Dissertation
submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by

M. Sc. Marius Nicolaus Felix

Born in: Freising, Germany
Oral Examination: 03.12.2021

# Proteomic Exploration of IDH-Mutant Gliomas using Mass Spectrometry

Referees:

Prof. Dr. Magnus von Knebel Doeberitz

Dr. Christiane Opitz

## Scientific contributions within the Scope of this thesis:

**Major parts of this thesis will be published in peer reviewed journals.**

- **Marius Felix**, Dennis Friedel, Ashok Kumar Jayavelu, Anne-Kathrin Reinhard, Uwe Warnken, Damian Stichel, Daniel Schrimpf, Andrey Korshunov, Yueting Wang, Tobias Kessler, Nima Etminan, Andreas Unterberg, Christel Herold-Mende, Laura Heikaus, Felix Sahm, Wolfgang Wick, Andreas von Deimling and David E. Reuss. **HIP1R and VIM immunohistochemistry predict 1p/19q status in IDH-mutant glioma. Manuscript in preparation.**

- **Marius Felix***, Dennis Friedel*, Ashok Kumar Jayavelu, Uwe Warnken, Damian Stichel, Daniel Schrimpf, Andrey Korshunov, Yueting Wang, Tobias Kessler, Andreas Unterberg, Christel Herold-Mende, Laura Heikaus, Felix Sahm, Wolfgang Wick, Andreas von Deimling and David E. Reuss
  **Copy number alterations profoundly impact the proteome of IDH-mutant Gliomas. Manuscript in preparation.**

## Oral presentations:

- Felix M. *et al.;* Annual Meeting of the German Pharmaceutical Society-DPHG, Heidelberg, Germany (2019)
- Felix M. *et al.;* DKFZ International PhD Program PhD-Retreat, Weil der Stadt, Germany (2019)
- Felix M. *et al.;* Bruker Virtual User Meeting for Mass Spectrometry (2021)

## Poster presentations:

- Felix M. *et al.;* DIA/SWATH Course, ETH, Zurich, Switzerland (2018)
- Felix M. *et al.;* DKFZ Phd Poster Presentation, Heidelberg, Germany (2019)
- Felix M. *et a*l.; US-HUPO 14$^{th}$ annual Conference (2021)

## Scientific peer reviewed articles beyond the scope of this thesis:

- Feller C, **Felix M**, Weiss T, Herold-Mende C, Zhang F, Kockmann T, Sahm F, Aebersold R, von Deimling A, Reuss DE. **Histone epiproteomic profiling distinguishes oligodendroglioma, IDH-mutant and 1p/19q co-deleted from IDH-mutant astrocytoma and reveals less tri-methylation of H3K27 in oligodendrogliomas. Acta Neuropathol**. 2020 Jan;139(1):211-213.

- Schmid D, Warnken U, Latzer P, Hoffmann DC, Roth J, Kutschmann S, Jaschonek H, Rübmann P, Foltyn M, Vollmuth P, Winkler F, Seliger C, **Felix M**, Sahm F, Haas J, Reuss D, Bendszus M, Wildemann B, von Deimling A, Wick W, Kessler T. **Diagnostic biomarkers from proteomic characterization of cerebrospinal fluid in patients with brain malignancies.**
  **J Neurochem**. 2021 Mar 18.doi: 10.1111/jnc.1535

- Abigail K. Suwala, **Marius Felix**, Dennis Friedel, Damian Stichel, Daniel Schrimpf, Ekkehard Hewer, Leonille Schweizer, Hildegard Dohmen, Ute Pohl, Daniel Erny, Christian Koelsche, Clinton Turner, Koray Özduman, Takahiro Ono, Marco Prinz, Till Acker, Christel Herold-Mende, Wolfgang Wick, David Capper, Pieter Wesseling, Felix Sahm, Andreas von Deimling, Christian Hartmann, David E. Reuss. **Oligosarcoma, IDH-mutant is a distinct aggressive type. Acta Neuropathol. Manuscript in revision.**

# Acknowledgement

At this point I would like to express my deepest gratitude to all who contributed to this work and made my PhD an unforgettable experience within and outside the lab.
First of all, I would like to thank Prof. Andreas von Deimling for giving me the opportunity to work on such an ambitious and exciting field in medical science within his lab with such an amazing team of various expertise.
Next, I would like to thank PD Dr. David Reuss for being an outstanding supervisor and mentor. Thank you for guiding me with high competence and experience throughout my PhD. Thank you for the fruitful discussions about brain tumor biology and diagnostics and for sharing your enthusiasm in cooking and your almost weekly anecdotes on ambitious and intriguing kitchen projects.
I would also like to thank Prof. Magnus von Knebel Doeberitz and Dr. Christiane Opitz for being my first and second referees of my thesis and for participating as members of my thesis advisory committee. Thank you for the valuable discussions and the constructive feedback during our TAC-meetings.
I also want to thank Prof. Jürgen Kopitz for his input as a member of my thesis advisory committee and for giving me guidance at the end phase of my PhD.
Next, I would like to thank Prof. Walter Haefeli for being chairman as well as Dr. Darjus Tschaharganeh for kindly accepting participation in my PhD examination committee.
Special thanks go to Dr. Uwe Warnken for supporting me in establishing the fresh frozen and FFPE protocols and keeping the mass spectrometers up and running at the DKFZ core facility while I was dealing with the sample preparations and bioinformatical analyses in the lab.
I would also like to thank Dr. Ashok Kumar Jayavelu who gave me constructive feedback on my datasets and the opportunity to measure my samples at the Max Planck Institute for Biochemistry in Martinsried.
Additionally, I would like to thank Dr. Annekathrin Reinhardt for microscopically examining the numerous amounts of slides for our dual biomarker approach and the constructive feedback which greatly improved the outcome of the manuscript.
I would also like to thank our bioinformaticians Dr. Daniel Schrimpf, Dr. Damian Stichel and Dennis Friedel in building up the data infrastructure needed for

proteomic analyses in the neuropathology and giving helpful feedback and support in data statistics and visualization. Especially I would like to thank Dennis Friedel for supporting me in generating various plots for our manuscripts and my thesis.
I would also thank the routine diagnostic lab of the neuropathology, particularly Viktoria Zeller, Ulrike Vogel and Sabrina Sprengart for preparing the countless numbers of slides and immunostains throughout my PhD.
Special thanks go as well to Jana Reuss and Christian Hagenlocher for being such great office buddies and supporting me in various aspects in the lab. The laughs and non-work-related discussions we had were truly refreshing and definitely helped cope with the struggles in the lab. .
I would also like to thank everyone in the neuropathology for creating such a welcoming atmosphere and supporting me throughout my experiments.
In addition, I would like to thank my two trainees Madelaine Stoll and Marlene Hoffarth, who were a great support during my PhD and helped improve various steps of the sample preparation protocols.
I would also like to thank all my PhD friends at the DKFZ that I met along the way, especially Jessica Devant, Monica Chandra, Anastasia Gkeka, Ines Gräßer, Laura Hartmann, and Sergej Lebedev for reminding me that the PhD-Life goes also beyond the Lab.
Many thanks go also to my friends in Tübingen, Dominic Gross, Marcel Dammann, Juliander Reiner and Luisa Kubin for all the great times we shared and the countless evenings we enjoyed together over bottles of wine and delicious food. Thank you for being a safe harbor during tough times.
Last but not least I would like to thank my parents and my sister for their continuing support and guidance along the way. Thank you for encouraging me throughout my studies and to follow my passion in science. In particular I would like to thank my father who inspired me to do my PhD in proteomics and for proof-reading the dissertation.

*To my family*

# Table of Contents

# List of Abbreviations

| | |
|---|---|
| ATRX | alpha thalassemia/mental retardation syndrome X-linked chromatin remodeler |
| CNN3 | calponin-3 |
| CNV | copy number variation |
| CPR | chromosomal protein ratio |
| DAP | differentially abundant proteins |
| DNA | deoxyribonucleic acid |
| EtOH | ethanol |
| FDR | false discovery rate |
| FF | fresh frozen |
| FFPE | formalin fixated paraffin embedded |
| FISH | fluorescence *in situ* hybridization |
| HE | hematoxylin and eosin |
| HIP1R | huntingtin-interacting protein 1-related protein |
| IHC | immunohistochemistry |
| h | hour(s) |
| HPLC | high performance liquid chromatography |
| IDH | isocitrate dehydrogenase |
| iRT | indexed retention time |
| LC | liquid chromatography |
| LFQ | label free quantification |
| MAD | median absolute deviation |
| min | minute(s) |
| MS | mass spectrometry |
| m/z | mass-to-charge ratio |
| PASEF | parallel accumulation serial fragmentation |
| PSM | peptide spectrum match |
| PTM | post translational modification |
| RNA | ribonucleic acid |

| | |
|---|---|
| ROS | reactive oxygen species |
| RPM | revolutions per min |
| RP-LC | reversed phase liquid chromatography |
| sec | second(s) |
| SYNM | synemin |
| TNC | tenascin |
| TOF | time-of-flight |
| VIM | vimentin |
| WHO | World Health Organization |
| WT | wildtype |

# Summary

The 2016 WHO Classification of Tumors of the Central Nervous System separates IDH-mutant gliomas into two distinct subtypes depending on the preservation or deletion of the chromosome arms of 1p and 19q termed “astrocytomas, IDH-mutant” and “oligodendrogliomas, IDH-mutant and 1p/19q-codeleted”. Currently, the assessment of the chromosomal copy number profile of 1p and 19q are essential to distinguish between the two tumor entities. An additional challenge is the WHO grading of astrocytomas which shows low correlation with survival. Biomarkers for facilitating the differential diagnosis of IDH-mutant glioma are of high clinical relevance. Therefore, there is a need for new markers and screening methods to effectively detect tumor malignancy and predict prognosis. Over the past decades mass spectrometry has become the method of choice for proteomic investigations in medical scientific areas. This study integrates a proteomic workflow into a clinical environment and explores the proteome of IDH-mutant gliomas with the aim of finding novel diagnostic and prognostic biomarkers for astrocytoma and oligodendroglioma.

For this, sample preparation techniques and protocols for fresh frozen as well as formalin-fixed-paraffin-embedded tissues were established and tumor samples analyzed using different mass spectrometry platforms. A dual surrogate biomarker for 1p/19q status in IDH-mutant gliomas was discovered which could simplify the diagnosis of these tumors, making it less dependent on elaborative genetic analyses. Furthermore, it was observed that protein abundances correlate in sum with copy number of their respective chromosomes in mutant gliomas. Potential prognostic biomarkers were found which could not only improve the grading of IDH-mutant astrocytomas, but also help us further understand how these tumors overcome oxidative stress.

