in Fig. 1 C.

Next, we characterized each GTS subgroup based on molecular and epigenetic features to identify putative biomarkers on epigenetic (DNA methylation and microRNA) and protein level (Fig. 2 A). We included

three samples each of slow, medium and rapid GTS as well as three samples of TER without a growth trend as controls (Fig. 1 A, asterisks).

Histomorphologically, the GTS subgroups showed typical TER features with cells of all three germ layers, resulting in detection of tissue structures like cartilage, smooth muscle or endothelium (Fig. S1 A - C). We stained all GTS groups for Ki67 to identify proliferating cell populations that might fuel GTS growth (Fig. S1 D). We detected mainly Ki67<sup>+</sup> endothelial cells and single Ki67<sup>+</sup> cells within the tumor or at its borders, but without significant differences between the GTS subgroups (Fig. S1 D). Thus, the distribution of Ki67+ cells in the GTS subgroups does not explain the differences in growth rates or tumor volume.

By using 850k DNA methylation arrays, we analyzed the DNA methylation (5mC) landscape of TER and GTS samples (Data S1 A). Genome-wide, all analyzed samples showed a highly comparable distribution of DNA methylation (Fig. 2 B, inset). Regarding site-specific DNA methylation, we identified hypo- (<20 % 5mC) and hyper-methylated (>80 % 5mC) CpG dinucleotides for each GTS group and analyzed their distribution across the genome (in gene coding and CpG island context) (Fig. S2 A; Data S1 B). In all groups, hypermethylated CpG dinucleotides were mainly found within gene bodies and open sea context, while hypomethylated CpG dinucleotides were mainly found in transcription start sites (TSS; TSS200, TSS1500) and CpG island context (Fig. S2 A). By Venn diagrams, we compared all hypo- and hyper-methylated CpG dinucleotides found in the different groups to each other and identified commonly shared and individual CpG dinucleotides (Fig. S2 B). To identify putative epigenetic biomarkers, we identified CpG dinucleotides exclusively hypermethylated (>10-fold change in 5mC) in the different subgroups (GTS<sup>slow</sup> vs. TER, GTS<sup>rapid</sup> vs. TER, GTS<sup>rapid</sup> vs. GTS<sup>slow</sup>, GTS<sup>slow/rapid</sup> vs. TER) (Fig. 2 B, Data S1 B).

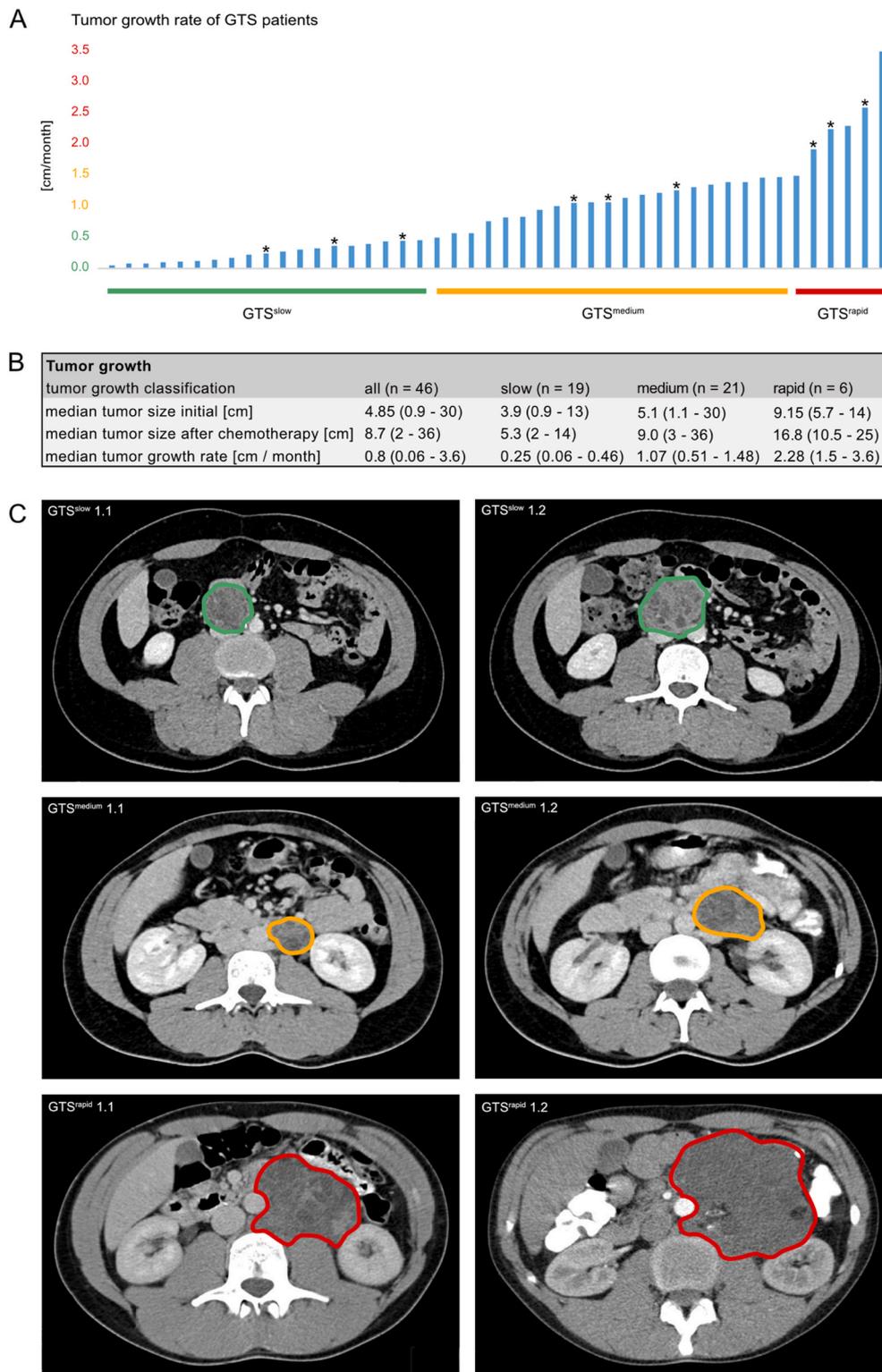
To identify further epigenetic biomarkers, we performed microRNAseq and demonstrated that the global microRNA expression profile of GTS<sup>slow</sup> and GTS<sup>rapid</sup> was more similar to each other than to TER (Fig. 2 C, inset). We identified microRNAs able to distinguish GTS<sup>slow</sup> from TER [13], GTS<sup>rapid</sup> from TER [8], and GTS<sup>rapid</sup> from GTS<sup>slow</sup> [12] (FC >2; p-value <0.05) (Fig. 2 C; Data S1 C).

Additionally, by LC-MS, we analyzed the secretomes of GTS and TER samples to identify secreted biomarkers (Data S1 D). A heatmap including unsupervised hierarchical clustering demonstrated that on secretome level, the GTS<sup>slow</sup> were more similar to TER than to GTS<sup>rapid</sup> (Fig. 2 D, inset). We screened for putative secreted biomarkers (signaling peptides and unconventional protein secretion (UPS)) indicative for the different GTS subgroups versus TER (with a FC >2) (Data S1 E). We identified 19 proteins specifically detected in GTS<sup>slow</sup> vs. TER, 53 in GTS<sup>rapid</sup> vs. TER, and 10 proteins in GTS<sup>slow/rapid</sup> vs. TER, which might serve as individual biomarkers (Fig. 2 C). Additionally, we identified putative biomarkers able to discriminate GTS<sup>rapid</sup> from GTS<sup>slow</sup> (Fig. 2 D, inset).

Of note, biomarkers specifically identifying TER are lacking. Thus, to identify general biomarkers for TER, we detected all secreted proteins (secreted peptides [31] and UPS (50)) with high intensity in all analyzed samples (TER+GTS; threshold 10×10<sup>6</sup>) (Fig. S2 B). To confirm our findings, we screened the TCGA GCT cohort for the mRNA expression of the putative biomarkers, demonstrating that indeed most of these factors were highly expressed in TER and mixed GCT with TER component (Fig. S3, asterisks).

So far, our study highlighted putative biomarkers on DNA methylation, microRNA and secretome level.

Now, we further extended our molecular characterization of the GTS by performing LC-MS of FFPE tissue slides (Data S1 F). Proteomes of GTS and TER were compared by a principle component analysis (PCA). Here, all GTS samples clustered apart from TER (with exception of 1 sample) (Fig. 3 A). A Pearson's correlation matrix (PCM) showed a decreasing similarity of GTS to TER with increasing growth speed, i. e. GTS<sup>rapid</sup> shows the least similarity to TER within the GTS groups (Fig. 3 B). To identify unique molecular features of GTS, we focused the analysis on the two subgroups GTS<sup>slow</sup> and GTS<sup>rapid</sup> as extremes within the spectrum



**Fig. 1.** A, B) Subtyping of GTS (slow, medium, rapid) based on the speed of growth (cm/month). C) Examples of slow (green), medium (orange) and rapid (red) GTS growth in CT scans before and after chemotherapy.

of GTS samples (Data S1 G). We compared their proteomes by using volcano plots, where 144 and 131 proteins were enriched compared to TER, respectively (Fig. 3 C). In contrast, 133 and 223 proteins were depleted compared to TER (Fig. 3 C). By using the STRING and DAVID algorithms, we screened for putative protein interactions and functional clustering (Fig. 3 D, E). Proteins found enriched in all GTS are involved in processes like DNA replication and unwinding, DNA repair, ATP-

related processes and the cell cycle, while mainly proteins associated with an immune response and the immune system in general were depleted compared to TER (Fig. 3 D). Moreover, exclusively in GTS<sup>rapid</sup>, further immune-related processes like the complement and coagulation cascade, the adaptive and humoral immune response, neutrophil extracellular trap formation, and the inflammatory or defense response were depleted. Additionally, apoptosis-associated processes were



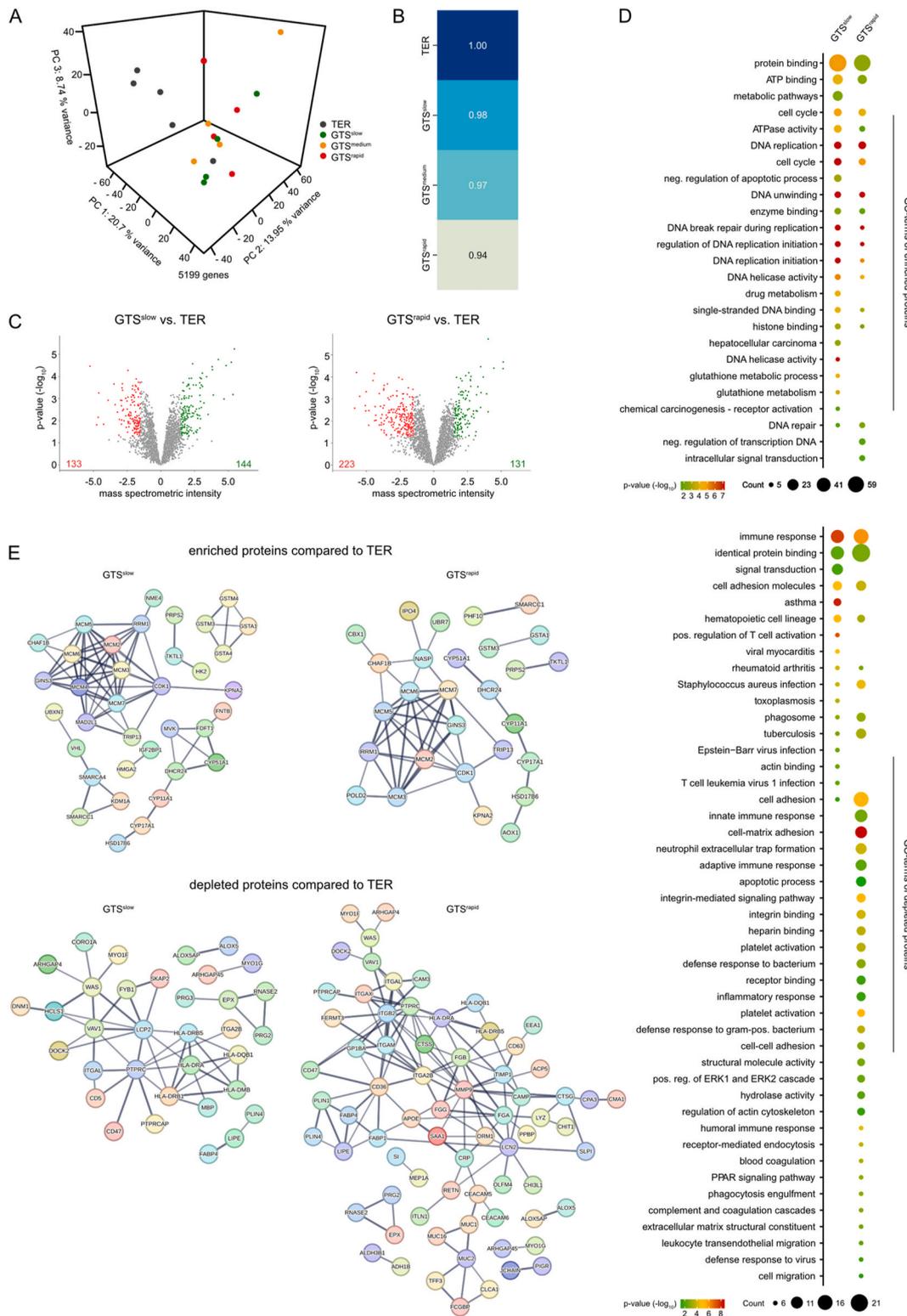
**Fig. 2.** A) Overview of the experiments performed in this study. B) Exclusively hypermethylated (>10-fold change in 5mC) CpG dinucleotides in the different GTS subgroups serving as possible biomarkers. Inlay: % distribution of global 5mC levels in GTS and TER samples. C) Differentially expressed microRNAs between the GTS groups and compared to TER (FC >2). Inlay: A heatmap including hierarchical clustering shows similarities between the microRNA expression profiles. D) Individually secreted biomarkers for the different GTS subgroups. A Venn diagram demonstrates differences and similarities between the secretomes of the different GTS subgroups (compared to TER). Inlay left side: A heatmap including hierarchical clustering shows similarities between the secretomes. Inlay middle: Biomarkers indicative for GTS<sup>rapid</sup> versus GTS<sup>slow</sup>.

depleted (Fig. 3 D). The STRING algorithm predicted interaction of enriched and depleted proteins in each GTS group compared to TER (Fig. 3 E).