Taken together, this study shows the potential impact of proteomic pipelines in a clinical environment and how it can complement already existing diagnostic infrastructures in the neuropathology. Importantly, it shows how a novel technology can help reveal novel potential biomarkers, which could be used by institutions where no genetic or proteomic analyses are available.

# Zusammenfassung

Die WHO-Klassifikation von Tumoren des Zentralnervensystems von 2016 klassifiziert IDH-mutierte Gliome in „Astrozytome, IDH-mutiert“ und „Oligodendrogliome IDH-mutiert und 1p/19q ko-deletiert“. Zur Unterscheidung der beiden Entitäten ist die Untersuchung des chromosomalen Kopienzahlprofils von 1p und 19q essenziell. Eine zusätzliche Herausforderung ist die WHO-Gradierung von Astrozytomen, welche eine geringe Korrelation mit dem Überleben zeigt. Daher besteht ein Bedarf an neuen Markern und Screening-Methoden, um die Bösartigkeit von Tumoren wirksam zu erkennen und die Prognose vorherzusagen. Die Massenspektrometrie hat sich zur Methode der Wahl für proteomische Untersuchungen im medizinisch-wissenschaftlichen Bereich entwickelt. Diese Studie integriert die Proteomik in eine klinische Umgebung und untersucht das Proteom von IDH-mutierten Gliomen mit dem Ziel, neue diagnostische und prognostische Biomarker für Astrozytome und Oligodendrogliome zu finden.

Dazu wurden Probenvorbereitungstechniken und Protokolle für frisch-gefrorene, sowie Formalin-fixierte Paraffin-eingebettete Gewebe etabliert und Tumorproben mit unterschiedlichen Massenspektrometrie-Plattformen analysiert. Es wurde ein dualer Surrogat-Biomarker für den 1p/19q-Status bei IDH-mutierten Gliomen identifiziert, der in Kombination mit ATRX die Diagnose dieser Tumoren vereinfachen und weniger abhängig von den genetischen Analysen machen könnte. Darüber hinaus wurde festgestellt, dass die Proteinabundanz in Summe mit der Kopiezahl ihrer jeweiligen Chromosomen in IDH-mutanten und Wildtyp Gliomen korreliert. Zusätzlich wurden neue potenzielle prognostische Biomarker gefunden, die die WHO Gradierung von IDH-mutierten Astrozytomen verbessern könnte, aber auch helfen zu verstehen, wie diese Tumoren oxidativen Stress überwinden.

Zusammenfassend zeigt diese Arbeit die zentrale Bedeutung der Proteomik in einem klinischen Umfeld und wie diese auch bereits bestehende diagnostische Infrastrukturen in der Neuropathologie ergänzen kann. Es zeigt vor allem wie eine neuartige Technologie dazu beitragen kann, neue potenzielle Biomarker zu entdecken, die von Einrichtungen genutzt werden könnten, die keine genetischen oder proteomischen Infrastruktur zur Verfügung haben.

# 1. Introduction

## 1.1 Neuropathology

It is well known that biologically different tumors have distinctive clinical characteristics. By characterizing these tumors, a prognosis of the clinical behavior can be made which is crucial for optimal therapeutic decision making. For tumors of the central nervous system this began in the beginning of 19$^{th}$ century, where gross morphological observations of autopsy material without microscopes was carried out. Later, Rudolf Virchow made first comprehensive histomorphological investigations identifying strong histological differences in brain tumors, pathing the way of how these tumors could later on be categorized [3]. Virchow's and his contemporary's descriptions were later on refined by Harvey Cushing and Percival Bailey. Their observations resulted in the foundation on which the first edition of WHO classification for brain tumors was built on [4].

The introduction of commercially available immune-histochemical markers led to further developments in the phenotypical classification. Tumors could be classified by their differentiation, activity in proliferation and cell growth. Additionally, over the years newer methods in molecular biology brought more in-depth data. With the development of genetic and epigenetic investigations, the biological understanding of brain tumors sustainably changed. Today, genetic data can be used as a molecular classification system that is independent from the existing phenotypic classification workflow. Both methods usefully combined resulted in the implementation of genetic parameters in the 2016 Classification of Tumors of the Central Nervous System [5]. Tumor entities are now more biologically homogenous and narrowly defined leading to higher diagnostic accuracy and possibly clinical treatment efficiency [1, 6].

### 1.1.1 Diffuse Astrocytic and Oligodendroglial Tumors

Diffuse astrocytic and oligodendroglial tumors are a heterogenous group of primary brain tumors that all show pronounced infiltrative growth. In this group, the WHO classification of 2016 discriminates astrocytomas, oligodendrogliomas WHO Grades II and III and additionally glioblastomas as WHO grade IV. The WHO grading system is based on histological features such as necrosis, mitotic activity,

cellularity and vascular proliferation. Through tumor progression a WHO grade II or III tumor can evolve to a WHO grade III or IV. If an astrocytoma with WHO grade II or III becomes WHO grade IV, it is defined as a secondary tumor with the diagnosis glioblastoma WHO grade IV [5].

The discovery of the IDH1-mutation and the less common IDH2-mutations, which are often found in oligodendrogliomas as well as astrocytomas and secondary glioblastomas, brought a major breakthrough in understanding diffuse gliomas [7, 8]. Further investigations showed that gliomas with an IDH-mutation have a better clinical course as IDH-WT gliomas with the same histology [9]. Whole genome analyses from these IDH-mutant gliomas revealed precisely separated groups that are roughly congruent with the histological categorization of astrocytomas and oligodendrogliomas.

In addition to the IDH mutation, oligodendroglial differentiated tumors have a common loss of the chromosome arms 1p and 19q (1p/19q co-deletion) which show a favorable clinical course [10, 11]. A more thorough investigation of this phenomenon revealed that only tumors with a complete loss of both arms show a favorable clinical course. In contrast, if an incomplete loss is observed the clinical course is strikingly worse. There are assumptions that the cause for the 1p/19q co-deletion is an unbalanced t (1; 19) (q10; p10) translocation [12, 13]. Studies have also shown that a complete loss of the chromosome arms 1p and 19q are linked to *TERT*-promotor mutations [14, 15].

In comparison to oligodendrogliomas, astrocytomas show mutations of *ATRX* and *TP53* but remain their wild type 1p 19q status. Although the mutation of *ATRX* is poorly understood, there are reports that it may correlate with alternative lengthening of telomeres (ALT) which leads to unlimited cellular growth [16, 17].

Considering the histologic group of oligoastrocytomas, a molecular equivalent was not found. Instead, these tumors either corresponded molecularly to the astrocytoma or oligodendroglioma group. This questions the existence of this entity entirely [18].

In summary all of these findings were later on implemented in the 2016 WHO classification system either using newly developed immunohistochemical markers or state-of-the-art genome sequencing.

## 1.1.2 The Algorithm of Classification of Diffuse Glioma

For decades the classification of brain tumors consisted of morphological investigations of tissue sections using light microscopy. Tumors with similar histological characteristics would be classified in to tumor entities using hematoxylin-eosin and immunohistochemically stained sections for feature analysis. With the publication of the 2016 WHO Classification of Tumors of the Central Nervous System, a conceptually new dimension in diagnostics was introduced into the classification of brain tumors. The development of genetic parameters to classify brain tumors gave the neuropathologist an additional tool to distinguish tumors in a more precise and objective matter [5]. Now, diagnosing a tumor requires next to a histological initial diagnosis an additional genetic evaluation which results in the final integrated diagnosis. During this procedure the initial diagnosis can deviate from the final integrative diagnosis which underlines the importance of the additional genetic parameters for characterization as can be seen in Figure 1 [1].

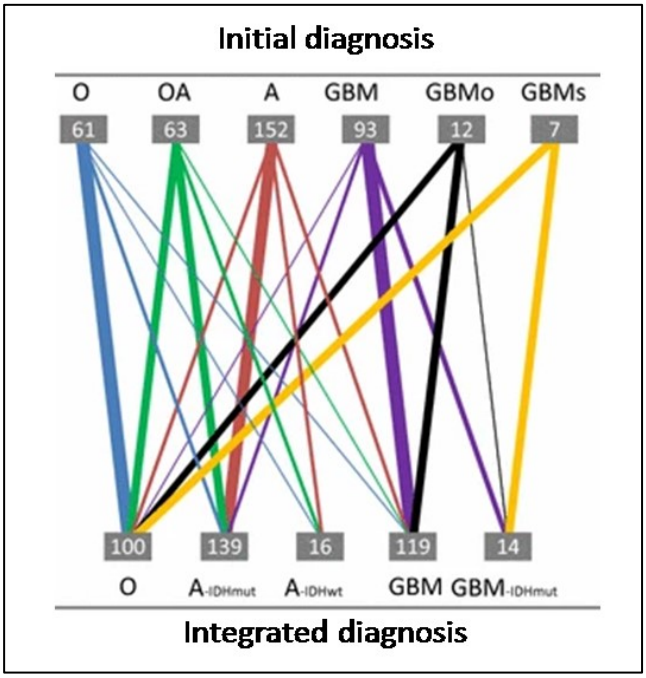


**Figure 1: Shift of diagnosis between initial and integrated diagnosis from 405 patients.** The widths of the bars represent the number of patients remaining or shifting to a specific diagnosis. **A** astrocytoma, **OA** oligoastrocytoma, **O** oligodendroglioma, **GBM** glioblastoma, **GBMo** glioblastoma with oligodendroglial component, **GBMs** secondary glioblastoma. *Adapted from [1]*

A simplified version of the classification algorithm of diffuse gliomas is depicted in Figure 2. In practice, to get an initial overview of the patient sample, histological examinations with HE-stained tissue sections are carried out. Here, the neuropathologist can already make a first assumption of what kind of tumor it can be. To verify the presumption, additional investigations using IHC are carried out. In the case of a diffuse gliomas, IHC investigation begins with the inspection of the *IDH1* mutation R132H which is present in 90% of all IDH-mutant tumors. The antibody used for this procedure was developed in the department of neuropathology in Heidelberg [19, 20]. If a positive staining is acquired, depending on identified morphological features, further investigations of *ATRX* status via immunostaining or genetic analysis of the 1p/19q status by FISH or DNA methylation profiling is needed. If the staining of ATRX is negative, the tumor is most possibly an astrocytoma, IDH-mutant. If the sample shows a co-deletion of the chromosome arms 1p and 19q, the tumor is an oligodendroglioma, IDH-mutant and 1p19q-codeleted. If the immunostaining of IDH1 R132H and ATRX are negative, additional IDH1/2 gene hotspot sequencing should be done to check for possible rare mutations of *IDH1* such as R132C and R132S or all *IDH2* R140 and R172 mutations. If a mutation is detected, the tumor is most likely an astrocytoma, IDH-mutant.