Of note, in the sets of enriched or depleted proteins found in GTS<sup>medium</sup> samples, we detected similar corresponding functional terms

as found in both, the GTS<sup>slow</sup> and GTS<sup>rapid</sup> samples (Fig. S4 A–C).

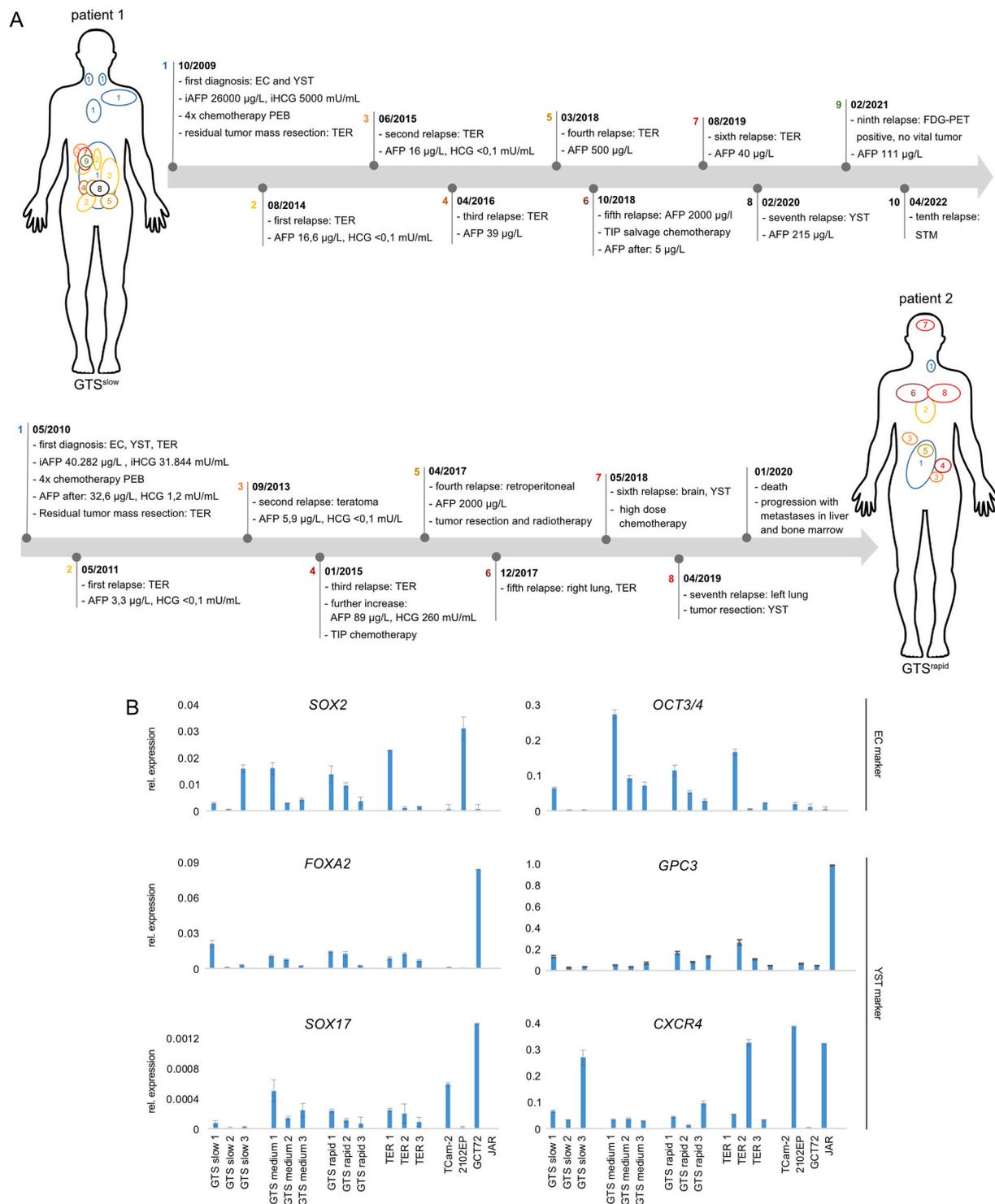
Finally, we asked how GTS develops in a patient over time. Therefore, we included longitudinal data of two GTS patients treated in our Department of Urology (UKD); with patient 1 and 2 representative for GTS<sup>slow</sup> and GTS<sup>rapid</sup> based on the initial growth dynamics of the GTS,



**Fig. 3.** A) A PCA of the proteome data from TER and the different GTS subgroups. B) A PCM compares the GTS subgroups to TER. C) Volcano plots illustrate proteins significantly enriched (green) or depleted (red) in GTS<sup>slow</sup> and GTS<sup>rapid</sup> compared to TER. D) Biological processes and functions of proteins found enriched in GTS compared to TER predicted by the DAVID algorithm. E) STRING-based protein-protein-interaction prediction of proteins found enriched/depleted in GTS<sup>slow</sup> and GTS<sup>rapid</sup> compared to TER.

respectively (Fig. 4 A). Both patients were diagnosed with a primary metastatic GTS. In patient 1, the initial tumor manifested mainly retroperitoneal, but was also found mediastinal and supraclavicular (Fig. 4 A). Over a time period of 13 years, patient 1 had developed 9 recurrences in the retroperitoneum only, starting with multiples TER,

followed by YST and then a somatic-type malignancy (STM) (Fig. 4 A). During the relapse period, AFP was always elevated between 16.6 µg/l and 500 µg/l. Patient 2 had 8 recurrences in 8 years (Fig. 4 A). After receiving several systemic therapies due to persistent high AFP level, the tumor metastasized into the lung and brain, showing YST for the first

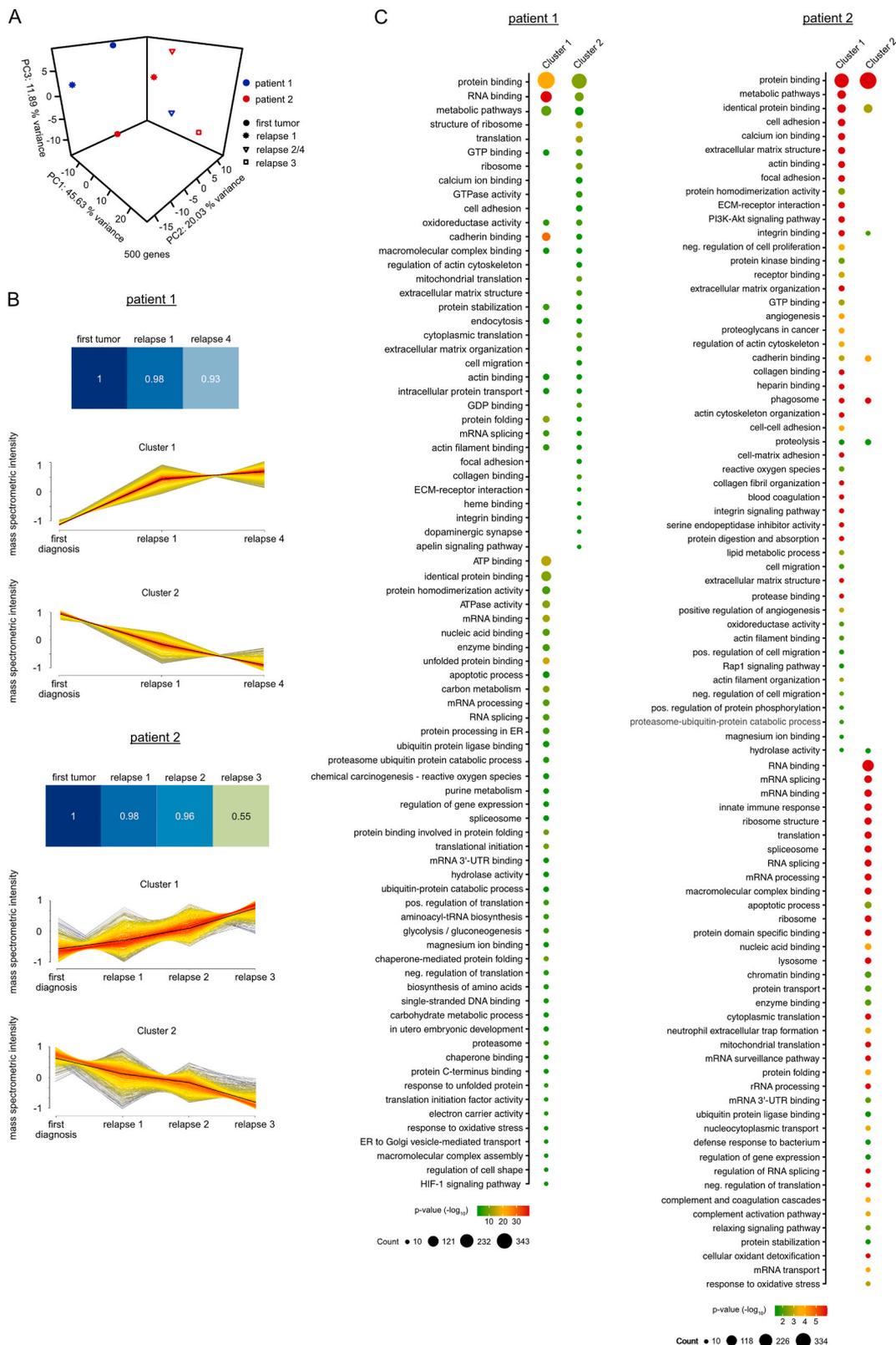


**Fig. 4.** A) Longitudinal clinical data of two GTS patients treated at the Department of Urology (UKD). Localization of each tumor/relapse is given in the pictogram of a human. B) qRT-PCR analysis of EC (*SOX2*, *OCT3/4*) and YST (*FOXA2*, *GPC3*, *SOX17*, *CXCR4*) marker genes in GTS and TER. GCT cell lines served as controls (TCam-2 = SEM; 2102EP = EC; GCT72 = YST; JAR = CC). *GAPDH* and *beta-ACTIN* served as housekeeping genes and were used for normalization.

time (Fig. 4 A). The tumor progressed rapidly with infiltration of the liver and bone marrow, subsequently, patient 2 died unfortunately. Taking both longitudinal cases into account, we observed teratomatous recurrences although AFP was elevated and steadily increasing over time, resulting in formation of YST (both patients) and even a STM (patient 1). These longitudinal data support the idea of a residual, but pathological not detectable (occult) subpopulation of EC or YST within the GTS. A hypothesis confirmed by us by qRT-PCR analysis of EC and

YST markers in GTS tissues (Fig. 4 B). Moderate but detectable expression of *SOX2* and *OCT3/4* (EC markers) as well as *FOXA2*, *GPC3*, *SOX17* and *CXCR4* (YST markers) suggests that most of these GTS samples also harbor an occult EC and/or YST component (Fig. 4 B) [20–22].

To gain insight into the underlying molecular mechanisms driving the GTS progression longitudinally, FFPE slides of the initial tumor and selected relapses were analyzed by LC-MS and compared by a PCA and PCM (Fig. 5 A, B). In both patients, the first recurrences were highly



**Fig. 5.** A) A PCA of the longitudinal tumor samples of patient 1 and 2. B) A PCM compares the proteome of the first tumor to the relapses for each patient. Clusters are showing the dynamics and kinetics of protein production during disease progression. C) The DAVID-based analysis of each cluster predicts underlying molecular processes.

similar to the first tumor (98 %) (Fig. 5 B). Over time, the similarity decreased with each following recurrence, showing a considerable change in patient 2 with only 55 % identity between the first tumor and its third relapse (Fig. 5 B). Further, we used a clustering software to analyze the dynamics of protein production over time. For both patients, we detected proteins enriched (cluster 1) or depleted (cluster 2) during disease progression (Fig. 5 B). Additional clusters representing the different dynamics are given in Data S1 H. We identified the underlying molecular processes associated with the proteins found in each cluster by the DAVID algorithm (Fig. 5 C). In patient 1, proteins linked to cluster 1 (steadily increasing in intensity) can be associated with RNA and protein regulation and processing (mRNA/nucleic acid/enzyme/protein binding, regulation of gene expression and translation), metabolic pathways, and embryonic development. Proteins linked to cluster 2 (steadily decreasing in intensity) can be associated with ECM interaction (e. g. ECM structure and organization, ECM-receptor interaction, cadherin/integrin/collagen binding), cell migration and adhesion (cell adhesion, focal adhesion). In patient 2, (in contrast to patient 1) proteins linked to cluster 1 can be associated with ECM interaction (e. g. ECM structure and organization, ECM-receptor interaction, cadherin/integrin/collagen binding and signaling), migration (regulation of migration, cell migration) and adhesion (cell adhesion, focal adhesion, cell-matrix adhesion). Proteins linked to cluster 2 can be associated with RNA and

protein regulation and processing (e. g. mRNA/enzyme/protein binding, regulation of gene expression and translation).

#### 4. Discussion

For metastatic GCT, GTS is a rare but serious condition having a poor prognosis. First described by Logothetis et al. [2], not much is known about GTS and its pathogenesis due to a small number of available cases so far. Treatment options are limited as GTS are resistant to chemo- and radiotherapy. Failure to detect GTS at early time points of manifestation leads to higher morbidity and mortality as GTS is growing quite fast and space demanding [3]. Thus, GTS patients harbor a higher risk to develop a STM [23,24]. This is the first study characterizing the clinical and molecular features of GTS on an (epi)genetic, transcriptional and proteome level. By this, we aimed at gaining insights into the underlying molecular pathogenesis of GTS and at identifying new biomarkers indicative for GTS.