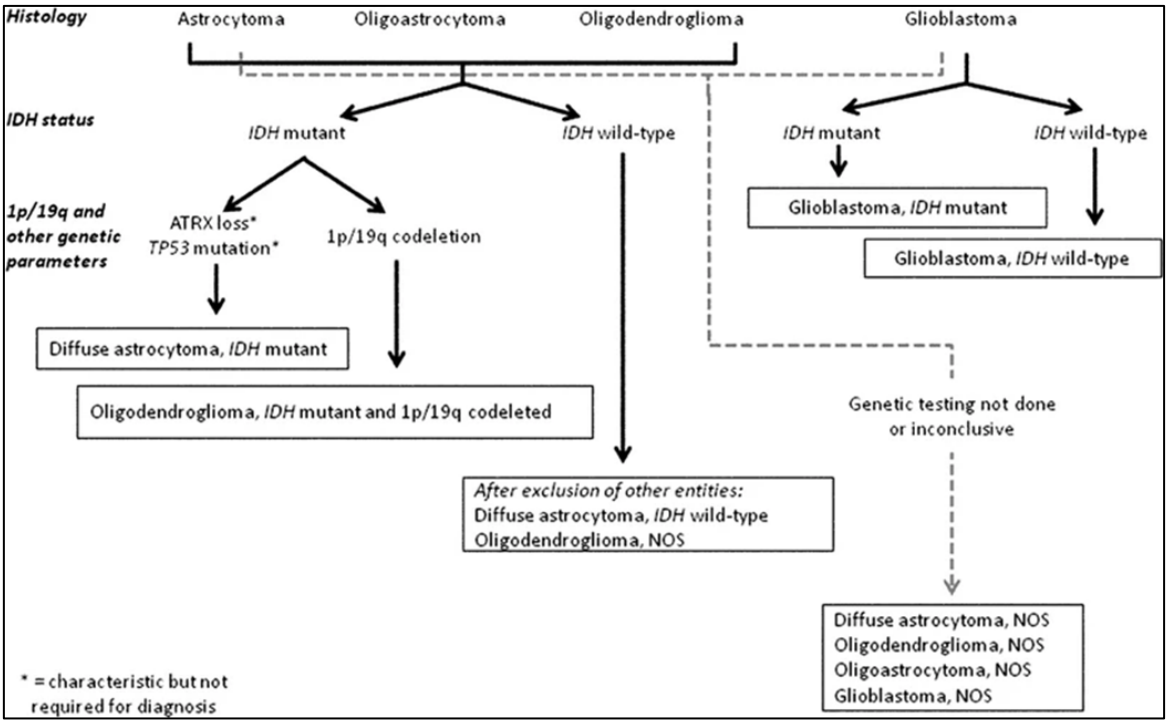


**Figure 2: Simplified algorithm for the classification of diffuse gliomas.** The diagnostic "flow" does not necessarily start with histology and proceed with molecular genetic features, since molecular features can overweigh histologic characteristics in achieving the "integrated Diagnosis". *Adapted from [5]*

### 1.1.3 Grading Criteria for IDH-Mutant Astrocytomas

As mentioned in chapter 1.1.1 IDH-mutant astrocytomas can be further distinguished as WHO grade II diffuse astrocytomas, WHO grade III anaplastic astrocytomas, and WHO grade IV glioblastomas. The grading of these tumors is currently performed by investigating morphological features such as microvascular proliferation, necrosis, mitotic activity and anaplasia [21]. Specifically, when trying to distinguish between WHO grade II and III, the microscopic investigation of mitotic activity plays a central role. If for example 2-3 mitoses are found on a tumor slice, the diagnose is an anaplastic astrocytoma. If it is lower, then it is most likely a diffuse astrocytoma. However, if the tumor region is rather small as for example in a biopsy punch, even one mitosis is enough to diagnose it as an anaplastic astrocytoma. In general, the number of mitoses needed to distinguish between the diffuse and anaplastic astrocytoma is not strictly defined which brings additional challenges in

giving a precise diagnosis.

These grading criteria were developed during a period where IDH-mutant gliomas were not discovered yet, which resulted in a “good” correlation with prognosis at the time. However, with the discovery of the *IDH* mutation the legacy classification using mitotic activity has become less accurate [22, 23]. As mentioned above, IDH-mutated gliomas have a better overall survival compared to IDH-WT but they also show less necrosis and mitotic activity. This means that earlier lower graded gliomas were in fact IDH-mutant gliomas, which was the reason why the grading correlated so well with prognosis. Now, when investigating IDH-mutant gliomas, mitotic activity does not correlate well with prognosis. Although the grading using mitosis for IDH-mutant gliomas will remain in the next issue of the WHO classification, the grading is slowly being replaced by novel diagnostic methods and biomarkers. Observations of possible genetic alterations such as deletions of *CDKN2A/B* were recently implemented into the grading of IDH-mutant gliomas [24-26]. Although the introduction of genetic parameters improved the grading to some extent, there is still much room for improvement. Presence of *CDKNA/B* deletions corresponds with a clinical behavior of a WHO grade IV tumor but does not play a major role when diagnosing between WHO grade II and III. *CDK4*, *PDGFRA* and *MYCN* amplifications may indicate a worse prognosis but more studies need to be performed to verify these findings [21]. Here, proteomics as a novel method for diagnostics, could reveal biomarkers that correlate with survival better than current classification methods.

### 1.1.4 The IDH Mutation and the Resulting Metabolic Shift

IDH is a mitochondrial enzyme which oxidizes isocitrate to α-ketoglutarate and is part of the TCA cycle. The TCA cycle is responsible for producing NADPH, which is needed for oxidative phosphorylation in order to create ATP via the ATP synthase but also to regenerate glutathione for $H_2O_2$ detoxification. When *IDH* is mutated the resulting enzyme further reduces α-ketoglutarate to the oncometabolite D-2-hydroxyglutarate under NADPH consumption which results in a dramatic shift in various metabolic processes. D-2-hydroxyglutarate is an inhibitor of the DNA-demethylase which leads to a higher prevalence of hypermethylated promotor

regions. Additionally, by consuming NADPH for D-2-hydroxyglutarate synthesis, less reductive species are present in order to regenerate glutathione. Consequently, the incapability of ROS detoxification leads to oxidative stress which is typical for IDH-mutant glioma [27].

## 1.2 Proteomics

In 1958 a key publication from Francis Crick named “The central dogma of molecular biology” laid the cornerstone of understanding of how biological systems express their genetic code [28]. Through transcription of DNA as a structural blueprint, corresponding RNA strands are generated which are then further translated into proteins. The entire set of proteins in a biological system is termed as the proteome. During the biosynthesis of proteins, various processes modulate the flux of information from the genetic code to the resulting protein with each step increasing the complexity of possible molecules that can be generated by a single gene. Alternative splicing of mRNA for example leads to various proteins, so called *isoforms* that originate from the same gene. The complexity of the proteome increases even more dramatically when additional post translational modifications are added to the equation, resulting in a complexity increase of almost three orders of magnitude. Thus, it is understandable that the resulting millions of *proteoforms* originating from the different splicing variants and modifications easily exceed the currently estimated ~20,000 protein coding genes of the human genome [29]. Proteomics addresses this complexity by large scale investigations of protein isoforms, protein localization, protein-protein interactions, post-translational modifications, protein abundance and turnover (Figure 3). Over the past decades mass spectrometry has risen to become the next potential

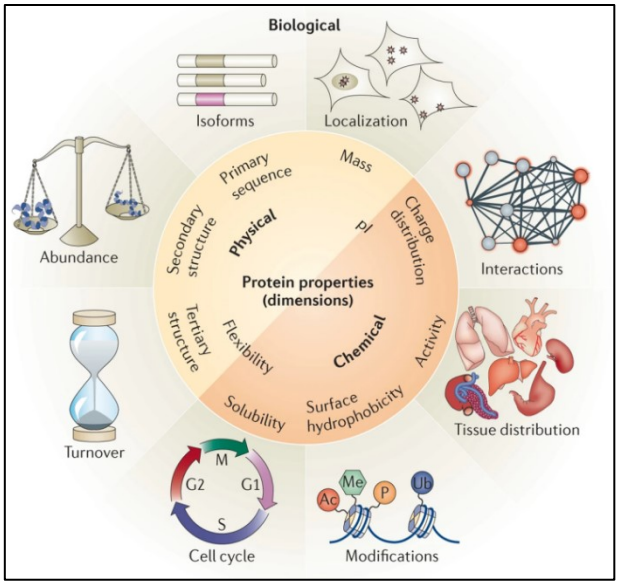


**Figure 3: Depth of information in proteomics.** Biological aspects that can be addressed using proteomic analyses that are otherwise not possible with genomic or transcriptomic approaches. *Adapted from [2]*

technology to enable the analysis of the proteome in great depth and could help investigate questions which have been so far untouched by genomic and transcriptomic analyses [30].

### 1.2.1 Bottom-up Proteomics

In comparison to *top-down* proteomics where intact proteins are investigated, *bottom-up* approaches as depicted in Figure 4, concentrate on the identification and quantification of peptides. In principle, proteins are extracted from any kind of biological origin such as cells, tissue or body fluids and are later on digested by proteases into a highly complex mixture of various peptides (see Figure 4A). Depending on the experiment, additional fractionation steps such as high pH-fractionation or gel-electrophoresis can be used to find lower abundant peptides in a sample, resulting into a deeper investigation of the proteome. Protein digestion is carried out using amino acid specific proteases. The most frequently used protease is the serine protease trypsin which cleaves at the carboxyl side of the basic amino acids lysine and arginine, except if a C-terminally bounded proline is present [31]. The frequency of lysine and arginine in the human proteome is at around five percent each; meaning that one amino acid out of ten in a peptide sequence can be one of the two amino acids [32]. This results in peptide lengths of approximately 10 amino acids and basic residues at the C-terminus. Both characteristics bring strong advantages for the suitability in mass spectrometry in terms of ionization and “flyability” [33]. Additionally to trypsin, the combination with other proteases such as LysC, GluC and chymotrypsin can help increase the sequence coverage [34].
Due to the high complexity of the peptide mixtures a mass spectrometer alone would not be able to detect peptides in a proper fashion. In order to decrease sample complexity, proteomic approaches are carried out on a HPLC that is coupled online to a mass spectrometer (Figure 4b). After the spectral data has been generated, the peptides are theoretically reconstructed to their protein of origin by a database search engine (Figure 4c).