Overall, patients diagnosed with GTS mainly present at primary metastatic stage with good prognosis and were treated with standard platin-based chemotherapy. However, GTS can also appear as metastatic recurrence or present with poor prognosis. Therefore, GTS seems to be unpredictable regarding clinical appearance, risk classification, received chemotherapy and tumor marker kinetics. Regarding the primary

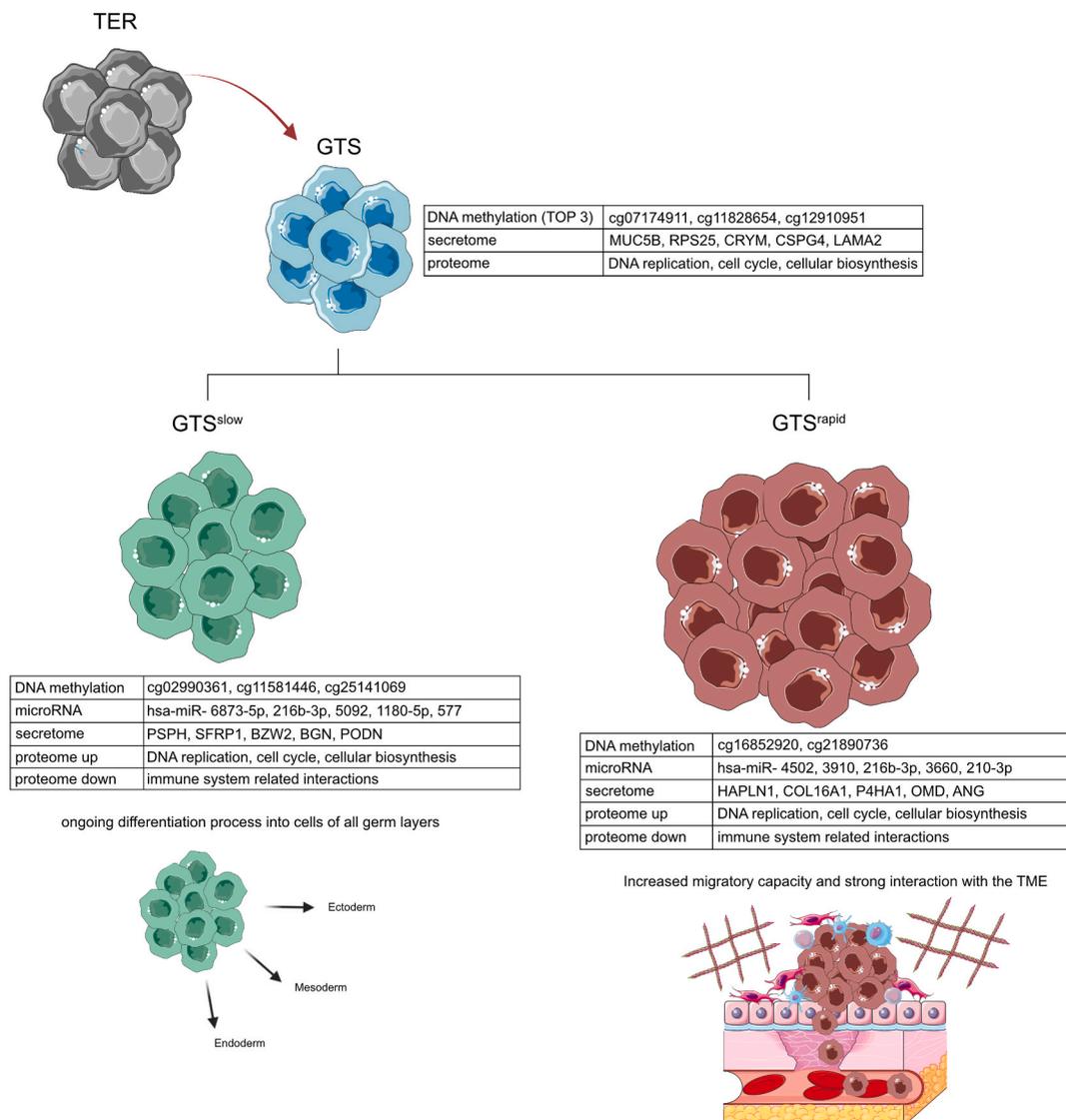


Fig. 6. Model summarizing molecular and epigenetic key findings of this study and highlighting most promising biomarkers.

orchietomy histology, no pattern was found to predict the presence or development of GTS, as all histologic specimen contain teratomatous components (cells of all three germ layers). However, analyzing the distribution of the different histological GCT entities within our tumor growth classification, we found a higher proportion of YST in GTS<sup>rapid</sup> (67 %) compared to GTS<sup>slow</sup> (16 %) (Table 1), indicating YST as possible predictive component for a faster tumor progression.

On an epigenetic perspective, the overall DNA methylation landscape between the different GTS subgroups and TER were highly similar. Nevertheless, we could highlight individual hypermethylated CpG dinucleotides exclusively found in the different GTS subtypes (GTS<sup>slow</sup> and GTS<sup>rapid</sup>) as putative epigenetic biomarkers, as found in other tumor types [25–27] (Fig. 6). Additionally, we offer a set of microRNAs that might stratify between GTS<sup>slow/rapid</sup> and TER as well as GTS<sup>slow</sup> and GTS<sup>rapid</sup> (Fig. 6).

We extended our pool of possible circulating biomarkers by a LC-MS-based analysis of the secretome of GTS and TER. By this, we have deduced putative biomarkers for the different GTS groups or TER classified as ‘secreted peptides’ or ‘unconventional protein secretion’ (UPS) (Fig. 6). Especially UPS factors are of high interest and offer new opportunities for biomarker identification, since novel mechanism of protein shedding or secretion has been postulated for these factors normally detectable in the cytoplasm, leading to secretion of small peptide sequences from the full protein [28,29]. These putative biomarkers on epigenetic and secretome level might allow to detect GTS formation (at an early stage), and therefore rendering it possible to adjust the therapeutic concept in time to prevent further outgrowth of this space demanding tumor. Taken together, our study offers a set of secreted biomolecules as putative biomarkers for GTS/TER, setting the stage for future biomarker screenings.

Of note, Nestler et al. recently identified biomarkers (AGR2, KRT19) able to distinguish viable teratoma elements from necrosis [30]. We detected AGR2 in the proteome of TER and GTS patients, while KRT19 could be detected in the secretome only (Fig. S4 D; Data S1 D, F). Nevertheless, overall LC-MS intensities were quite low compared to the other putative biomarkers identified in this study and measurements did not reach the significance threshold (ANOVA p-value/q-value).

Furthermore, we analyzed GTS tissues on proteome and secretome level by LC-MS to identify molecular features of GTS. Consistent with its clinical behavior and continuous tumor growth compared to TER without a growth trend, proteins found enriched in GTS were involved in DNA replication and regulation, cell cycle and cellular biosynthetic processes (Fig. 6). Additionally, proteins interacting with the immune system and mediating pro-apoptotic processes were significantly depleted compared to TER, suggesting that GTS utilize mechanisms to escape the immune system and apoptotic processes, enabling an unimpeded tumor growth (Fig. 6).

From our 50 GTS patients, 8 suffered from teratoma recurrence, of which 3 developed a STM and 1 patient a considerable AFP elevation. In both longitudinal cases, we observed teratomatous recurrences and steadily increasing AFP levels over the time, resulting in formation of YST (both patients) and eventually even a STM (patient 1). Additionally, the majority of analyzed patients had a mixed GCT history at first diagnosis with increased serum markers AFP and beta-hCG (Table 1). Further, we detected expression of pluripotency and YST factors within the tested GTS samples, indicative of EC and/or YST subpopulations. These data support the idea of a residual, but pathological not detectable (occult) subpopulation of EC, YST or STM within the GTS. Although these occult elements might not be detectable in every patient, the definition by Logothetis et al. needs to be updated as it defines GTS as a pure and mature teratoma. We suggest the following definition:

**The GTS describes a continuously growing teratoma that might harbor occult non-seminomatous components considerably reduced during therapy but outgrowing over time again.**

Additionally, in the future our identified putative biomarkers, once validated to be suitable for detecting GTS in routine diagnostics, might

be combined with detection of *microRNA371* or AFP to early detect re-growth of occult EC or YST elements, respectively [31–34].

We observed a considerably decreasing similarity in the proteome with each longitudinal recurrence in the GTS<sup>rapid</sup> patient [2], showing a similarity of only 55 % to the first tumor after four years. In contrast, the recurrence of the GTS<sup>slow</sup> patient [1] still showed a similarity of 93 % after nine years. Regarding the metastatic localization, GTS<sup>rapid</sup> metastasized also into lung and brain, while GTS<sup>slow</sup> relapsed only inside the retroperitoneum. These data suggest that patients of GTS<sup>rapid</sup> subgroup do not only suffer from a faster tumor growth, but also from a more aggressive disease progression with the ability to migrate more easily into other parts of the body. The characterization of the molecular processes associated with the changes in the proteome in both longitudinal patients suggests that in the GTS<sup>slow</sup> patient, the tumor is considerably altering its gene expression profile and that a differentiation process into the three germ layers is still ongoing, while a diminished interaction with the tumor microenvironment (TME) and migratory capacity could be observed over time, which is in line with an ongoing differentiation process and a limited tendency to metastasize. In contrast, in the GTS<sup>rapid</sup> patient, a strongly enhanced interaction with TME including migration could be observed, putatively explaining its tendency to metastasize quickly.

In summary and translating our findings to the clinic, our study shed light on the poorly understood molecular and epigenetic features of GTS and we updated the definition of a GTS. We provide a repertoire of biomolecules that might serve as secreted biomarkers on molecular and epigenetic level for future pathological routine diagnostics. The GTS subgroup-specific biomarkers may offer valuable guidance to physicians in their decision-making, suggesting immediate surgical intervention in cases of GTS<sup>rapid</sup>, while recommending active surveillance for GTS<sup>slow</sup> as the more appropriate approach. By this, the quality of life for patients is improved by avoiding unnecessary treatments. Nevertheless, suitability of these biomarkers needs to be validated in future studies first. Additionally, our molecular analysis may help to identify GTS-specific therapeutic concepts, e. g. we identified cell cycle-associated and regulating proteins to be enriched in GTS samples, thus, testing cell cycle inhibitors might be a reasonable approach in future studies.

## Ethics approval and consent to participate

The ethics committees of the Heinrich Heine University Düsseldorf and the University of Cologne raised no concerns on utilizing the tissue samples and corresponding clinical data for research (votes 2020-1247 and 21-1108, respectively).

## Consent for publication

All authors are aware of this article and agreed on publication.

## Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files or can be requested from the corresponding author. 850k DNA methylation array and microRNA sequencing data have been uploaded to Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) (GSE24009, GSE251975). LC-MS data generated in this study can be accessed via ProteomeXchange (<http://www.proteomexchange.org>) (PXD043529).

## Funding

P. Pongratanakul was supported by the ‘Junior Clinician Scientist’ program of the ‘Forschungskommission’ of the Medical Faculty of the Heinrich Heine University, Düsseldorf. F. Bremmer was supported by the ‘Wilhelm Sander-Stiftung’ (2016.041.1/.2/.3). MicroRNAseq analysis were performed under support by the ‘DFG Research

Infrastructure NGS\_CC' (project 407495230) as part of the 'Next Generation Sequencing Competence Network' (project 423957469).

### CRediT authorship contribution statement

**Pailin Pongratanakul:** Writing – review & editing, Writing – original draft, Visualization, Validation, Investigation, Funding acquisition, Formal analysis. **Felix Bremmer:** Resources, Investigation, Funding acquisition, Formal analysis, Conceptualization, Writing – review & editing. **Stella Pauls:** Software, Investigation, Formal analysis, Data curation. **Gereon Poschmann:** Software, Methodology, Investigation, Formal analysis, Data curation. **Catena Kresbach:** Investigation, Formal analysis. **Fatma Parmaksiz:** Investigation. **Margaretha A. Skowron:** Formal analysis, Visualization. **Janina Fuß:** Formal analysis, Investigation, Software. **Alexa Stephan:** Methodology, Investigation. **Pia Paffenholz:** Formal analysis. **Kai Stühler:** Resources, Methodology, Writing – review & editing. **Ulrich Schüller:** Resources, Methodology, Writing – review & editing. **Philipp Ströbel:** Resources, Writing – review & editing. **Axel Heidenreich:** Resources, Writing – review & editing. **Yue Che:** Writing – original draft, Investigation, Writing – review & editing. **Peter Albers:** Resources, Writing – review & editing. **Daniel Nettersheim:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Formal analysis, Data curation, Conceptualization, Investigation.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

We kindly thank Anna Pehlke and Olga Dschun for excellent technical assistance.

We kindly thank Prof. Tilman Rau and Mara Walther (Institute for Pathology, UKD) for their support and access to the local biobank.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.canlet.2024.216673>.