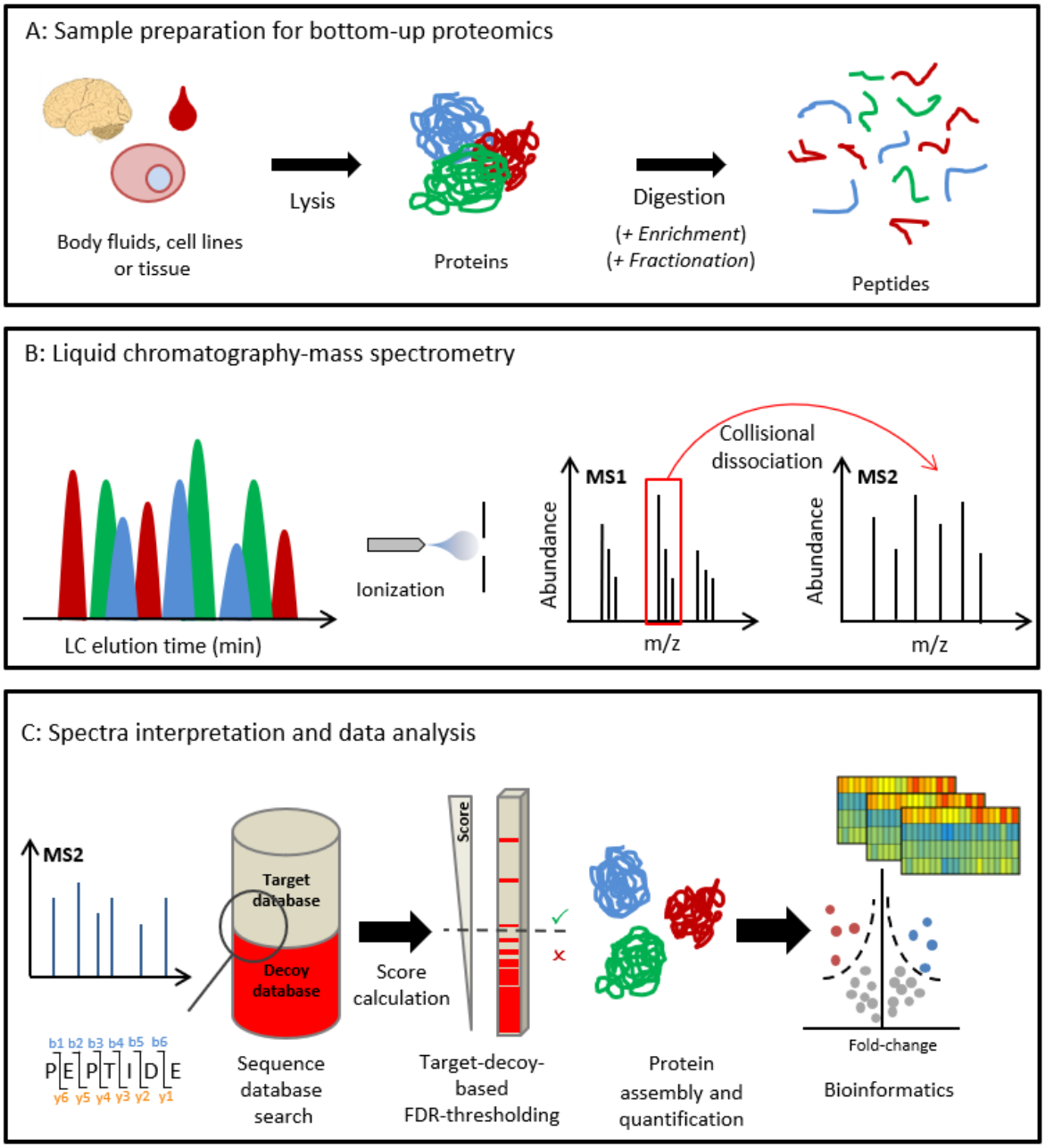


**Figure 4: Experimental workflow for a typical bottom-up proteomic analysis. A:** Sample preparation for bottom-up proteomics: Acquired tissue, cells or body fluid are lysed and resulting proteins digested to peptides. **B:** Liquid chromatography-mass spectrometry: Resulting peptides are separated by HPLC and then injected *on line* into the mass spectrometer where ms1 and ms2 spectra are acquired. **C:** Spectra interpretation and data analysis: Full scan data containing peptide mass, intensity, PTM, isotopes and fragment scans with information of b- and y-ions are needed for the identification of peptides via database searching. A score threshold is defined that leads to a percentage of false positives that remain in the hit population. Peptide identifications are reassembled to proteins and quantified followed by data by data interpretation using bioinformatics. *Adapted from [35]*

#### 1.2.1.1 Sample Preparation of FFPE Samples

For over a century, patient samples obtained from biopsies and resections have been preserved using formalin fixation and paraffin-embedding [36]. It is routinely used in the tissue banks and pathological institutes due to its capabilities for long term storage and accessibility for downstream analysis such as IHC and genome sequencing [37, 38]. FFPE is an economical choice for preservation since it can be stored at room temperature without the need of freezers or liquid nitrogen tanks. There have been major trends in the field proteomics to analyze these types of samples since in comparison to FF tissue, FFPE samples are highly abundant and well documented concerning clinical information such as diagnosis, genetic status, medication and overall survival. Although these types of samples have great potential in proteomics to discover new findings, there are major difficulties in analyzing these samples.

As formalin comes in contact with fresh tissue, it penetrates the in ethanol dehydrated tissue and causes uncontrollable intra and intermolecular crosslinks between peptide chains. Specifically, formalin reacts with the side chain amino group of lysine resulting in the formation of methylene bridges [39]. Additionally, crosslinking between aminomethylol groups and phenol, indole and imidazole containing side chains are possible in form of a Mannich reaction. Consequently, this means that the amino acids arginine, tyrosine, asparagine, histidine, glutamine and serine are additionally affected by formaldehyde. In summary, the crosslinking can modify α-helices and β-sheets of the secondary structure and therefore alter the tertiary structure of fixed proteins and most possibly link nearby proteins to one another [40].

The main challenge for analyzing these samples in proteomics is to resolve the crosslinks formed during the fixation process. Keeping the crosslinks would lead to lower identification rates of peptides due to the mass shift resulting from the various types of crosslinks and the connection partners to it. Many theories and preparation methods have been developed to tackle this issue with all of them having more or less the following components in common: Heat, high or low pH, presence of detergent or chaotropic substances and high pressure [39, 41-43]. Summarized, all protocols deal with crosslinks by trying to force hydrolysis. The principle used is

based on common workflows that have already been established since the existence of immunostaining. Here, so called antigen retrieval is carried out in order to make the binding of the primary antibody to an antigen on the tissue section surface possible. This is always either carried out in acidic (pH ~6) or as basic conditions (pH ~9) with the combination of heat (~95°C).
Newer methods for FFPE proteomics call for the use of pressure to improve cleavage efficiency. Although the mechanistic is not completely understood, it is assumed that high pressure forces water into the lipophilic pockets of the protein resulting in protein unfolding and better hydrolysis [44, 45]. This approach is also used for protein digestion since proteases catalyze hydrolysis, accelerating the digestion significantly [42, 46].

#### 1.2.1.2 Identification of Proteins

As mentioned in chapter 1.2.1, peptide sequences can be inferred by analyzing fragment ion spectra which have been generated by LC-MS approaches. What used to be done by *de novo* sequencing, is nowadays carried out using algorithms for database searching where measured fragment ion spectra are compared to calculated ones [47]. In short the algorithm generates theoretical ion-spectra of *in silico* digested proteins that derive from protein sequence databases such as the most frequently used UniProt Knowledgebase [48]. Before searching through raw data, one must define specific criteria for the search such as i) proteolytic enzyme, ii) number of missed cleavages, iii) fixed and variable post translational modifications, iv) mass tolerance and FDR cutoff. A score is calculated for every spectrum that is matched to a calculated spectrum of a peptide and the best scoring spectrum is then reported as a PSM. To keep in mind is that the peptide centric nature of bottom up proteomics defines the identification of a protein solely by the presents of one or multiple unique peptide sequences that it corresponds to. However, this sequence can also belong to several other isoforms (see chapter 1.2) leading to ambiguous identifications of proteins also known as protein inference [49]. This could lead to difficulties in protein quantification and data interpretation. Therefore, a summary of the different isoforms is reported in form of a protein group. Another aspect to keep in mind is that not every PSM is a true spectrum match. For

large scale proteomics it is required to use automated scoring algorithms that can define false positive spectrum matches and distinguish false and true peptide assignments. The already mentioned database score gives information of how well the measured spectrum matches its theoretical counterpart but doesn't give information if it is a false positive or not. For this, a target-decoy search strategy is used to estimate the number of false positives that are present in the correctly identified cohort (Figure 4C). Briefly, in addition to the true protein database a so-called decoy database is searched for possible spectrum matches. The decoy database contains proteins sequences of the original protein database that have been scrambled, shuffled or reversed. Peptides identified by the decoy database are defined as false positive hits and enable the calculation of the FDR:

$$FDR = \frac{FP}{TP + FP}$$

FP: false positive identification TP: true positive identification

By defining a certain FDR threshold an error probability is given for each identified peptide. In general, the FDR filtering is applied on the peptide and protein level [50]. After FDR thresholding, identified peptides are reassembled to their corresponding proteins and quantified.

#### 1.2.1.3 Quantification of Proteins

Over the years the main goal in proteomics has shifted from identifying the most proteins as possible in a sample to precisely quantifying proteins in a large-scale matter [51, 52]. In summary, quantification of proteins can be carried out using two main approaches called relative und absolute quantification. The first approach consists of a relative comparison of peptide abundances of two or more samples through protein fold-change whereas absolute quantification determines the exact concentration of a protein or peptide of interest. Quantification in mass spectrometry assumes that abundance levels of mass peaks are direct proportional to the concentration of the peptide. However, it is important to keep in mind that this

assumption can only be considered when comparing the same peptides signals between samples due to the different physiochemical characteristics and biases introduced by sample preparation.
Currently, one of the most used strategies in relative quantitative proteomics is label-free quantification where peptide quantities are compared over different mass spectrometry measurements. The accuracy of this approach has greatly improved over the years due to continuous improvements in computational methodology and software optimization [53, 54]. Nevertheless, samples are only combined *in silico* which makes sample preparation in highly reproducible and simplified fashion essential to maintain the comparability of different samples over long periods of time. Additionally experimental designs with corresponding protocols must be well worked out to avoid batch effects or other confounding factors [55].