### References

1. L. Cheng, P. Albers, D.M. Berney, D.R. Feldman, G. Daugaard, T. Gilligan, et al., Testicular cancer, *Nat. Rev. Dis. Prim.* 4 (2018) 29.
2. C.J. Logothetis, M.L. Samuels, A. Trindade, D.E. Johnson, The growing teratoma syndrome, *Cancer* 50 (1982) 1629–1635.
3. P. Paffenholz, D. Pfister, V. Matveev, A. Heidenreich, Diagnosis and management of the growing teratoma syndrome: a single-center experience and review of the literature, *Urol. Oncol.: Seminars and Original Investigations* 36 (2018) 529. e23–529.e30.
4. D.J. Lee, H. Djaladat, N.N. Tadros, M. Movvassaghi, T. Tejura, V. Duddalwar, et al., Growing teratoma syndrome: clinical and radiographic characteristics, *Int. J. Urol.* 21 (2014) 905–908.
5. M. Stella, A. Gandini, P. Meeus, I. Aleksic, A. Flechon, C. Cropet, et al., Retroperitoneal vascular surgery for the treatment of giant growing teratoma syndrome, *Urology* 79 (2012) 365–370.
6. P.E. Spiess, W. Kassouf, G.A. Brown, A.M. Kamat, P. Liu, J.A. Gomez, et al., Surgical management of growing teratoma syndrome: the M. D. Anderson cancer center experience, *J. Urol.* 177 (2007) 1330–1334.
7. F. Bremmer, H. Bohnenberger, S. Küffer, T. Oellerich, H. Serve, H. Urlaub, et al., Proteomic comparison of malignant human germ cell tumor cell lines, *Dis. Markers* 2019 (2019) 1–14.
8. F. Bremmer, P. Pongratanakul, M. Skowron, Y. Che, A. Richter, S. Küffer, et al., Characterizing the mutational burden, DNA methylation landscape, and proteome of germ cell tumor-related somatic-type malignancies to identify the tissue-of-origin, mechanisms of therapy resistance, and druggable targets, *Br. J. Cancer* 129 (2023) 1580–1589.
9. M.A. Skowron, M. Vermeulen, A. Winkelhausen, T.K. Becker, F. Bremmer, P. Petzsch, et al., CDK4/6 inhibition presents as a therapeutic option for paediatric and adult germ cell tumours and induces cell cycle arrest and apoptosis via canonical and non-canonical mechanisms, *Br. J. Cancer* 123 (2020) 378–391.
10. M.D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data, *Genome Biol.* 11 (2010) R25.
11. M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics* 26 (2010) 139–140.
12. K. Ikeda, T. Monden, T. Kanoh, M. Tsujie, H. Izawa, A. Haba, et al., Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections, *J. Histochem. Cytochem.* 46 (1998) 397–403.
13. C.S. Hughes, S. Moggridge, T. Müller, P.H. Sorensen, G.B. Morin, J. Krijgsveld, Single-pot, solid-phase-enhanced sample preparation for proteomics experiments, *Nat. Protoc.* 14 (2019) 68–85.
14. K. Brenig, L. Grube, M. Schwarzländer, K. Köhrer, K. Stühler, G. Poschmann, The proteomic landscape of cysteine oxidation that underpins retinoic acid-induced neuronal differentiation, *J. Proteome Res.* 19 (2020) 1923–1940.
15. V. Demichev, C.B. Messner, S.I. Vernardis, K.S. Lilley, M. Ralsler, Dia-Nn, Neural networks and interference correction enable deep proteome coverage in high throughput, *Nat. Methods* 17 (2020) 41–44.
16. V.G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, *Proc. Natl. Acad. Sci. USA* 98 (2001) 5116–5121.
17. L. Zhao, G. Poschmann, D. Waldera-Lupa, N. Rafiee, M. Kollmann, K. Stühler, OutCyte: a novel tool for predicting unconventional protein secretion, *Sci. Rep.* 9 (2019) 19448.
18. Gao J, Arman Aksoy B, Dogrusoz U, Dresdner G, Gross B, Onur Sumer S, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. [cited 2023 Dec 5]; Available from: <http://www.adobe.com/products/illustrator.html>.
19. E. Cerami, J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, A. Aksoy, et al., The cBio Cancer Genomics Portal: an Open Platform for Exploring Multidimensional Cancer Genomics Data [Internet], *CANCER DISCOVERY*, 2012, 401. Available from: <http://aacrjournals.org/cancerdiscovery/article-pdf/2/5/401/1817590/401.pdf>.
20. J. de Jong, L.H.J. Looijenga, Stem cell marker OCT3/4 in tumor biology and germ cell tumor diagnostics: history and future, *Crit. Rev. Oncog.* 12 (2006) 171–203.
21. W. Wruck, F. Bremmer, M. Kotthoff, A. Fichtner, M.A. Skowron, S. Schönberger, et al., The pioneer and differentiation factor FOXA2 is a key driver of yolk-sac tumour formation and a new biomarker for paediatric and adult yolk-sac tumours, *J. Cell Mol. Med.* 25 (2021) 1394–1405.
22. J. de Jong, H. Stoop, A. Gillis, R. van Gorp, G.-J. van de Geijn, M de Boer, et al., Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications, *J. Pathol.* 215 (2008) 21–30.
23. C.V. Comiter, A.S. Kibel, J.P. Richie, M.R. Nucci, A.A. Renshaw, Prognostic features of teratomas with malignant transformation: a clinicopathological study of 21 cases, *J. Urol.* 159 (1998) 859–863.
24. R.J. Motzer, A. Amsterdam, V. Prieto, J. Sheinfeld, V.V.V.S. Murty, M. Mazumdar, et al., Teratoma with malignant transformation: diverse malignant histologies arising in men with germ cell tumors, *J. Urol.* 159 (1998) 133–138.
25. A. Schröck, A. Leisse, L. de Vos, H. Gevensleben, F. Dröge, A. Franzen, et al., Free-circulating methylated DNA in blood for diagnosis, staging, prognosis, and monitoring of head and neck squamous cell carcinoma patients: an observational prospective cohort study, *Clin. Chem.* 63 (2017) 1288–1296.
26. D. Dietrich, O. Hasinger, L.L. Bañez, L. Sun, G.J. van Leenders, T.M. Wheeler, et al., Development and clinical validation of a real-time PCR Assay for PITX2 DNA methylation to predict prostate-specific antigen recurrence in prostate cancer patients following radical prostatectomy, *J. Mol. Diagn.* 15 (2013) 270–279.
27. B. Schmidt, V. Liebenberg, D. Dietrich, T. Schlegel, C. Kneip, A. Seegebarth, et al., SHOX2 DNA Methylation is a Biomarker for the diagnosis of lung cancer based on bronchial aspirates, *BMC Cancer* 10 (2010) 600.
28. G. Poschmann, J. Bahr, J. Schrader, I. Stejerean-Todoran, I. Bogeski, K. Stühler, Secretomics—a key to a comprehensive picture of unconventional protein secretion, *Front. Cell Dev. Biol.* 10 (2022).
29. L. Zhao, G. Poschmann, D. Waldera-Lupa, N. Rafiee, M. Kollmann, K. Stühler, OutCyte: a novel tool for predicting unconventional protein secretion, *Sci. Rep.* 9 (2019) 19448.
30. T. Nestler, L. Kremer, M. von Brandenstein, M. Wittersheim, P. Paffenholz, S. Wagener-Rydzek, et al., Differentially expressed messenger RNA/proteins can distinguish teratoma from necrosis in postchemotherapy retroperitoneal lymph node dissection tissue, *Cancer* 129 (2023) 634–642.
31. K.-P. Dieckmann, A. Radtke, L. Geczi, C. Matthies, P. Anheuser, U. Eckardt, et al., Serum levels of MicroRNA-371a-3p (M371 test) as a new biomarker of testicular germ cell tumors: results of a prospective multicentric study, *J. Clin. Oncol.* 37 (2019) 1412–1423.
32. J.T. Lafin, N. Singla, S.L. Woldu, Y. Lotan, C.M. Lewis, K. Majmudar, et al., Serum MicroRNA-371a-3p levels predict viable germ cell tumor in chemotherapy-naïve patients undergoing retroperitoneal lymph node dissection, *Eur. Urol.* 77 (2020) 290–292.
33. J. Piao, J.T. Lafin, C.G. Scarpini, M.M. Nuño, I. Syring, K.-P. Dieckmann, et al., A multi-institutional pooled analysis demonstrates that circulating miR-371a-3p alone is sufficient for testicular malignant germ cell tumor diagnosis, *Clin. Genitourin. Cancer* 19 (2021) 469–479.
34. A.J.M. Gillis, M.A. Rijlaarsdam, R. Eini, L.C.J. Dorssers, K. Biermann, M.J. Murray, et al., Targeted serum miRNA (TSmiR) test for diagnosis and follow-up of (testicular) germ cell cancer patients: a proof of principle, *Mol. Oncol.* 7 (2013) 1083–1092.

## Author contributions to Manuscript 6

### **Assessing the risk to develop a growing teratoma syndrome based on molecular and epigenetic subtyping as well as novel secreted biomarkers**

#### Pailin Pongratanakul

Writing – review & editing, Writing – original draft, Visualization, Validation, Investigation, Funding acquisition, Formal analysis.

#### Felix Bremmer

Resources, Investigation, Funding acquisition, Formal analysis, Conceptualization, Writing – review & editing.

#### Stella Pauls

Software, Investigation, LC-MS/MS analysis for FFPE tissues, Proteomic data analysis.

#### Gereon Poschmann

Software, Methodology, Investigation, Formal analysis, Data curation.

#### Catena Kresbach

Investigation, Formal analysis.

#### Fatma Parmaksiz

Investigation.

#### Margaretha A. Skowron

Formal analysis, Visualization.

#### Janina Fuß

Formal analysis, Investigation, Software.

#### Alexa Stephan

Methodology, Investigation.

#### Pia Pfaffenholz

Formal analysis

#### Kai Stühler

Resources, Methodology, Writing – review & editing.

#### Ulrich Schüller

Resources, Methodology, Writing – review & editing.

Philipp Ströbel

Resources, Writing – review & editing.

Axel Heidenreich

Resources, Writing – review & editing.

Yue Che

Writing – original draft, Investigation, Writing – review & editing.

Peter Albers

Resources, Writing – review & editing.

Daniel Nettersheim

Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Formal analysis, Data curation, Conceptualization, Investigation.

## 11 Concluding Remarks

The methods presented in this work, including the optimized HIAR protocol, can be used flexibly for FFPE tissues with variable or very small sample amounts. The high sensitivity and reproducibility of the optimized sample preparation enable multi-omic approaches, such as combined proteomics/glyco-proteomics or integrated genomics/transcriptomics/proteomics strategies. This allows for analyses that generate a high information content even from very small sample quantities. In particular, the optimized conditions for tissue biopsies from various very small tissue changes, also known as precursor lesions, enable highly sensitive proteomic analyses, thus further elucidating the biology behind tumor development. The ability to detect subtle changes at the molecular level could lead to earlier diagnosis and better treatment options for patients. In the field of glyco-proteomics, there remains the challenge of providing high protein quantities for enrichment. The optimized sample preparation method also allows for the collection of glyco-proteomic data, which can be extracted from the proteomic dataset without additional measurements. This opens up new avenues for improving clinical diagnostics and personalized medicine. Not only are sample quantities limited in the context of PDAC precursor lesions, but there are also some very rare tissue changes in testicular carcinomas with highly variable sizes/amounts. It has also been demonstrated that the high reproducibility of the optimized method leads to very meaningful proteomic data, despite the significant fluctuations associated with clinical samples. Overall, the optimized sample preparation method can be applied in diverse ways and enables deep proteomic analyses to be conducted on very small sample amounts from FFPE tissues. This versatility is crucial for advancing our understanding of complex diseases and developing targeted therapeutic strategies. Future studies could further explore the integration of these methods with advanced bioinformatics tools, enhancing data interpretation and enabling novel discoveries in biomarker research.



## 12 Danksagung

Zuallererst bedanke ich mich bei Prof. Dr. Kai Stühler, der mir die Möglichkeit gegeben hat an diesem interessanten Thema zu arbeiten und dadurch tiefe Einblicke in die Massenspektrometrie und damit verbundenen Krebsforschung mittels Proteomics zu erlangen. Danke für die Unterstützung und die Möglichkeit meine Erfahrungen und mein Wissen so sehr zu erweitern. Des weiteren bedanke ich mich bei Prof. Dr. Andreas Reichert, der als Leiter des BMFZ und als Mentor meiner Arbeit einen großen Teil dazu beigetragen hat mir die Möglichkeit zu geben an diesen spannenden, interdisziplinären Themen zu arbeiten.

Bei Prof. Dr. Irene Esposito, Friederike Opitz, Sandra Biskup und den Mitarbeitern des Instituts für Pathologie bedanke ich mich für die Bereitstellung der mikrodisektierten FFPE Proben und die Möglichkeit am Thema des Pankreaskarzinoms arbeiten zu dürfen. Bei Prof. Dr. Nettersheim und den Mitarbeitern des Instituts für translationale UrOnkologie bedanke ich mich recht herzlich für die Bereitstellung der FFPE Proben im Bereich der Hodenkarzinome und die Möglichkeit an diesen interessanten Publikationen mitzuwirken.

Ganz besonders bedanke ich mich bei Dr. Anja Stefanski, bei der ich die Möglichkeit hatte mir professionelles Wissen im Bereich der Massenspektrometrie und Flüssigkeits-chromatographie anzueignen. Ich bin dankbar für all die erfolgreichen als auch frustrierenden Troubleshoots der Geräte, ohne die ich nie so viel lernen konnte. Weiterhin danke ich dir für den mentalen Support und die vielen fachlichen als auch persönlichen Dinge, die ich bei dir lernen durfte. Ich danke außerdem Dr. Nina Prescher für die Einarbeitung in den Geräteraum und die vielen hilfreichen Tips bezüglich der Promotion. Weiterhin bedanke ich mich bei Dr. Gereon Poschmann, der mir bei der Datenauswertung vieler verschiedener Projekte zur Seite stand und durch den ich viele verschiedene Methoden zur Auswertung lernen durfte. Thomas Lenz danke ich zum einen für die Bereitstellung des CaproMag Protokolls, mit dessen Hilfe ich die Probenvorbereitungsmethoden meiner Arbeit verbessern konnte, als auch für die vielen Nachhilfestunden im Bereich des Programmierens mit R, verbunden mit vielen anregenden Gesprächen während Raucherpausen und Feierabendbieren. Bei Dr. Marc Daniel Driessen bedanke ich mich für die Expertise im Bereich der Glyco-Proteomics und die vielen interessanten, wissenschaftlichen Gespräche. Weiterhin danke ich allen Beteiligten des Instituts des Molecular Proteomics Laboratory, im Besonderen dem technischen Personal: Eva Bruns, Vanessa Hoffmann, Christin Hafermann und Marion Gerke für ihre tatkräftige Unterstützung im Labor und Geräteraum.

Des Weiteren möchte ich mich bei Jens König bedanken für die vielen, hilfreichen Coachings, die emotionalen Gespräche und die damit verbundene Weiterentwicklung, die einen großen Teil zu meiner persönlichen Entwicklung beigetragen haben.

An dieser Stelle möchte ich mich bei meinen Unikolleginnen für die gemeinsame Studienzeit bedanken und die vielen Dinge, die wir gemeinsam gemeistert haben. Danke Helena Lazic für die aufregende Zeit der Masterarbeit und danke Bianca Axinte für die gemeinsame Doktorzeit

und danke euch beiden für die gemeinsame Studienzeit. Danke für all die mentale Unterstützung, die vielen gemeinsamen, verzweifelten Momente beim Protokolle schreiben, die wunderschönen Treffen, Parties und dem lauten Lachen, das uns die ganze Zeit begleitet und Kraft gegeben hat. Ich danke dir, Marcellino Caliqi, dafür dass du das ganze Studium an mich geglaubt hast und mich immer gepusht und aufgerüttelt hast, wenn ich es gebraucht habe. Danke für die große persönliche Entwicklung, die ich dank dir gemacht habe.