## 1.3 Proteomic Approach using Mass Spectrometry

With the ESI ionization technique the usage of liquid chromatography and mass spectrometry in an *online* set up has become the gold standard for proteomic analysis also known as LC-MS [56, 57]. Although LC-MS devices have different techniques for separation and detection of biomolecules, the main principles stay the same and can be divided in to four different steps: i) a chromatographic device to separate the biomolecules from each other according to their physiochemical properties, ii) an ion source for ionization in to the gas phase of the LC-separated biomolecules, iii) mass analyzers which can separate biomolecules depending on their mass-to-charge ratio (m/z) and iv) a detector that registers the ion current for the given m/z value. The following chapters describe the different principles in more detail.

### 1.3.1 High Performance Liquid Chromatography

After preparing samples for global proteome analysis, one is left with a peptide mixture containing an enormous number of different peptides that have wide dynamic range. Due to the high complexity in these samples, there is a great need in lowering it before the sample can be analyzed using a mass spectrometer. High

performance liquid chromatography offers the best solution for this problem by separating proteolytic peptides depending on their physiochemical properties and abilities in interacting with a stationary and mobile phase. During chromatography, peptide samples are dissolved in the mobile phase which passes through a chromatographic column. The column contains a stationary phase which is made out of small particles (mostly silicate). As the peptides pass through the column they interact with functional groups that are bound to the stationary phase. The extent of the interaction with the stationary phase is dependent on their physiochemical properties. If the interaction of the peptide with the stationary phase is stronger than with mobile phase, retention occurs. For a more efficient separation and improved peak resolution, the composition of the mobile phase can be changed over time. This approach termed gradient elution leads to a sequential elution of peptides with distinct physiochemical properties, facilitating their detection in the mass spectrometer.

In proteomics most experiments are carried out using reversed phase liquid chromatography. Here, the stationary phase is commonly a reversed phased column, which means the silicate particles are chemically modified with hydrophobic elements such as long chains of carbon (mostly C18) on the surface. Using a reversed phased column and a hydrophilic mobile phase, peptides are separated by their hydrophobicity [58, 59]. Thanks to the high resolution, peak capacity and the compatibility of the volatile solutions with ESI, RP-LC is most often used for LC-MS experiments [60].

### 1.3.2 Tandem Mass Spectrometry

The main principle for the identification of a peptide is acquiring information about the mass of the referred peptide and its fragment ions injected in to the mass spectrometer (Figure 4B). Typically, a mass spectrometer measures an initial full scan (MS1) of intact peptide precursors to identify m/z values and intensities followed by the selection of one precursor for collision induced fragmentation and a second full scan of the fragments generated (MS2). One of the most used picking strategies in bottom-up proteomics is data-dependent acquisition (DDA) which is an automated strategy for picking precursors for fragmentation after the initial MS1

scan. For this, a predefined number of precursors are picked based on their abundance in the MS1 scan and subsequently fragmented and analyzed in the MS2 scan (schematic overview shown in Figure 5A). A major drawback with this strategy is the stochastic nature of the precursor isolation since only the strongest signals are selected for fragmentation leading to biased information in terms of identification and limitations in reproducibility [61]. A rather new and attractive alternative strategy is data independent acquisition (DIA) [62]. Here, all precursor ions within a defined m/z window are selected for fragmentation. During the acquisition, the selection window wanders upwards until the entire m/z range has been measured and starts then again from the beginning. This results in complex MS2 spectra that can only be further analyzed using prior information in form of spectral libraries and specialized software solutions [63, 64]. Since the peptide selection is independent from any previous scans the identification and quantification of proteins is more robust [65].

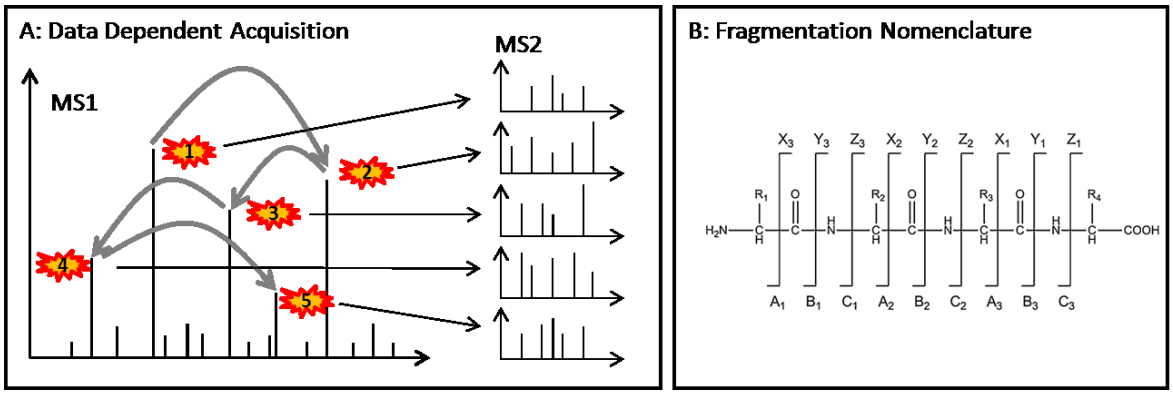


**Figure 5: Data dependent acquisition. A:** Schematic overview of a DDA scan cycle top five most intense precursors are selected for fragmentation and downstream MS2 sequencing. **B**: Fragmentation nomenclature suggested by Roepstorff and Fohlmann [66].

Various fragmentation methods have been developed in proteomics. The most commonly used for peptide fragmentation are collision-induced dissociation (CID) [67], high-energy collision dissociation (HCD) [68] and electron transfer dissociation (ETD) [69]. Both CID and HCD are based on the transfer of energy by collision with inert gas molecules (e.g. nitrogen) which cause peptide fragmentation on the back-bone structure producing mostly b- and y-ions (see Figure 5B). ETD is based on

generated anion radicals that donate electrons to the peptide cations leading to unstable radicals that immediately decay in to mostly c- and z-ions. ETD has the strong advantage of preserving labile post translational modifications such as phosphorylation which are usually cleaved by CID and HCD, making it more suitable for phosphoproteomic experiments.

## 1.4 Aim of the Thesis

With recent technological developments, proteomic mass spectrometry analyses have become the next potential tool to unravel unsolved questions in life sciences and diagnostics [30, 70-72]. Various sample preparation techniques for FFPE tissue have unlocked possibilities in discovering new findings that have been since now unreachable [73-75].

This study aims at the establishment and application of novel methodologies in developing a proteomic workflow for differential diagnostics of IDH-mutant glioma. The specific aims of this study are:

1. Establishing FF and FFPE sample preparation protocols for brain tumor tissue samples which show comparable results in the proteomic data.
2. Search and identification of new signatures and biomarkers for distinguishing between astrocytoma and oligodendroglioma and to identify links to already existing diagnostic biomarkers.
3. Development of an approach linking proteomic and genetic data with each other which can later on be used as a quality control procedure.
4. Identification of prognostic biomarkers which can potentially help improve the grading of astrocytoma.

# 2. Materials and Methods

# 2.1 Materials

## 2.1.1 Instrumentation

**Table 1: List of Instruments**

| **Liquid Chromatography** | |
|---|---|
| UltiMate 3000 HPLC ® | Thermo Scientific, Waltham, MA, USA |
| Easy nLC 1200 | Thermo Scientific, Waltham, MA, USA |
| Trap Column: | |
| Acclaim™ PepMap™ 100 C18-LC-Column;<br>100Å, 5µm, 0,3x5mm | Product number: 160454<br>Thermo Scientific, Waltham, MA, USA |
| PepMap™ trap holder 5mm 30µm nanoViper | Product number: 164649<br>Thermo Scientific, Waltham, MA, USA |
| **Mass Spectrometers** | |
| Orbitrap Exploris™ 480 | Thermo Scientific, Waltham, MA, USA |
| timsTOF Pro | Bruker Daltonics, Bremen, Germany |

| **Other Instruments** | |
|---|---|
| Analytical Balance<br>Model: Discovery | OHAUS Corporation, Parsippany, NJ, USA |
| Centrifuge<br>Model: Hareus™ Fresco 21™ Centrifuge | Thermo Scientific, Waltham, MA, USA |
| Mini Centrifuge<br>Model: Spectrafuge™ B-181 | neoLab Migge GmbH, Heidelberg, Germany |
| Microplate reader<br>Model: Fluostar Omega | BMG LABTECH, Ortenberg, Germany |
| Probe sonicator<br>Model: Sonopuls HD 2070 | BANDELIN electronic, Berlin, Germany |
| Bath sonicator<br>Model: Sonorex Super RK 52 | BANDELIN electronic, Berlin, Germany |
| Barocycler<br>Model: Barocycler® NEP2320-45K Extreme | Pressure Biosciences Inc., Easton, MA, USA |
| Thermomixer<br>Model: Comfort | Eppendorf, Hamburg, Germany |
| Tissue Homogenizer<br>Model: Precellys 24 | Bertin Technologies SAS, Montigny Le Bretonneux, France |
| Vacuum concentrator<br>Model: Concentrator plus | Eppendorf, Hamburg, Germany |
| VENTANA BenchMark ULTRA | Roche Diagnostics, Risch-Rotkreuz, Switzerland |

| | |
|---|---|
| Vortexer<br>Model: RS-VF-10 | PHOENIX Instrument, Garbsen, Germany |

## 2.1.2 Software

**Table 2: List of Software**

| | |
|---|---|
| **Database search** | |
| MaxQuant Version 1.6.17.0 | MPI of Biochemistry, Martinsried, Germany |
| **Spectrum/chromatogram visualization and interpretation** | |
| Xcalibur | Thermo Scientific, Waltham, MA, USA |
| Freestyle | Thermo Scientific, Waltham, MA, USA |
| Compass Hystar 5.1 | Bruker Daltonics, Bremen, Germany |
| Compass DataAnalysis | Bruker Daltonics, Bremen, Germany |
| **Data interpretation and visualization** | |
| Perseus | MPI of Biochemistry, Martinsried, Germany |
| R | R Core Team |

## 2.1.3 Chemicals

**Table 3: List of Chemicals**

| **Substance**, *purity* | |
|---|---|
| ***Liquid*** | |
| 1,1,1,3,3,3-Hexafluoro-2-propanol, *≥99%* | Product No.: 105228-100G<br>Sigma-Aldrich, St. Louis, MO, USA |
| Acetonitrile, *LC/MS* | Product No.: 10489553<br>Fischer Scientific, Schwerte, Germany |
| Ethanol, *≤100%* | Product No.: 1117271000<br>Merck, Darmstadt, Germany |
| Ethanol, *≥99,8 %* | Product No.: 9065.2<br>Carl Roth, Karlsruhe, Germany |
| Eukitt® | Product No.: 01086540<br>O. Kindler, Freiburg im Breisgau, Germany |
| Methanol, *LC/MS* | Product No.: 10031094<br>Fischer Scientific, Schwerte, Germany |
| Water, *LC/MS* | Product No.: 10728098<br>Fischer Scientific, Schwerte, Germany |
| Xylene, *>98 %* | Product No.: 9713<br>Carl Roth, Karlsruhe, Germany |

***Acid/Base***

| | |
|---|---|
| Acetic acid, *glacial* | Product No.: 1000631000<br>Merck, Darmstadt, Germany |
| Acetic acid (Rotipuran) | Product No.: 3738.1<br>Carl Roth, Karlsruhe, Germany |
| Formic acid, *90-100%* | Product No.: 822254<br>Merck, Darmstadt, Germany |
| Hydrochloric acid (1M) | Product No.: H1758<br>Sigma-Aldrich, St. Louis, MO, USA |
| Sodium hydroxide (1M) | Product No.: 35256<br>Honeywell, Morristown, NJ, USA |
| Trifluoroacetic acid, LC/MS | Product No.: T6508<br>Merck, Darmstadt, Germany |

***Solid/Powder***

| | |
|---|---|
| Ammonium bicarbonate | Product No.: 09830<br>Sigma-Aldrich, St. Louis, MO, USA |
| Citric acid monohydrate | Product No.: 3958.1<br>Carl Roth, Karlsruhe, Germany |
| Chloral hydrate | Product No.: K318.1<br>Carl Roth, Karlsruhe, Germany |
| Eosin | Product No.: 7089.2<br>Carl Roth, Karlsruhe, Germany |

| | |
|---|---|
| Hematoxylin | Product No.: 1.04302.0100<br>Merck, Darmstadt, Germany |
| Potassium aluminum sulfate | Product No.: 8896.1<br>Carl Roth, Karlsruhe, Germany |
| Sodium iodide | Product No.: 6525<br>Merck, Darmstadt, Germany |
| Trizma® hydrochloride, | Product No.: T3253<br>Sigma-Aldrich, St. Louis, MO, USA |

***Detergents/Reagents***

| | |
|---|---|
| Iodoacetamide | Product No.: I1149<br>Sigma-Aldrich, St. Louis, MO, USA |
| Halt™ Protease Inhibitor Cocktail (100X) | Product No.: 78430<br>Thermo Scientific, Waltham, MA, USA |
| Sodium dodecyl sulfate, ≥99% | Product No.: L4390<br>Sigma-Aldrich, St. Louis, MO, USA |
| Tris(2-carboxyethyl)phosphine | Product No.: C4706<br>Sigma-Aldrich, St. Louis, MO, USA |
| Trypsin/Lys-C Mix, *MS-Grade* | Product No.: V5073<br>Promega, Mannheim, Germany |

## 2.1.4 Consumables

**Table 4: List of Consumables**

| Consumables | |
|---|---|
| Pierce™BCA Protein Assay Kit | Product No.: 23225<br>Thermo Scientific, Waltham, MA, USA |
| Disposable biopsy punch, 1,5mm Ø | Product No.: BP-15f<br>Kai industries, Seki, Japan |
| iRT Kit | Ki-3002<br>Biognosys, Schlieren, Switzerland |
| Precellys lysing Kit | Product No.: P000918-LYSK0-A<br>Bertin Technologies SAS, Montigny Le Bretonneux, France |
| MicroCaps (50µl) | Product No.: MTWS-MC50-RK<br>Pressure Biosciences Inc., Easton, MA, USA |
| MicroCap Tool | Product No.: MTWS-CR-03<br>Pressure Biosciences Inc., Easton, MA, USA |
| MicroTubes | Product No.: MTWS-MT-RK<br>Pressure Biosciences Inc., Easton, MA, USA |
| Superfrost® Plus Adhesive slides | Product No.: 10149870<br>Thermo Scientific, Waltham, MA, USA |

### *Electrophoresis*

| | |
|---|---|
| Colloidal Blue Staining Kit | Product No.: LC6025<br>Thermo Scientific, Waltham, MA, USA |
| NuPAGE™ 4x LDS Sample Buffer | Product No.: NP0008 Thermo Scientific, Waltham, MA, USA |
| NuPAGE™ 10x Sample Reducing Agent | Product No.: NP0009 Thermo Scientific, Waltham, MA, USA |
| NuPAGE™ Antioxidant | Product No.: NP0005 Thermo Scientific, Waltham, MA, USA |
| NuPAGE™ 20x MOPS SDS Running Buffer | Product No.: NP0001<br>Thermo Scientific, Waltham, MA, USA |
| NuPAGE™ 4-12% Bis-Tris Protein Gels, 1.0 mm, 10-well | Product No.: NP0321BOX<br>Thermo Scientific, Waltham, MA, USA |
| PageRuler™ Prestained Protein Ladder, 10 to 180 kDa | Product No.: 26616<br>Thermo Scientific, Waltham, MA, USA |

### *Immunohistochemistry*

| | |
|---|---|
| Bluing Reagent | Product No.: 760-2037<br>Roche Diagnostics, Risch-Rotkreuz, Switzerland |
| Cell Conditioning 1 | Product No.: 950-124<br>Roche Diagnostics, Risch-Rotkreuz, Switzerland |

| | |
|---|---|
| Cell Conditioning 2 | Product No.: 950-123<br>Roche Diagnostics, Risch-Rotkreuz, Switzerland |
| Hematoxylin I | Product No.: 760-2021<br>Roche Diagnostics, Risch-Rotkreuz, Switzerland |
| Hematoxylin II | Product No.: 790-2208<br>Roche Diagnostics, Risch-Rotkreuz, Switzerland |
| OptiView Amplification Kit | Product No.: 760-099<br>Roche Diagnostics, Risch-Rotkreuz, Switzerland |
| OptiView Universal DAB Detection Kit | Product No.: 760-700<br>Roche Diagnostics, Risch-Rotkreuz, Switzerland |

### 2.1.5 Antibodies

**Table 5: List of Antibodies**

| | |
|---|---|
| VIM | Product No.: GA63061-2<br>Dako, Jena, Germany |
| HIP1R | Product No.: ab140608, Abcam, Cambridge, United Kingdom |
| ATRX | Product No.: BSB3296, BioSB, Santa Barbara, USA), |

## 2.1.6 Reagents for HE-staining

**Table 6: List of Reagents for HE-Staining**

| Chemical | Amount |
|---|---|
| **Hematoxylin Mayer's** | |
| Hematoxylin | 1 g |
| Aqua dest. | 1000 ml |
| Sodium iodide | 0.2 g |
| Potassium aluminium sulfate | 50 g |
| Mix well until fully dissolved then add: | |
| Chloral hydrate | 50 g |
| Citric acid | 1 g |
| **Eosin in alcohol** | |
| Eosin | 20 g |
| Alcohol 70% | 2000 ml |
| Acetic acid, glacial | 3-5 drops before use |

## 2.1.7 Buffers

All buffers are prepared with LC/MS grade water as a solvent in the appropriate volume

**Table 7: List of Buffers**

| Chemical | Final Concentration |
|---|---|
| **FFPE Tissue lysis buffer** | |
| Sodium dodecyl sulfate | 1% (w/v) |
| *tris*(hydroxymethyl)aminomethane hydrochloride | 300 mM<br>*Adjusted to pH 8.5* |
| **FF tissue lysis buffer** | |
| Sodium dodecyl sulfate | 1% (w/v) |
| Halt™ Protease Inhibitor Cocktail (100X) | 1X |
| EDTA (100X) | 1X |
| Ammonium bicarbonate | 100 mM |

## 2.2 Methods

### 2.2.1 Sample Acquisition for FF Tissue

FF samples were acquired after surgical removal of the tumor from the patient and cryopreserved at the division for Neurosurgical Research led by Prof. Christel Herold Mende until further use. The collection and processing as well as the data collection followed the local ethics regulations and approval (ethical vote number S-315/2019).

### 2.2.2 Formalin Fixation and Paraffin Embedding

*Procedures 2.2.2 and 2.2.3 were performed by the routine diagnostic lab of the Heidelberg Neuropathology. PD. Dr. David Reuss performed all HE-stain examinations*

After surgical removal, the tumor was placed in formalin for at least 6 h to ensure complete fixation of the tumor. Afterwards the tumor underwent a dehydration procedure where it was incubated in increasing concentrations of ethanol and xylene. Then, the tumor was placed in molten paraffin and formed into a solid formalin fixed and paraffin embedded (FFPE) tumor block was ready for further procedures.

### 2.2.3 FF Sample Collection

After, a trained neuropathologist confirmed the suitability of the FF tissue (see chapter 2.2.1) by histological examination using HE-stained slides, approximately 1-2 mm of frozen sample was grated in to a nitrogen-chilled 1.5 ml sample tube. The sample tube was then immediately placed in to liquid nitrogen and later on stored in a freezer at -80°C for long-term storage until further use.

### 2.2.4 FFPE Sample Collection

Before a tissue punch of a FFPE tumor block (see chapter 2.2.2) could be made, a trained neuropathologist examined the HE-stained slides of the corresponding blocks. The suitable area was then circled in, which should be later on further

examined. A disposable biopsy punch with a diameter of 1.5 mm was used to punch out the marked tissue area for further investigation.