Zusätzlich bedanke ich mich bei meinen zwei besten Freundinnen, Marine Kloecker und Yasemin Kocakaya. Danke Marine für deine mentale Unterstützung, all die wunderbaren Gespräche die zur persönlichen Entwicklung beigetragen haben und der gemeinsamen Erlangung unserer Abschlüsse, im Herzen verbunden, wenn auch örtlich getrennt. Danke Yasemin, dass du immer für mich da warst und mit deiner reinen Seele und deiner liebevollen Art mein Herz bereichert hast.

Mein ganz besonderer Dank gilt meiner gesamten Familie, insbesondere meinen Eltern, Isabella Arzt und Wilhelm Pauls für ihre mentale und finanzielle Unterstützung, ohne die ich das Studium und die Doktorzeit nicht überstanden hätte. Ich danke meinem Onkel Simon Kubanski für die ständige Unterstützung bei IT-Problemen, als auch dafür, dass ich beim Bewerbungsprozess so sehr auf deine Hilfe und deine Erfahrung vertrauen konnte. Meinem Patenonkel Dominik Pauls danke ich für all die Unterstützung, die er mir in schweren Zeiten geleistet hat.

Zu guter Letzt möchte ich mich im Besonderen bei meinem Bruder Benedikt Pauls bedanken. Danke, dass du immer für mich da warst, mein Leben so bereichert hast und mir in den schwersten Zeiten zur Seite standest. Ohne dich wäre ich nicht da, wo ich heute bin.

## 13 References

- (1) Patel, V.; Hood, B. L.; Molinolo, A. A.; Lee, N. H.; Conrads, T. P.; Braisted, J. C.; Krizman, D. B.; Veenstra, T. D.; Gutkind, J. S. Proteomic Analysis of Laser-Captured Paraffin-Embedded Tissues: A Molecular Portrait of Head and Neck Cancer Progression. *Clinical Cancer Research* **2008**, *14* (4), 1002-1014.
- (2) Johann, D. J.; Rodriguez-Canales, J.; Mukherjee, S.; Prieto, D. A.; Hanson, J. C.; Emmert-Buck, M.; Blonder, J. Approaching Solid Tumor Heterogeneity on a Cellular Basis by Tissue Proteomics Using Laser Capture Microdissection and Biological Mass Spectrometry. *Journal of Proteome Research* **2009**, *8* (5), 2310-2318.
- (3) Umar, A.; Kang, H.; Timmermans, A. M.; Look, M. P.; Meijer-van Gelder, M. E.; den Bakker, M. A.; Jaitly, N.; Martens, J. W. M.; Luider, T. M.; Foekens, J. A.; et al. Identification of a Putative Protein Profile Associated with Tamoxifen Therapy Resistance in Breast Cancer\*. *Molecular & Cellular Proteomics* **2009**, *8* (6), 1278-1294.
- (4) Wiśniewski, J. R.; Ostasiewicz, P.; Mann, M. High Recovery FASP Applied to the Proteomic Analysis of Microdissected Formalin Fixed Paraffin Embedded Cancer Tissues Retrieves Known Colon Cancer Markers. *Journal of Proteome Research* **2011**, *10* (7), 3040-3049.
- (5) Craven, R. A.; Cairns, D. A.; Zougman, A.; Harnden, P.; Selby, P. J.; Banks, R. E. Proteomic analysis of formalin-fixed paraffin-embedded renal tissue samples by label-free MS: Assessment of overall technical variability and the impact of block age. *PROTEOMICS – Clinical Applications* **2013**, *7* (3-4), 273-282.
- (6) Longuespée, R.; Fléron, M.; Pottier, C.; Quesada-Calvo, F.; Meuwis, M.-A.; Baiwir, D.; Smargiasso, N.; Mazzucchelli, G.; De Pauw-Gillet, M.-C.; Delvenne, P. Tissue proteomics for the next decade? Towards a molecular dimension in histology. *Omics: a journal of integrative biology* **2014**, *18* (9), 539-552.
- (7) Djuric, U.; Rodrigues, D. C.; Batruch, I.; Ellis, J.; Shannon, P.; Diamandis, P. Spatiotemporal proteomic profiling of human cerebral development. *Molecular & Cellular Proteomics* **2017**, *16* (9), 1548-1562.
- (8) Föll, M. C.; Fahrner, M.; Oria, V. O.; Kühs, M.; Biniossek, M. L.; Werner, M.; Bronsert, P.; Schilling, O. Reproducible proteomics sample preparation for single FFPE tissue slices using acid-labile surfactant and direct trypsinization. *Clinical Proteomics* **2018**, *15* (1).
- (9) Friedrich, C.; Schallenberg, S.; Kirchner, M.; Ziehm, M.; Niquet, S.; Haji, M.; Beier, C.; Neudecker, J.; Klauschen, F.; Mertins, P. Comprehensive micro-scaled proteome and phosphoproteome characterization of archived retrospective cancer repositories. *Nature Communications* **2021**, *12* (1), 3576.
- (10) Ye, Z.; Mao, Y.; Clausen, H.; Vakhrushev, S. Y. Glyco-DIA: a method for quantitative O-glycoproteomics with in silico-boosted glycopeptide libraries. *Nature Methods* **2019**, *16* (9), 902-910.
- (11) Fox, C. H.; Johnson, F. B.; Whiting, J.; Roller, P. P. Formaldehyde fixation. *Journal of Histochemistry & Cytochemistry* **1985**, *33* (8), 845-853.
- (12) *Pancreatic cancer statistics*. World Cancer Research Fund International, 2020. <https://www.wcrf.org/cancer-trends/pancreatic-cancer-statistics/> (accessed 11.01.2024).
- (13) Siegel, R. L.; Miller, K. D.; Wagle, N. S.; Jemal, A. Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians* **2023**, *73* (1), 17-48.
- (14) UK, P. C. Pancreatic Cancer UK, 2022. <https://www.pancreaticcancer.org.uk/information/just-diagnosed-with-pancreatic-cancer/pancreatic-ductal-adenocarcinoma-and-other-exocrine-tumours/> (accessed 12.01.2024).
- (15) Nakanuma, Y.; Sato, Y. Hilar cholangiocarcinoma is pathologically similar to pancreatic duct adenocarcinoma: suggestions of similar background and development. *Journal of Hepato-Biliary-Pancreatic Sciences* **2014**, *21* (7), 441-447.
- (16) Nakanuma, Y.; Kakuda, Y.; Sugino, T.; Sato, Y.; Fukumura, Y. Pathologies of Precursor Lesions of Biliary Tract Carcinoma. *Cancers* **2022**, *14* (21), 5358.
- (17) Nakanuma, Y. A novel approach to biliary tract pathology based on similarities to pancreatic counterparts: Is the biliary tract an incomplete pancreas? *Pathology International* **2010**, *60* (6), 419-429.
- (18) Distler, M.; Aust, D.; Weitz, J.; Pilarsky, C.; Grützmann, R. Precursor lesions for sporadic pancreatic cancer: PanIN, IPMN, and MCN. *Biomed Res Int* **2014**, *2014*, 474905.
- (19) Klöppel, G.; Bommer, G.; Rückert, K.; Seifert, G. Intraductal proliferation in the pancreas and its relationship to human and experimental carcinogenesis. *Virchows Archiv A* **1980**, *387* (2), 221-233.

- (20) Aichler, M.; Seiler, C.; Tost, M.; Siveke, J.; Mazur, P. K.; Da Silva-Buttkus, P.; Bartsch, D. K.; Langer, P.; Chiblak, S.; Durr, A.; et al. Origin of pancreatic ductal adenocarcinoma from atypical flat lesions: a comparative study in transgenic mice and human tissues. *J Pathol* **2012**, *226* (5), 723-734.
- (21) Sipos, B.; Frank, S.; Gress, T.; Hahn, S.; Klöppel, G. Pancreatic Intraepithelial Neoplasia Revisited and Updated. *Pancreatology* **2008**, *9* (1-2), 45-54.
- (22) Wang, L.; Xie, D.; Wei, D. Pancreatic acinar-to-ductal metaplasia and pancreatic cancer. *Pancreatic Cancer: Methods and Protocols* **2019**, 299-308.
- (23) Zamboni, G.; Hirabayashi, K.; Castelli, P.; Lennon, A. M. Precancerous lesions of the pancreas. *Best Practice & Research Clinical Gastroenterology* **2013**, *27* (2), 299-322.
- (24) Habbe, N.; Shi, G.; Meguid, R. A.; Fendrich, V.; Esni, F.; Chen, H.; Feldmann, G.; Stoffers, D. A.; Konieczny, S. F.; Leach, S. D.; et al. Spontaneous induction of murine pancreatic intraepithelial neoplasia (mPanIN) by acinar cell targeting of oncogenic Kras in adult mice. *Proceedings of the National Academy of Sciences* **2008**, *105* (48), 18913-18918.
- (25) Esposito, I.; Seiler, C.; Bergmann, F.; Kleeff, J.; Friess, H.; Schirmacher, P. Hypothetical Progression Model of Pancreatic Cancer With Origin in the Centroacinar-Acinar Compartment. *Pancreas* **2007**, *35* (3), 212-217.
- (26) Ban, S.; Naitoh, Y.; Mino-Kenudson, M.; Sakurai, T.; Kuroda, M.; Koyama, I.; Lauwers, G. Y.; Shimizu, M. Intraductal Papillary Mucinous Neoplasm (IPMN) of the Pancreas: Its Histopathologic Difference Between 2 Major Types. *The American Journal of Surgical Pathology* **2006**, *30* (12), 1561-1569.
- (27) Lüttges, J.; Zamboni, G.; Longnecker, D.; Klöppel, G. The Immunohistochemical Mucin Expression Pattern Distinguishes Different Types of Intraductal Papillary Mucinous Neoplasms of the Pancreas and Determines Their Relationship to Mucinous Noncystic Carcinoma and Ductal Adenocarcinoma. *The American Journal of Surgical Pathology* **2001**, *25* (7), 942-948.
- (28) Kim, G. E.; Bae, H. I.; Park, H. U.; Kuan, S. F.; Crawley, S. C.; Ho, J. J. L.; Kim, Y. S. Aberrant expression of MUC5AC and MUC6 gastric mucins and sialyl Tn antigen in intraepithelial neoplasms of the pancreas. *Gastroenterology* **2002**, *123* (4), 1052-1060.
- (29) Lüttges, J.; Reinecke-Lüthge, A.; Möllmann, B.; Menke, M. A. O. H.; Clemens, A.; Klimpfinger, M.; Sipos, B.; Klöppel, G. Duct changes and K-ras mutations in the disease-free pancreas: analysis of type, age relation and spatial distribution. *Virchows Archiv* **1999**, *435* (5), 461-468.
- (30) Lühr, M.; Klöppel, G.; Maisonneuve, P.; Lowenfels, A. B.; Lüttges, J. Frequency of K-ras Mutations in Pancreatic Intraductal Neoplasias Associated with Pancreatic Ductal Adenocarcinoma and Chronic Pancreatitis: A Meta-Analysis. *Neoplasia* **2005**, *7* (1), 17-23.
- (31) Kanda, M.; Matthaei, H.; Wu, J.; Hong, S. M.; Yu, J.; Borges, M.; Hruban, R. H.; Maitra, A.; Kinzler, K.; Vogelstein, B.; et al. Presence of Somatic Mutations in Most Early-Stage Pancreatic Intraepithelial Neoplasia. *Gastroenterology* **2012**, *142* (4), 730-733.e739.
- (32) Hruban, R. H.; Wilentz, R. E.; Kern, S. E. Genetic Progression in the Pancreatic Ducts. *The American Journal of Pathology* **2000**, *156* (6), 1821-1825.
- (33) Almoguera, C.; Shibata, D.; Forrester, K.; Martin, J.; Arnheim, N.; Perucho, M. Most human carcinomas of the exocrine pancreas contain mutant cK-ras genes. *Cell* **1988**, *53* (4), 549-554.
- (34) Mohri, D.; Asaoka, Y.; Ijichi, H.; Miyabayashi, K.; Kudo, Y.; Seto, M.; Ohta, M.; Tada, M.; Tanaka, Y.; Ikenoue, T.; et al. Different subtypes of intraductal papillary mucinous neoplasm in the pancreas have distinct pathways to pancreatic cancer progression. *Journal of Gastroenterology* **2012**, *47* (2), 203-213.
- (35) Kawabata, H.; Ono, Y.; Tamamura, N.; Oyama, K.; Ueda, J.; Sato, H.; Takahashi, K.; Taniue, K.; Okada, T.; Fujibayashi, S.; et al. Mutant GNAS limits tumor aggressiveness in established pancreatic cancer via antagonizing the KRAS-pathway. *Journal of Gastroenterology* **2022**, *57* (3), 208-220.
- (36) Dal Molin, M.; Matthaei, H.; Wu, J.; Blackford, A.; Debeljak, M.; Rezaee, N.; Wolfgang, C. L.; Butturini, G.; Salvia, R.; Bassi, C. Clinicopathological correlates of activating GNAS mutations in intraductal papillary mucinous neoplasm (IPMN) of the pancreas. *Annals of surgical oncology* **2013**, *20*, 3802-3808.
- (37) Day, J. D.; Diguseppe, J. A.; Yeo, C.; Lai-Goldman, M.; Anderson, S. M.; Goodman, S. N.; Kern, S. E.; Hruban, R. H. Immunohistochemical evaluation of HER-2/neu expression in pancreatic adenocarcinoma and pancreatic intraepithelial neoplasms. *Human Pathology* **1996**, *27* (2), 119-124.
- (38) Satoh, K.; Sasano, H.; Shimosegawa, T.; Koizumi, M.; Yamazaki, T.; Mochizuki, F.; Kobayashi, N.; Okano, T.; Toyota, T.; Sawai, T. An immunohistochemical study of the c-erbB-2

- oncogene product in intraductal mucin-hypersecreting neoplasms and in ductal cell carcinomas of the pancreas. *Cancer* **1993**, 72 (1), 51-56.
- (39) Wilentz, R. E.; Geradts, J.; Maynard, R.; Offerhaus, G. J. A.; Kang, M.; Goggins, M.; Yeo, C. J.; Kern, S. E.; Hruban, R. H. Inactivation of the p16 (INK4A) tumor-suppressor gene in pancreatic duct lesions: loss of intranuclear expression. *Cancer research* **1998**, 58 (20), 4740-4744.
- (40) Cicenas, J.; Kvederaviciute, K.; Meskinyte, I.; Meskinyte-Kausiliene, E.; Skeberdyte, A.; Cicenas, J. KRAS, TP53, CDKN2A, SMAD4, BRCA1, and BRCA2 Mutations in Pancreatic Cancer. *Cancers* **2017**, 9 (5), 42.
- (41) Wilentz, R. E.; Iacobuzio-Donahue, C. A.; Argani, P.; McCarthy, D. M.; Parsons, J. L.; Yeo, C. J.; Kern, S. E.; Hruban, R. H. Loss of Expression of Dpc4 in Pancreatic Intraepithelial Neoplasia: Evidence That DPC4 Inactivation Occurs Late in Neoplastic Progression<sup>1</sup>. *Cancer Research* **2000**, 60 (7), 2002-2006.
- (42) Hwang, R. F.; Gordon, E. M.; Anderson, W. F.; Parekh, D. Gene therapy for primary and metastatic pancreatic cancer with intraperitoneal retroviral vector bearing the wild-type p53 gene. *Surgery* **1998**, 124 (2), 143-151.
- (43) Brody, L. C.; Biesecker, B. B. Breast Cancer Susceptibility Genes:BRCA1andBRCA2. *Medicine* **1998**, 77 (3), 208-226.
- (44) Friedenson, B. BRCA1 and BRCA2 pathways and the risk of cancers other than breast or ovarian. *MedGenMed* **2005**, 7 (2), 60.
- (45) Huang, L.; Wu, C.; Yu, D.; Wang, C.; Che, X.; Miao, X.; Zhai, K.; Chang, J.; Jiang, G.; Yang, X.; et al. Identification of common variants in BRCA2 and MAP2K4 for susceptibility to sporadic pancreatic cancer. *Carcinogenesis* **2013**, 34 (5), 1001-1005.
- (46) Park, H.-U.; Kim, J.-W.; Kim, G. E.; Bae, H.-I.; Crawley, S. C.; Yang, S. C.; Gum, James R. Jr.; Batra, S. K.; Rousseau, K.; Swallow, D. M.; et al. Aberrant Expression of MUC3 and MUC4 Membrane-Associated Mucins and Sialyl Lex Antigen in Pancreatic Intraepithelial Neoplasia. *Pancreas* **2003**, 26 (3), e48-e54.
- (47) Wesseling, J.; van der Valk, S. W.; Vos, H. L.; Sonnenberg, A.; Hilkens, J. Episialin (MUC1) overexpression inhibits integrin-mediated cell adhesion to extracellular matrix components. *Journal of Cell Biology* **1995**, 129 (1), 255-265.
- (48) Yonezawa, S.; Sato, E. Expression of mucin antigens in human cancers and its relationship with malignancy potential. *Pathology International* **1997**, 47 (12), 813-830.
- (49) Kaur, S.; Kumar, S.; Momi, N.; Sasson, A. R.; Batra, S. K. Mucins in pancreatic cancer and its microenvironment. *Nature Reviews Gastroenterology & Hepatology* **2013**, 10 (10), 607-620.
- (50) Suh, H.; Pillai, K.; Morris, D. L. Mucins in pancreatic cancer: biological role, implications in carcinogenesis and applications in diagnosis and therapy. *Am J Cancer Res* **2017**, 7 (6), 1372-1383.
- (51) Singh, P. K.; Hollingsworth, M. A. Cell surface-associated mucins in signal transduction. *Trends in Cell Biology* **2006**, 16 (9), 467-476.
- (52) Singh, A. P.; Moniaux, N.; Chauhan, S. C.; Meza, J. L.; Batra, S. K. Inhibition of MUC4 Expression Suppresses Pancreatic Tumor Cell Growth and Metastasis. *Cancer Research* **2004**, 64 (2), 622-630.
- (53) Thompson, C. M.; Cannon, A.; West, S.; Ghersi, D.; Atri, P.; Bhatia, R.; Smith, L.; Rachagani, S.; Wichman, C.; Kumar, S.; et al. Mucin Expression and Splicing Determine Novel Subtypes and Patient Mortality in Pancreatic Ductal Adenocarcinoma. *Clinical Cancer Research* **2021**, 27 (24), 6787-6799.
- (54) Adsay, N. V.; Merati, K.; Basturk, O.; Iacobuzio-Donahue, C.; Levi, E.; Cheng, J. D.; Sarkar, F. H.; Hruban, R. H.; Klimstra, D. S. Pathologically and Biologically Distinct Types of Epithelium in Intraductal Papillary Mucinous Neoplasms: Delineation of an "Intestinal" Pathway of Carcinogenesis in the Pancreas. *The American Journal of Surgical Pathology* **2004**, 28 (7), 839-848.
- (55) Nagata, K.; Horinouchi, M.; Saitou, M.; Higashi, M.; Nomoto, M.; Goto, M.; Yonezawa, S. Mucin expression profile in pancreatic cancer and the precursor lesions. *Journal of Hepato-Biliary-Pancreatic Surgery* **2007**, 14 (3), 243-254.
- (56) Jensen, P. H.; Kolarich, D.; Packer, N. H. Mucin-type O-glycosylation – putting the pieces together. *The FEBS Journal* **2010**, 277 (1), 81-94.
- (57) Wells, L.; Vosseller, K.; Hart, G. W. Glycosylation of Nucleocytoplasmic Proteins: Signal Transduction and O-GlcNAc. *Science* **2001**, 291 (5512), 2376-2378.

- (58) Chou, T.-Y.; Hart, G. W.; Dang, C. V. c-Myc Is Glycosylated at Threonine 58, a Known Phosphorylation Site and a Mutational Hot Spot in Lymphomas \*. *Journal of Biological Chemistry* **1995**, *270* (32), 18961-18965.
- (59) Kelly, W. G.; Dahmus, M. E.; Hart, G. W. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *Journal of Biological Chemistry* **1993**, *268* (14), 10416-10424.
- (60) Comer, F. I.; Hart, G. W. O-Glycosylation of Nuclear and Cytosolic Proteins: DYNAMIC INTERPLAY BETWEEN O-GlcNAc AND O-PHOSPHATE. *Journal of Biological Chemistry* **2000**, *275* (38), 29179-29182.
- (61) Friedenreich, V. The Thomsen hemagglutination phenomenon: Production of a specific receptor quality in red corpuscles by bacterial activity. (*No Title*) **1930**.
- (62) Dausset, J.; Moullec, J.; Bernard, J. Acquired hemolytic anemia with polyagglutinability of red blood cells due to a new factor present in normal human serum (Anti-Tn). *Blood* **1959**, *14* (10), 1079-1093.
- (63) Springer, G. F. Immunoreactive T and Tn epitopes in cancer diagnosis, prognosis, and immunotherapy. *Journal of Molecular Medicine* **1997**, *75* (8), 594-602.
- (64) Springer, G. F. T and Tn Pancarcinoma Markers: Autoantigenic Adhesion Molecules in Pathogenesis, Prebiopsy Carcinoma-Detection, and Long-Term Breast Carcinoma Immunotherapy. **1995**, *6* (1), 57-85.
- (65) Springer, G. F.; Desai, P. R.; Ghazizadeh, M.; Tegtmeier, H. T/Tn pancarcinoma autoantigens: fundamental, diagnostic, and prognostic aspects. *Cancer Detect Prev* **1995**, *19* (2), 173-182.
- (66) Kim, Y. S.; Gum, J.; Brockhausen, I. Mucin glycoproteins in neoplasia. *Glycoconjugate Journal* **1996**, *13* (5), 693-707.
- (67) Itzkowitz, S.; Kjeldsen, T.; Frieria, A.; Hakomori, S.-I.; Yang, U.-S.; Kim, Y. S. Expression of Tn, sialosyl Tn, and T antigens in human pancreas. *Gastroenterology* **1991**, *100* (6), 1691-1700.
- (68) Ho, J. J. L.; Bi, N.; Siddiki, B.; Chung, Y.-s.; Yuan, M.; Kim, Y. S. Multiple Forms of Intracellular and Secreted Mucins in a Pancreatic Cancer Cell Line1. *Cancer Research* **1993**, *53* (4), 884-890.
- (69) Ho, J. J. L.; Chung, Y.-s.; Fujimoto, Y.; Bi, N.; Ryan, W.; Yuan, S.-z.; Byrd, J. C.; Kim, Y. S. Mucin-like Antigens in a Human Pancreatic Cancer Cell Line Identified by Murine Monoclonal Antibodies SPan-1 and YPan-11. *Cancer Research* **1988**, *48* (14), 3924-3931.
- (70) Ho, J. J.; Siddiki, B.; Kim, Y. S. Association of sialyl-Lewis<sub>x</sub> and sialyl-Lewis<sub>x</sub> with MUC-1 apomucin in a pancreatic cancer cell line. *Cancer research* **1995**, *55* (16), 3659-3663.
- (71) Kim, Y. S.; Itzkowitz, S. H.; Yuan, M.; Chung, Y.-s.; Satake, K.; Umeyama, K.; Hakomori, S.-i. Lex and Ley antigen expression in human pancreatic cancer. *Cancer research* **1988**, *48* (2), 475-482.
- (72) Mansson, J.-E.; Fredman, P.; Nilsson, O.; Lindholm, L.; Holmgren, J.; Svennerholm, L. Chemical structure of carcinoma ganglioside antigens defined by monoclonal antibody C-50 and some allied gangliosides of human pancreatic adenocarcinoma. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* **1985**, *834* (1), 110-117.
- (73) Nilsson, O.; Månsson, J.-E.; Lindholm, L.; Holmgren, J.; Svennerholm, L. Sialosyllactotetraosylceramide, a novel ganglioside antigen detected in human carcinomas by a monoclonal antibody. *FEBS Letters* **1985**, *182* (2), 398-402.
- (74) Goonetilleke, K. S.; Siriwardena, A. K. Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. *European Journal of Surgical Oncology (EJSO)* **2007**, *33* (3), 266-270.
- (75) Banfi, G.; Zerbi, A.; Pastori, S.; Parolini, D.; Di Carlo, V.; Bonini, P. Behavior of tumor markers CA19.9, CA195, CAM43, CA242, and TPS in the diagnosis and follow-up of pancreatic cancer. *Clinical Chemistry* **1993**, *39* (3), 420-423.
- (76) Richter, J. M.; Christensen, M. R.; Rustgi, A. K.; Silverstein, M. D. The Clinical Utility of the CA19-9 Radioimmunoassay for the Diagnosis of Pancreatic Cancer Presenting as Pain or Weight Loss: A Cost-Effectiveness Analysis. *Archives of Internal Medicine* **1989**, *149* (10), 2292-2297.
- (77) Ni, X. G.; Bai, X. F.; Mao, Y. L.; Shao, Y. F.; Wu, J. X.; Shan, Y.; Wang, C. F.; Wang, J.; Tian, Y. T.; Liu, Q.; et al. The clinical value of serum CEA, CA19-9, and CA242 in the diagnosis and prognosis of pancreatic cancer. *European Journal of Surgical Oncology (EJSO)* **2005**, *31* (2), 164-169.

- (78) Harmenberg, U.; Wahren, B.; Wiechel, K.-L. Tumor Markers Carbohydrate Antigens CA 19-9 and CA-50 and Carcinoembryonic Antigen in Pancreatic Cancer and Benign Diseases of the Pancreatobiliary Tract. *Cancer Research* **1988**, *48* (7), 1985-1988.
- (79) Barone, D.; Onetto, M.; Conio, M.; Paganuzzi, M.; Saccomanno, S.; Aste, H.; Pugliese, V. Ca 19-9 Assay in Patients with Extrahepatic Cholestatic Jaundice. *The International Journal of Biological Markers* **1988**, *3* (2), 95-100.
- (80) *Cancer Stat Facts: Testicular Cancer*. National Cancer Institute, <https://seer.cancer.gov/statfacts/html/testis.html> (accessed 03.07.2024).
- (81) Gill, M. S.; Shah, S. H.; Soomro, I. N.; Kayani, N.; Hasan, S. H. Morphological pattern of testicular tumors. *Journal of Pakistan medical association* **2000**, *50* (4), 110.
- (82) Cheng, L.; Albers, P.; Berney, D. M.; Feldman, D. R.; Daugaard, G.; Gilligan, T.; Looijenga, L. H. J. Testicular cancer. *Nature Reviews Disease Primers* **2018**, *4* (1), 29.
- (83) Oosterhuis, J. W.; Looijenga, L. H. J. Testicular germ-cell tumours in a broader perspective. *Nature Reviews Cancer* **2005**, *5* (3), 210-222.
- (84) Oosterhuis, J. W.; Looijenga, L. H. J.; van Echten, J.; de Jong, B. Chromosomal constitution and developmental potential of human germ cell tumors and teratomas. *Cancer Genetics and Cytogenetics* **1997**, *95* (1), 96-102.
- (85) Skowron, M. A.; Kotthoff, M.; Bremmer, F.; Ruhnke, K.; Parmaksiz, F.; Richter, A.; Küffer, S.; Reuter-Jessen, K.; Pauls, S.; Stefanski, A.; et al. Targeting CLDN6 in germ cell tumors by an antibody-drug-conjugate and studying therapy resistance of yolk-sac tumors to identify and screen specific therapeutic options. *Molecular Medicine* **2023**, *29* (1), 40.
- (86) Skowron, M. A.; Oing, C.; Bremmer, F.; Ströbel, P.; Murray, M. J.; Coleman, N.; Amatruda, J. F.; Honecker, F.; Bokemeyer, C.; Albers, P.; et al. The developmental origin of cancers defines basic principles of cisplatin resistance. *Cancer Letters* **2021**, *519*, 199-210.
- (87) Ushiku, T.; Shinozaki-Ushiku, A.; Maeda, D.; Morita, S.; Fukayama, M. Distinct expression pattern of claudin-6, a primitive phenotypic tight junction molecule, in germ cell tumours and visceral carcinomas. *Histopathology* **2012**, *61* (6), 1043-1056.
- (88) Johansson, M. P.; Maaheimo, H.; Ekholm, F. S. New insight on the structural features of the cytotoxic auristatins MMAE and MMAF revealed by combined NMR spectroscopy and quantum chemical modelling. *Scientific Reports* **2017**, *7* (1), 15920.
- (89) Wruck, W.; Bremmer, F.; Kotthoff, M.; Fichtner, A.; Skowron, M. A.; Schönberger, S.; Calaminus, G.; Vokuhl, C.; Pfister, D.; Heidenreich, A.; et al. The pioneer and differentiation factor FOXA2 is a key driver of yolk-sac tumour formation and a new biomarker for paediatric and adult yolk-sac tumours. *Journal of Cellular and Molecular Medicine* **2021**, *25* (3), 1394-1405.
- (90) Ahmed, T.; Bosl, G. J.; Hajdu, S. I. Teratoma with malignant transformation in germ cell tumors in men. *Cancer* **1985**, *56* (4), 860-863.
- (91) Comiter, C. V.; Kibel, A. S.; Richie, J. P.; Nucci, M. R.; Renshaw, A. A. Prognostic features of teratomas with malignant transformation: a clinicopathological study of 21 cases. *The journal of Urology* **1998**, *159* (3), 859-863.
- (92) Hwang, M. J.; Hamza, A.; Zhang, M.; Tu, S.-M.; Pisters, L. L.; Czerniak, B.; Guo, C. C. Somatic-type malignancies in testicular germ cell tumors: a clinicopathologic study of 63 cases. *The American journal of surgical pathology* **2022**, *46* (1), 11-17.
- (93) Bremmer, F.; Pongratanakul, P.; Skowron, M.; Che, Y.; Richter, A.; Küffer, S.; Reuter-Jessen, K.; Bohnenberger, H.; Pauls, S.; Kresbach, C.; et al. Characterizing the mutational burden, DNA methylation landscape, and proteome of germ cell tumor-related somatic-type malignancies to identify the tissue-of-origin, mechanisms of therapy resistance, and druggable targets. *British Journal of Cancer* **2023**.
- (94) Logothetis, C. J.; Samuels, M. L.; Trindade, A.; Johnson, D. E. The growing teratoma syndrome. *Cancer* **1982**, *50* (8), 1629-1635.
- (95) Paffenholz, P.; Pfister, D.; Matveev, V.; Heidenreich, A. Diagnosis and management of the growing teratoma syndrome: A single-center experience and review of the literature. *Urologic Oncology: Seminars and Original Investigations* **2018**, *36* (12), 529.e523-529.e530.
- (96) Pongratanakul, P.; Bremmer, F.; Pauls, S.; Poschmann, G.; Kresbach, C.; Parmaksiz, F.; Skowron, M. A.; Fuß, J.; Stephan, A.; Paffenholz, P.; et al. Assessing the risk to develop a growing teratoma syndrome based on molecular and epigenetic subtyping as well as novel secreted biomarkers. *Cancer Letters* **2024**, 216673.
- (97) Stanta, G. Archive Tissues. 2011; pp 3-4.