### 2.2.5 FFPE Sample Punch Pulverization and Deparaffinization

The sample punch (see chapter 2.2.4) was placed in a bead tube from the Precellys lysis kit and filled with xylene. After 5 min of incubation in the thermocycler (30°C, 600 rpm) the xylene was discarded and replaced with 100% EtOH. The incubation in EtOH with the thermocycler (30°C, 600 rpm) was carried out two times with the EtOH being changed after the first incubation step. Next, the Precellys tube was placed into the tissue homogenizer (2x20 sec, 6000 rpm). After bead homogenization, the tube was incubated again in the thermocycler (30°C, 600 rpm, 2 min). The cloudy suspension was pipetted into a fresh 1,5 ml sample tube and the Precellys tube washed with EtOH and added to the sample tube. The sample tube was then centrifuged at 21000 g for 10 min and the supernatant replaced with xylene. The pellet was resuspended and incubated in the thermocycler (30°C, 600 rpm, 30 min). After an additional centrifugation (21000 g, 10 min) the supernatant was exchanged by xylene and the incubation-centrifugation step repeated. Next, the supernatant was exchanged by EtOH and incubated in the thermocycler (30°C, 600 rpm, 5 min). After an additional centrifugation (21000 g, 10 min) the supernatant was exchange by fresh EtOH and the incubation-centrifugation step repeated. The supernatant was removed and the pellet dried in a vacuum concentrator for 15 min at 30°C.

## 2.2.6 Histochemistry for FFPE Tissue

*Histochemistry were performed by the routine diagnostic lab of the Heidelberg neuropathology.*

HE-staining and Immunohistochemistry were carried out on 4 µm thick FFPE tissue sections mounted on Superfrost® Plus slides. Immuno stains were performed on a BenchMark Ultra immunostainer with a following incubation in increasing concentrations of EtOH and xylene and a final mounting using Eukitt® (see final five steps for HE-Staining in Table 8).

**Table 8: Protocols for HE- and Immunostains.**

| Stain type | Protocol |
|---|---|
| **HE-staining** | 3 x 5 min in xylene<br>1-10 sec in EtOH, ≥99.8 %<br>1-10 sec in EtOH, 96%<br>1-10 sec in EtOH, 70%<br>1-10 sec in Aqua Dest.<br>Hematoxylin Mayer's 3 min<br>Tap water 3 min<br>Eosin in alcohol 1-2 min<br>1-10 sec in EtOH, 70%<br>1-10 sec in EtOH, 96%<br>1-10 sec in EtOH, ≥99,8 %<br>2x10 sec in Xylene<br>mounting of slides with Eukitt® |

| | |
|---|---|
| **VIM-staining** | Ventana Deparaffinization protocol<br>Incubation in Cell Conditioner 1 for 64 min<br>Incubation in VIM-antibody (1:300) for 1 h<br>Ventana ultraWash protocol<br>Incubation in Hematoxylin I for 8 min<br>Incubation in Bluing Reagent for 4 min |
| **HIP1R-staining** | Ventana Deparaffinization protocol<br>Incubation in Cell Conditioner 2 for 56 min<br>Incubation in HIP1R-antibody (1:200) for 1 h<br>Ventana ultraWash protocol<br>Incubation in Hematoxylin I for 8 min<br>Incubation in Bluing Reagent for 4 min |
| **ATRX-staining** | Ventana Deparaffinization protocol<br>Incubation in Cell Conditioner 2 for 64 min<br>Incubation in ATRX-antibody (1:2000) for 1 h<br>Incubation in HQ Linker for 8 min<br>Incubation in HRP Multimer for 8 min<br>Incubation with Amplification $H_2O_2$ and Amplifier for 4 min<br>Incubation in Amplification Multimer for 4 min<br>Incubation in Hematoxylin II for 8 min<br>Incubation in Bluing Reagent for 4 min |

## 2.2.7 Protein Extraction

### 2.2.7.1 FF Tissue Lysis

Lysates from FF tissue (see chapter 2.2.3) was performed using FF tissue lysis buffer. Incubation was carried out in a thermomixer for 5 min at 95°C and 600 rpm. Immediately after the sample was placed on ice for 5 min. Afterwards the sample was sonicated using the probe sonicator with two short bursts (1 sec each) and placed on ice.

#### 2.2.7.2 FFPE Tissue Lysis

The sample pellet (see chapter 2.2.5) was resuspended in FFPE tissue lysis buffer and transferred in to a MicroTube. MicroTubes were closed off using a MicroCap (50 µl) with the help of the MicroCap Tool. The lysis was carried out under pressure cycling in a Barocycler 2320EXT (60 cycles, 45000 psi for 50 s per cycle and 14.7 psi for 10 s per cycle, 95°C). Afterwards protein quantity was determined using the Pierce™ BCA Protein Assay Kit according to the manufacturer's protocol

### 2.2.8 Protein Digestion using Pressure Cycling Technology

For proteome analysis 50 µg of protein lysate (chapter 2.2.7.1 and 2.2.7.2) were used for further preparation. Protein reduction and alkylation were carried out adding Tris (2-carboxyethyl) phosphine (10 mM) and iodoacetamide (40 mM) subsequently to the solution for 30 min in the dark at 25°C. To remove SDS from the protein solution, proteins were precipitated by adding acetone (80%) to the solution and incubated at -20°C for 1 h. After centrifugation of the tubes for 5 min at 21000 g the supernatant was carefully removed without dislodging the pellet. The pellet was shortly washed with 90% acetone and centrifuged again for 5 min at 21000 g. The supernatant was carefully removed and the pellet left to dry for a maximum of 3 min. Afterwards the pellet was resuspended in 100 mM ammonium bicarbonate using an ultrasonic bath until the pellet was completely disaggregated. The resulting suspension was transferred into MicroTubes. For digestion Trypsin/Lys-C Mix was added to each sample (enzyme-to-substrate-ratio, 1:25). Tubes were closed with MicroCaps (50 µl). The digestion was carried out under pressure cycling in a Barocycler NEP2320EXT (150 cycles, 45000 psi for 50 sec/cycle and 14.7 psi for 10 sec/cycle at 33°C). Samples were dried in a vacuum concentrator for 1 h or longer if needed at 30°C. Before MS analysis a solution of 0.1% trifluoroacetic acid and 2.5% hexafluoroisopropanol was added to the dried pellet and placed in a sonication bath for 5 min. The peptide solution was quantified using Pierce™ Quantitative Colorimetric Peptide Assay according to the manufacturer's protocol and the concentration adjusted to 0.2 µg/µl. 1 µg of peptide was added to a HPLC Vial and spiked in with iRT peptides according to manufactures protocol.

## 2.2.9 Tandem Mass Spectrometry using the Orbitrap Exploris™ 480

### 2.2.9.1 Liquid Chromatography

LC-MS samples (see chapter 2.2.8) were separated using an UltiMate 3000 HPLC. Prior separation samples were directed through an Acclaim PepMap 100 trap column (C18 100 Å, 5 µm, 300 µm x 55 mm) for trapping. Peptides were separated on nanoLC nanoEase 200 mm M/Z BEH C18, 130 Å, 1.7 µm column. Solvent A contained 99.9% H2O and 0.1% formic acid and solvent B contained 80% ACN, 19,9% H2O and 0,1% formic acid, for low pH peptide separation. For each analysis 1 µg of peptide was loaded onto the column and separated using a 3.5h gradient. This was set as follows: at 2% of B for 3 min, from 2 to 8% B in 15 min, from 8 to 25% B in 125 min, from 25 to 40% B in 30 min, from 40 to 95% B in 1 min, 5 min at 95% B, from 95 to 2% B in 1 min, 30 min at 2%B.

### 2.2.9.2 Mass Spectrometry with Data Dependent Acquisition

The Ultimate 3000 HPLC was coupled online to an Orbitrap Exploris™ 480 Mass Spectrometer. The sample was ionized using 2.2 kV spray voltage and a capillary temperature of 275°C. The mass spectrometer was operated in data-dependent mode to automatically measure MS1 and MS2. Full scan MS1 spectra were acquired at a resolution of 120.000 at 400 m/z in the orbitrap covering the mass range of 375-1700 m/z. MS2 spectra were acquired at a resolution of 30.000 at 400 m/z in cycles after every full scan, where the 20 most abundant precursor ions with a normalized collision energy of 27 were selected for fragmentation. Dynamic exclusion of 60 sec was used to prevent extensive MS2 acquisition of high abundant peptides.

### 2.2.9.3 Database Searching with MaxQuant

For the analysis of data dependent raw data, MaxQuant (version 1.6.17.0) [43] with the integrated Andromeda search engine was used for peptide identification [50]. The raw data was searched against the human UniProt database (downloaded

05/09/2018; 20373 sequences). Precursor tolerance was set to 20 ppm for the first search and the isotope match tolerance to 2 ppm. Trypsin/P was specified as protease with a maximum of 2 missed cleavages. Carbamidomethyl (C) was set as a fixed modification and oxidation (M), Acetyl (Protein N-Term), as variable modifications. False Discovery Rate (FDR) was set to 1% on PSM and protein level, match between runs was enabled. A protein was considered quantified when it had at least 2 unique and/or 2 razor peptides corresponding to the given protein identified. Protein abundances were log2 transformed for linearization of the data and in order to remove any confounding factors, proteins which had more that 66% missing values in the complete sample cohort were excluded from further analysis. Remaining missing values were imputed with values from a -1.8 downshifted standard deviation and a width of 30% for the gaussian distribution in order to simulate a low abundance of the given protein.

### 2.2.10 Tandem Mass Spectrometry using the timsTOF Pro

*Following subchapters were performed by Dr. Ashok Kumar Jayavelu, at the Max Planck Institute of Biochemistry in Martinsried, Germany. Text passages of following subchapters have been taken from Felix et al "HIP1R and VIM immunohistochemistry predict 1p/19q status in IDH-mutant glioma." (manuscript in preparation).*

#### 2.2.10.1 Mass Spectrometry with PASEF Data Dependent Acquisition

The timsTOF Pro mass spectrometer was operated in data dependent (DDA)-PASEF mode. For a single TIMS-MS survey scan, 10 PASEF MS/MS scans were acquired per topN acquisition cycle. Ion accumulation and ramp time in the dual TIMS analyzer was set to 50 ms each and ion mobility range from 1/K0 = 1.6 Vs cm-2 to 0.6 Vs cm-2 was analyzed. Suitable precursor ions for PASEF-MS/MS were selected in real time from the TIMS-MS scan. Precursors for PASEF MS/MS were picked at an intensity threshold of 1500 arbitrary units (a.u.) and resequenced the low abundant precursors until reaching a 'target value' of 20000 a.u taking into account a dynamic exclusion of 40 sec elution. MS and MS/MS were recorded from m/z 100 to 1700. Precursor ions for MS/MS analysis were isolated with a 2 Th window for m/z < 700 and 3 Th for m/z >700 in a total m/z range of 100-1700 by

synchronizing quadrupole switching events with the precursor elution profile from the TIMS device. The collision energy was lowered linearly as a function of increasing mobility starting from 59 eV at 1/K0 = 1.6 VS cm-2 to 20 eV at 1/K0 = 0.6 Vs cm-2. By applying the polygon filter, m/z and ion mobility plane representing peptide precursor was selected and singly charged precursor ions were excluded.