- (98) Bevilacqua, G.; Bosman, F.; Dassel, T.; Höfler, H.; Janin, A.; Langer, R.; Larsimont, D.; Morente, M. M.; Riegman, P.; Schirmacher, P.; et al. The role of the pathologist in tissue banking: European Consensus Expert Group Report. *Virchows Archiv* **2010**, *456* (4), 449-454.
- (99) Gustafsson, O. J. R.; Arentz, G.; Hoffmann, P. Proteomic developments in the analysis of formalin-fixed tissue. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2015**, *1854* (6), 559-580.
- (100) Ilyin, S. E.; Belkowski, S. M.; Plata-Salamán, C. R. Biomarker discovery and validation: technologies and integrative approaches. *Trends in Biotechnology* **2004**, *22* (8), 411-416.
- (101) Palmer-Toy, D. E.; Krastins, B.; Sarracino, D. A.; Nadol, J. B.; Merchant, S. N. Efficient Method for the Proteomic Analysis of Fixed and Embedded Tissues. *Journal of Proteome Research* **2005**, *4* (6), 2404-2411.
- (102) Ikeda, K.; Monden, T.; Kanoh, T.; Tsujie, M.; Izawa, H.; Haba, A.; Ohnishi, T.; Sekimoto, M.; Tomita, N.; Shiozaki, H. Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections. *Journal of Histochemistry & Cytochemistry* **1998**, *46* (3), 397-403.
- (103) Wolff, C.; Schott, C.; Porschewski, P.; Reischauer, B.; Becker, K.-F. Successful Protein Extraction from Over-Fixed and Long-Term Stored Formalin-Fixed Tissues. *Plos One* **2011**, *6* (1), e16353.
- (104) Addis, M. F.; Tanca, A.; Pagnozzi, D.; Crobu, S.; Fanciulli, G.; Cossu-Rocca, P.; Uzzau, S. Generation of high-quality protein extracts from formalin-fixed, paraffin-embedded tissues. *PROTEOMICS* **2009**, *9* (15), 3815-3823.
- (105) Bravo, S.; Garcia-Vence, M.; Chantada-Vazquez, M. d. P.; Sosa-Fajardo, A.; Agra, R.; Barcia de la Iglesia, A.; Otero-Glez, A.; Garcia-Gonzalez, M. A.; Cameselle-Teijeiro, J.; Nuñez, C. Protein extraction from FFPE kidney tissue samples: A review of the literature and characterization of techniques. *Frontiers in Medicine* **2021**, *8*, 553.
- (106) Shi, S.-R.; Taylor, C. R.; Fowler, C. B.; Mason, J. T. Complete solubilization of formalin-fixed, paraffin-embedded tissue may improve proteomic studies. *PROTEOMICS – Clinical Applications* **2013**, *7* (3-4), 264-272.
- (107) Stanta, G. *Guidelines for molecular analysis in archive tissues*; Springer Science & Business Media, 2011.
- (108) Bussolati, G.; Chiusa, L.; Cimino, A.; D'Armento, G. Tissue transfer to pathology labs: under vacuum is the safe alternative to formalin. *Virchows Archiv* **2008**, *452* (2), 229-231.
- (109) Blow, N. Tissue issues. *Nature* **2007**, *448* (7156), 959-960.
- (110) Hallmans, G.; Vaught, J. B. Best practices for establishing a biobank. In *Methods in Biobanking*, Springer, 2011; pp 241-260.
- (111) Pitt, K. E.; Campbell, L. D.; Skubitz, A. P.; Aamodt, R. L.; Anouna, A.; Baird, P.; Beck, J. C.; Bledsoe, M. J.; De Souza, Y.; Grizzle, W. E. Best practices for repositories I: Collection, storage, and retrieval of human biological materials for research. *Cell Preservation Technology* **2005**, *3* (1), 5-48.
- (112) Xie, R.; Chung, J.-Y.; Ylaya, K.; Williams, R. L.; Guerrero, N.; Nakatsuka, N.; Badie, C.; Hewitt, S. M. Factors Influencing the Degradation of Archival Formalin-Fixed Paraffin-Embedded Tissue Sections. *Journal of Histochemistry & Cytochemistry* **2011**, *59* (4), 356-365.
- (113) Fraenkel-Conrat, H.; Brandon, B. A.; Olcott, H. S. The reaction of formaldehyde with proteins: IV. Participation of indole groups. Gramicidin. *Journal of Biological Chemistry* **1947**, *168* (1), 99-118.
- (114) Fraenkel-Conrat, H.; Olcott, H. S. Reaction of formaldehyde with proteins: VI. Cross-linking of amino groups with phenol, imidazole, or indole groups. *Journal of Biological Chemistry* **1948**, *174* (3), 827-843.
- (115) Hoffman, E. A.; Frey, B. L.; Smith, L. M.; Auble, D. T. Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J Biol Chem* **2015**, *290* (44), 26404-26411.
- (116) Magdeldin, S.; Yamamoto, T. Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues. *PROTEOMICS* **2012**, *12* (7), 1045-1058.
- (117) Sutherland, B. W.; Toews, J.; Kast, J. Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions. *Journal of Mass Spectrometry* **2008**, *43* (6), 699-715.
- (118) Orlando, V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends in biochemical sciences* **2000**, *25* (3), 99-104.
- (119) Shi, S. R.; Key, M. E.; Kalra, K. L. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. *Journal of Histochemistry & Cytochemistry* **1991**, *39* (6), 741-748.

- (120) Jones, M.; Banks, P.; Caron, B. Transition metal salts as adjuncts to formalin for tissue fixation. *Lab Invest* **1981**, *44* (32).
- (121) Shi, S.-R.; Shi, Y.; Taylor, C. R. Antigen retrieval immunohistochemistry: review and future prospects in research and diagnosis over two decades. *Journal of Histochemistry & Cytochemistry* **2011**, *59* (1), 13-32.
- (122) Chu, W.-S.; Liang, Q.; Liu, J.; Wei, M. Q.; Winters, M.; Liotta, L.; Sandberg, G.; Gong, M. A nondestructive molecule extraction method allowing morphological and molecular analyses using a single tissue section. *Laboratory investigation* **2005**, *85* (11), 1416-1428.
- (123) Prieto, D. A.; Hood, B. L.; Darfler, M. M.; Guiel, T. G.; Lucas, D. A.; Conrads, T. P.; Veenstra, T. D.; Krizman, D. B. Liquid Tissue™: proteomic profiling of formalin-fixed tissues. *BioTechniques* **2005**, *38* (6S), S32-S35.
- (124) Yamashita, S.; Okada, Y. Mechanisms of heat-induced antigen retrieval: analyses in vitro employing SDS-PAGE and immunohistochemistry. *Journal of Histochemistry & Cytochemistry* **2005**, *53* (1), 13-21.
- (125) Fowler, C. B.; O'leary, T. J.; Mason, J. T. Techniques of protein extraction from FFPE tissue/cells for mass spectrometry. *Antigen Retrieval Immunohistochemistry Based Research and Diagnostics* **2010**, 333-346.
- (126) Jiang, X.; Jiang, X.; Feng, S.; Tian, R.; Ye, M.; Zou, H. Development of Efficient Protein Extraction Methods for Shotgun Proteome Analysis of Formalin-Fixed Tissues. *Journal of Proteome Research* **2007**, *6* (3), 1038-1047.
- (127) Proc, J. L.; Kuzyk, M. A.; Hardie, D. B.; Yang, J.; Smith, D. S.; Jackson, A. M.; Parker, C. E.; Borchers, C. H. A Quantitative Study of the Effects of Chaotropic Agents, Surfactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin. *Journal of Proteome Research* **2010**, *9* (10), 5422-5437.
- (128) Hervey; Strader, M. B.; Hurst, G. B. Comparison of Digestion Protocols for Microgram Quantities of Enriched Protein Samples. *Journal of Proteome Research* **2007**, *6* (8), 3054-3061.
- (129) Poulsen, J. W.; Madsen, C. T.; Young, C.; Poulsen, F. M.; Nielsen, M. L. Using Guanidine-Hydrochloride for Fast and Efficient Protein Digestion and Single-step Affinity-purification Mass Spectrometry. *Journal of Proteome Research* **2013**, *12* (2), 1020-1030.
- (130) Ferro, M.; Seigneurin-Berny, D.; Rolland, N.; Chapel, A.; Salvi, D.; Garin, J.; Joyard, J. Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins. *ELECTROPHORESIS* **2000**, *21* (16), 3517-3526.
- (131) Chertov, O.; Biragyn, A.; Kwak, L. W.; Simpson, J. T.; Boronina, T.; Hoang, V. M.; Prieto, D. A.; Conrads, T. P.; Veenstra, T. D.; Fisher, R. J. Organic solvent extraction of proteins and peptides from serum as an effective sample preparation for detection and identification of biomarkers by mass spectrometry. *PROTEOMICS* **2004**, *4* (4), 1195-1203.
- (132) Wang, H.; Qian, W.-J.; Mottaz, H. M.; Clauss, T. R. W.; Anderson, D. J.; Moore, R. J.; Camp, D. G.; Khan, A. H.; Sforza, D. M.; Pallavicini, M.; et al. Development and Evaluation of a Micro- and Nanoscale Proteomic Sample Preparation Method. *Journal of Proteome Research* **2005**, *4* (6), 2397-2403.
- (133) Zhang, H.; Lin, Q.; Ponnusamy, S.; Kothandaraman, N.; Lim, T. K.; Zhao, C.; Kit, H. S.; Arijit, B.; Rauff, M.; Hew, C.-L.; et al. Differential recovery of membrane proteins after extraction by aqueous methanol and trifluoroethanol. *PROTEOMICS* **2007**, *7* (10), 1654-1663.
- (134) Wigley, W. C.; Vijayakumar, S.; Jones, J. D.; Slaughter, C.; Thomas, P. J. Transmembrane Domain of Cystic Fibrosis Transmembrane Conductance Regulator: Design, Characterization, and Secondary Structure of Synthetic Peptides m1-m6. *Biochemistry* **1998**, *37* (3), 844-853.
- (135) Blonder, J.; Goshe, M. B.; Moore, R. J.; Pasa-Tolic, L.; Masselon, C. D.; Lipton, M. S.; Smith, R. D. Enrichment of Integral Membrane Proteins for Proteomic Analysis Using Liquid Chromatography-Tandem Mass Spectrometry. *Journal of Proteome Research* **2002**, *1* (4), 351-360.
- (136) Goshe, M. B.; Blonder, J.; Smith, R. D. Affinity Labeling of Highly Hydrophobic Integral Membrane Proteins for Proteome-Wide Analysis. *Journal of Proteome Research* **2003**, *2* (2), 153-161.
- (137) Deshusses, J. M. P.; Burgess, J. A.; Scherl, A.; Wenger, Y.; Walter, N.; Converset, V.; Paesano, S.; Corthals, G. L.; Hochstrasser, D. F.; Sanchez, J.-C. Exploitation of specific properties of trifluoroethanol for extraction and separation of membrane proteins. *PROTEOMICS* **2003**, *3* (8), 1418-1424.

- (138) Zuobi-Hasona, K.; Crowley, P. J.; Hasona, A.; Bleiweis, A. S.; Brady, L. J. Solubilization of cellular membrane proteins from *Streptococcus mutans* for two-dimensional gel electrophoresis. *ELECTROPHORESIS* **2005**, *26* (6), 1200-1205.
- (139) Tian, Y.; Gurley, K.; Meany, D. L.; Kemp, C. J.; Zhang, H. N-linked glycoproteomic analysis of formalin-fixed and paraffin-embedded tissues. *Journal of proteome research* **2009**, *8* (4), 1657-1662.
- (140) Sprung, R. W., Jr.; Brock, J. W. C.; Tanksley, J. P.; Li, M.; Washington, M. K.; Slebos, R. J. C.; Liebler, D. C. Equivalence of Protein Inventories Obtained from Formalin-fixed Paraffin-embedded and Frozen Tissue in Multidimensional Liquid Chromatography-Tandem Mass Spectrometry Shotgun Proteomic Analysis. *Molecular & Cellular Proteomics* **2009**, *8* (8), 1988-1998.
- (141) Coscia, F.; Doll, S.; Bech, J. M.; Schweizer, L.; Mund, A.; Lengyel, E.; Lindebjerg, J.; Madsen, G. I.; Moreira, J. M.; Mann, M. A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis. *J Pathol* **2020**, *251* (1), 100-112.
- (142) Shi, S. R.; Imam, S. A.; Young, L.; Cote, R. J.; Taylor, C. R. Antigen retrieval immunohistochemistry under the influence of pH using monoclonal antibodies. *Journal of Histochemistry & Cytochemistry* **1995**, *43* (2), 193-201.
- (143) Shi, S.-R.; Shi, Y.; Taylor, C. R.; Gu, J. New Dimensions of Antigen Retrieval Technique: 28 Years of Development, Practice, and Expansion. *Applied Immunohistochemistry & Molecular Morphology* **2019**, *27* (10), 715-721.
- (144) O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry* **1975**, *250* (10), 4007-4021.
- (145) Sitek, B.; Sipos, B.; Alkatout, I.; Poschmann, G.; Stephan, C.; Schulenburg, T.; Marcus, K.; Luttgies, J.; Dittert, D. D.; Baretton, G.; et al. Analysis of the pancreatic tumor progression by a quantitative proteomic approach and immunohistochemical validation. *J Proteome Res* **2009**, *8* (4), 1647-1656.
- (146) Sitek, B.; Luttgies, J.; Marcus, K.; Kloppel, G.; Schmiegel, W.; Meyer, H. E.; Hahn, S. A.; Stuhler, K. Application of fluorescence difference gel electrophoresis saturation labelling for the analysis of microdissected precursor lesions of pancreatic ductal adenocarcinoma. *Proteomics* **2005**, *5* (10), 2665-2679.
- (147) Rabilloud, T.; Lelong, C. Two-dimensional gel electrophoresis in proteomics: A tutorial. *Journal of Proteomics* **2011**, *74* (10), 1829-1841.
- (148) Rabilloud, T. Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains. *PROTEOMICS* **2002**, *2* (1), 3-10.
- (149) Lee, P. Y.; Saraygord-Afshari, N.; Low, T. Y. The evolution of two-dimensional gel electrophoresis - from proteomics to emerging alternative applications. *Journal of Chromatography A* **2020**, *1615*, 460763.
- (150) Zhang, Z.; Dubiak, K. M.; Huber, P. W.; Dovichi, N. J. Miniaturized Filter-Aided Sample Preparation (MICRO-FASP) Method for High Throughput, Ultrasensitive Proteomics Sample Preparation Reveals Proteome Asymmetry in *Xenopus laevis* Embryos. *Analytical Chemistry* **2020**, *92* (7), 5554-5560.
- (151) Wiśniewski, J. R.; Zougman, A.; Mann, M. Combination of FASP and StageTip-Based Fractionation Allows In-Depth Analysis of the Hippocampal Membrane Proteome. *Journal of Proteome Research* **2009**, *8* (12), 5674-5678.
- (152) Wiśniewski, J. R. Quantitative Evaluation of Filter Aided Sample Preparation (FASP) and Multienzyme Digestion FASP Protocols. *Analytical Chemistry* **2016**, *88* (10), 5438-5443.
- (153) Ni, M.-w.; Wang, L.; Chen, W.; Mou, H.-z.; Zhou, J.; Zheng, Z.-g. Modified filter-aided sample preparation (FASP) method increases peptide and protein identifications for shotgun proteomics. *Rapid Communications in Mass Spectrometry* **2017**, *31* (2), 171-178.
- (154) Tang, L.; Wu, Z.; Wang, J.; Zhang, X. Formaldehyde Derivatization: An Unexpected Side Reaction During Filter-Aided Sample Preparation. *Analytical Chemistry* **2020**, *92* (18), 12120-12125.
- (155) Gomez-Zepeda, D.; Michna, T.; Ziesmann, T.; Distler, U.; Tenzer, S. HowDirty: An R package to evaluate molecular contaminants in LC-MS experiments. *PROTEOMICS* **2023**, *n/a* (n/a), 2300134.
- (156) Ostasiewicz, P.; Zielinska, D. F.; Mann, M.; Wiśniewski, J. R. Proteome, Phosphoproteome, and N-Glycoproteome Are Quantitatively Preserved in Formalin-Fixed Paraffin-Embedded Tissue and Analyzable by High-Resolution Mass Spectrometry. *Journal of Proteome Research* **2010**, *9* (7), 3688-3700.

- (157) Wiśniewski, J. R.; Zielinska, D. F.; Mann, M. Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method. *Analytical Biochemistry* **2011**, *410* (2), 307-309.
- (158) Hughes, C. S.; Foehr, S.; Garfield, D. A.; Furlong, E. E.; Steinmetz, L. M.; Krijgsveld, J. Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular Systems Biology* **2014**, *10* (10), 757.
- (159) Hughes, C. S.; Moggridge, S.; Müller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols* **2019**, *14* (1), 68-85.
- (160) Sielaff, M.; Kuharev, J.; Bohn, T.; Hahlbrock, J.; Bopp, T.; Tenzer, S.; Distler, U. Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *Journal of Proteome Research* **2017**, *16* (11), 4060-4072.
- (161) Kassem, S.; van der Pan, K.; de Jager, A. L.; Naber, B. A.; de Laat, I. F.; Louis, A.; van Dongen, J. J.; Teodosio, C.; Díez, P. Proteomics for Low Cell Numbers: How to Optimize the Sample Preparation Workflow for Mass Spectrometry Analysis. *Journal of Proteome Research* **2021**.
- (162) Griesser, E.; Wyatt, H.; Ten Have, S.; Stierstorfer, B.; Lenter, M.; Lamond, A. I. Quantitative Profiling of the Human Substantia Nigra Proteome from Laser-capture Microdissected FFPE Tissue. *Mol Cell Proteomics* **2020**, *19* (5), 839-851.
- (163) Zecha, J.; Satpathy, S.; Kanashova, T.; Avanesian, S. C.; Kane, M. H.; Clauser, K. R.; Mertins, P.; Carr, S. A.; Kuster, B. TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach\*[S]. *Molecular & Cellular Proteomics* **2019**, *18* (7), 1468-1478.
- (164) Pappireddi, N.; Martin, L.; Wühr, M. A review on quantitative multiplexed proteomics. *Chembiochem* **2019**, *20* (10), 1210-1224.
- (165) Cheung, T. K.; Lee, C.-Y.; Bayer, F. P.; McCoy, A.; Kuster, B.; Rose, C. M. Defining the carrier proteome limit for single-cell proteomics. *Nature Methods* **2021**, *18* (1), 76-83.
- (166) Bekker-Jensen, D. B.; Bernhardt, O. M.; Högrefe, A.; Martinez-Val, A.; Verbeke, L.; Gandhi, T.; Kelstrup, C. D.; Reiter, L.; Olsen, J. V. Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nature Communications* **2020**, *11* (1), 787.
- (167) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis\*. *Molecular & Cellular Proteomics* **2012**, *11* (6), O111.016717.
- (168) Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods* **2020**, *17* (1), 41-44.
- (169) Gandhi, T.; Verbeke, L.; Bernhardt, O. M.; Muntel, J.; Müller, S.; Bruderer, R. M.; and Lukas Reiter, Y. X. Best of Both Worlds: Spectronaut Combines the Depth of Resource with iRT-Precision of Project Specific Data.
- (170) Pino, L. K.; Searle, B. C.; Bollinger, J. G.; Nunn, B.; MacLean, B.; MacCoss, M. J. The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrometry Reviews* **2020**, *39* (3), 229-244.
- (171) Zhang, F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T. Data-independent acquisition mass spectrometry-based proteomics and software tools: a glimpse in 2020. *Proteomics* **2020**, *20* (17-18), 1900276.
- (172) Kim, Y. J.; Sweet, S. M. M.; Egertson, J. D.; Sedgewick, A. J.; Woo, S.; Liao, W.-I.; Merrihew, G. E.; Searle, B. C.; Vaske, C.; Heaton, R.; et al. Data-Independent Acquisition Mass Spectrometry To Quantify Protein Levels in FFPE Tumor Biopsies for Molecular Diagnostics. *Journal of Proteome Research* **2019**, *18* (1), 426-435.
- (173) Stanojevic, A.; Samiotaki, M.; Lygirou, V.; Marinkovic, M.; Nikolic, V.; Stojanovic-Rundic, S.; Jankovic, R.; Vlahou, A.; Panayotou, G.; Fijneman, R. J. A.; et al. Data independent acquisition mass spectrometry (DIA-MS) analysis of FFPE rectal cancer samples offers in depth proteomics characterization of response to neoadjuvant chemoradiotherapy. *medRxiv* **2023**, 2023.2005.2012.23289671.
- (174) Weke, K.; Kote, S.; Faktor, J.; Al Shboul, S.; Uwugiaren, N.; Brennan, P. M.; Goodlett, D. R.; Hupp, T. R.; Dapic, I. DIA-MS proteome analysis of formalin-fixed paraffin-embedded glioblastoma tissues. *Analytica Chimica Acta* **2022**, *1204*, 339695.
- (175) Keam, S. P.; Gulati, T.; Gamell, C.; Caramia, F.; Huang, C.; Schittenhelm, R. B.; Kleifeld, O.; Neeson, P. J.; Haupt, Y.; Williams, S. G. Exploring the oncoproteomic response of human

- prostate cancer to therapeutic radiation using data-independent acquisition (DIA) mass spectrometry. *The Prostate* **2018**, *78* (8), 563-575.
- (176) O'Rourke, M. B.; Padula, M. P. Analysis of formalin-fixed, paraffin-embedded (FFPE) tissue via proteomic techniques and misconceptions of antigen retrieval. *BioTechniques* **2016**, *60* (5), 229-238.
- (177) Tanca, A.; Pagnozzi, D.; Addis, M. F. Setting proteins free: Progresses and achievements in proteomics of formalin-fixed, paraffin-embedded tissues. *PROTEOMICS – Clinical Applications* **2012**, *6* (1-2), 7-21.
- (178) Hwang, S. I.; Thumar, J.; Lundgren, D. H.; Rezaul, K.; Mayya, V.; Wu, L.; Eng, J.; Wright, M. E.; Han, D. K. Direct cancer tissue proteomics: a method to identify candidate cancer biomarkers from formalin-fixed paraffin-embedded archival tissues. *Oncogene* **2007**, *26* (1), 65-76.
- (179) Paulo, J. A.; Lee, L. S.; Banks, P. A.; Steen, H.; Conwell, D. L. Proteomic analysis of formalin-fixed paraffin-embedded pancreatic tissue using liquid chromatography tandem mass spectrometry. *Pancreas* **2012**, *41* (2), 175-185.
- (180) Nirmalan, N. J.; Hughes, C.; Peng, J.; McKenna, T.; Langridge, J.; Cairns, D. A.; Harnden, P.; Selby, P. J.; Banks, R. E. Initial development and validation of a novel extraction method for quantitative mining of the formalin-fixed, paraffin-embedded tissue proteome for biomarker investigations. *J Proteome Res* **2011**, *10* (2), 896-906.
- (181) Crockett, D. K.; Lin, Z.; Vaughn, C. P.; Lim, M. S.; Elenitoba-Johnson, K. S. J. Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS. *Laboratory Investigation* **2005**, *85* (11), 1405-1415.
- (182) Bauden, M.; Kristl, T.; Andersson, R.; Marko-Varga, G.; Ansari, D. Characterization of histone-related chemical modifications in formalin-fixed paraffin-embedded and fresh-frozen human pancreatic cancer xenografts using LC-MS/MS. *Laboratory Investigation* **2017**, *97* (3), 279-288.
- (183) Fowler, C.; O'Leary, T.; Mason, J. Improving the Proteomic Analysis of Archival Tissue by Using Pressure-Assisted Protein Extraction: A Mechanistic Approach. *Journal of proteomics & bioinformatics* **2014**, *7*, 151-157.
- (184) Lenz, T. Devices, systems and methods for separating magnetic particles. Google Patents: 2012.
- (185) Doellinger, J.; Schneider, A.; Hoeller, M.; Lasch, P. Sample Preparation by Easy Extraction and Digestion (SPEED) - A Universal, Rapid, and Detergent-free Protocol for Proteomics Based on Acid Extraction. *Molecular & Cellular Proteomics* **2020**, *19* (1), 209-222.
- (186) Takikita, M.; Chung, J.-Y.; Hewitt, S. M. Tissue microarrays enabling high-throughput molecular pathology. *Current Opinion in Biotechnology* **2007**, *18* (4), 318-325.
- (187) Hewitt, S. M. Design, construction, and use of tissue microarrays. In *Protein Arrays*, Springer, 2004; pp 61-72.
- (188) Espina, V.; Mehta, A. I.; Winters, M. E.; Calvert, V.; Wulfkuhle, J.; Petricoin III, E. F.; Liotta, L. A. Protein microarrays: Molecular profiling technologies for clinical specimens. *PROTEOMICS* **2003**, *3* (11), 2091-2100.
- (189) Kumar, L.; Futschik, M. E. Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* **2007**, *2* (1), 5.
- (190) Rodrigues, E.; Macauley, M. S. Hypersialylation in Cancer: Modulation of Inflammation and Therapeutic Opportunities. *Cancers* **2018**, *10* (6), 207.
- (191) Karsten, U.; Goletz, S. What controls the expression of the core-1 (Thomsen—Friedenreich) glycotope on tumor cells? *Biochemistry (Moscow)* **2015**, *80*, 801-807.