#### 2.2.10.2 Liquid Chromatography

An ultra-high pressure nanoflow chromatography, Easy nLC 1200 system was coupled online to timsTOF Pro. Peptides (equivalent to 500 ng) were loaded onto a 50 cm, in house paced reversed-phase column (75 µm inner diameter) with a pulled emitter tip, packed with 1.9 µm C18-coated porous silica beads (Dr. Maisch, Ammerbuch-Entringen, Germany). The column temperature was maintained at 60°C and chromatography was performed with a constant flow of 400 nl/min on a binary buffer system. Mobile phases A (0.1% formic acid (v/v)) and B (80/20/0.1% ACN/water/formic acid (v/v/vol)) respectively, for low pH peptide separation. In 120 min gradient, peptides were separated with a linear gradient from 7.5 to 27.5% B within 60 min, followed by an increase to 37.5% B within 30 min and further to 55% within 10 min, followed by a washing step at 95% B and re-equilibration.

#### 2.2.10.3 Database Searching with MaxQuant

The TIMS raw data were analyzed in the MaxQuant environment (version 1.6.17.0) which extracts features from four-dimensional isotope patterns and associated MS/MS spectra. The data was searched against the human UniProt database (August 2018 release). False-discovery rates was set at 1% both on peptide spectral match (PSM) and protein levels. Enzyme specificity was set to trypsin cleaving c-terminal to arginine and lysine. A maximum of two missed cleavages were allowed. Peptides with a minimum length of seven amino acids were considered for the search including N-terminal acetylation and methionine oxidation as variable modifications and cysteine carbamido methylation as fixed modification, while limiting the maximum peptide mass to 4600 Da. Maximum precursor and fragment ion mass tolerance were set as default, which is the first search peptide tolerance to 70 and main search tolerance to 20. Peptide identifications by MS/MS were

transferred by matching four-dimensional isotope patterns between the runs (MBR) with a 0.7-min retention-time match window and a 0.05 1/K0 ion mobility window. Label-free quantification was performed with the MaxLFQ algorithm and a minimum ratio count of one. If proteins could not be distinguished on the basis of unique peptides, they were merged by MaxQuant as one protein group.

### 2.2.11 Data Processing

*Data processing were performed by Dennis Friedel, Institute of Pathology, Department of Neuropathology, Heidelberg and myself. Text passages of this section have been taken from Felix et al. "HIP1R and VIM immunohistochemistry predict 1p/19q status in IDH-mutant glioma". (manuscript in preparation).*

For data transformation and visualization, the freely available software called Perseus (version 1.6.14.0) [76] and R (version 4.0.4, https://www.r-project.org/index.html, last accessed 14.04.21) were used. Proteins that were identified by a single peptide, as well as potential contaminants and proteins identified from the decoy database containing reversed protein sequences were excluded from further data analysis. Proteins with more than 10% missing values within a condition were discarded to reduce data missingness which could interfere with upcoming statistical analysis. The functions from the R-package 'DEP' were used for normalization, missing value imputation and for the following differential expression analysis. Here, technical variability between samples was corrected with vsn. After normalization, missing values were substituted by using a downshift imputation (shift =1.5, scale=0.5). Differential expression was examined by the R-package 'limma' and Benjamini-Hochberg was used for FDR controlling of the resulting p-values. Proteins were considered to be significantly regulated if their abundance had a fold change value higher than 1.5 between astrocytoma and oligodendroglioma and an adjusted p-value <0.01

# 3. Results

## 3.1 Proteomic Exploration of IDH-Mutant Gliomas

In this study IDH-mutant glioma were analyzed with the scope to reveal differences in protein abundances between astrocytoma and oligodendroglioma for biomarker discovery and to link the proteomic data with already existing genomic datasets. Since the FF tissue sample preparation is simpler and better established than for FFPE tissue, 41 FF tissue samples were used in the discovery phase. Once an initial picture of the different tumor entities was made, an FFPE tissue protocol was established and benchmarked against the FF tissue results. Afterwards, a more in-depth analyses of a larger FFPE cohort of 72 samples was carried out and linked with existing methylation datasets.

### 3.1.1 Establishing a Link from FF to FFPE Tissue of IDH-Mutant Glioma

#### 3.1.1.1 Proteomic Landscape of FF Tissue of IDH Mutant Glioma

The here investigated FF cohort of brain tumors consisted out of 20 astrocytomas IDH-mutant and 21 oligodendrogliomas IDH-mutant, 1p19q-codeleted. After all samples were prepared for proteomic analysis, 1 µg of peptide from each sample were submitted for proteomic analysis using the Dionex Ultimate 3000 in configuration with the Thermo Exploris 480 Orbitrap mass spectrometer (subchapters of 2.2.9).

The MS-analysis revealed 8,388 proteins and 75,612 peptides in all FF samples combined. The mean identification rate was 5,003 (SD: 248) proteins and 33,534 (SD: 3,322) peptides with an FDR of below 1% per sample. In order to find significant deregulated proteins between astrocytomas and oligodendrogliomas, differential expression analysis was performed. Means of the log2 protein abundances of the two sample cohorts were calculated and then subtracted from one another for fold change calculation. Protein abundances from each cohort were checked for significance by a 2-sided t-test. The resulting p-values were adjusted by permutation-based FDR using the Benjamini Hochberg procedure. Proteins were considered as significantly deregulated if they had an FDR <1% and a fold-change

value >1.5. A summary of significantly deregulated proteins between astrocytoma and oligodendroglioma in FF tissue is depicted as a volcano plot in Figure 6.

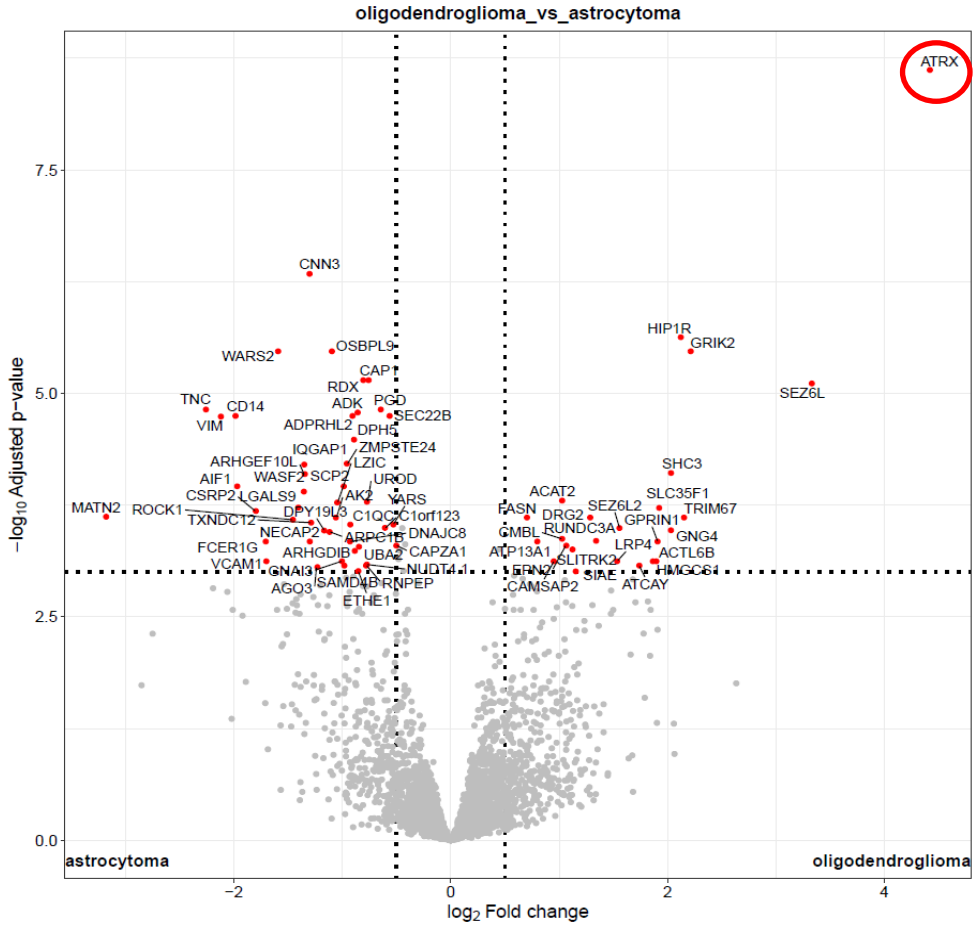


**Figure 6: Volcano plot of significantly deregulated proteins between astrocytomas and oligodendrogliomas.** Proteins above the defined FDR of 1% and fold change of 1.5 are depicted as red dots. Proteins significantly higher abundant in astrocytomas are on the left side of the plot, Proteins significantly higher in oligodendrogliomas are on the right side of the plot. A clear confirmation of ATRX (circled red) previously reported as a biomarker for differentiation between astrocytomas and oligodendrogliomas [1].

In total 69 proteins were identified that were above the given threshold parameters. ATRX was identified with the highest fold change value and significance which has been previously recognized as a loss of function biomarker for the diagnosis of astrocytomas [1]. Additionally, proteins such as HIP1R, GRIK2, SEZ6L, TNC, CNN3 and VIM showed a distinct higher significance and fold change then other proteins in the volcano plot. Later experiments with FFPE confirmed these findings with measurements being carried out on a timsTOF pro platform (chapter 3.1.2.1).

### 3.1.1.2 Proteomic comparison of FF and FFPE Tissue of IDH-Mutant Gliomas

In order to see how comparable FF and FFPE samples are to one another, a direct comparison of tissue from the same tumor is best. For this a group of 16 FF–FFPE sample pairs was assembled, which derived from the IDH-mutant glioma cohort in chapter 3.1.1.1. The investigation was focused on how similar these samples were in terms of peptide and protein identification and if proteins identified in both sample pairs had similar LFQ intensities.

As can be seen in Figure 7 (left), peptide identifications were approximately ~33% and protein identifications ~25% higher in FF samples than in FFPE samples. Additionally, higher fluctuations can be seen on peptide and protein level when looking into the FFPE group.

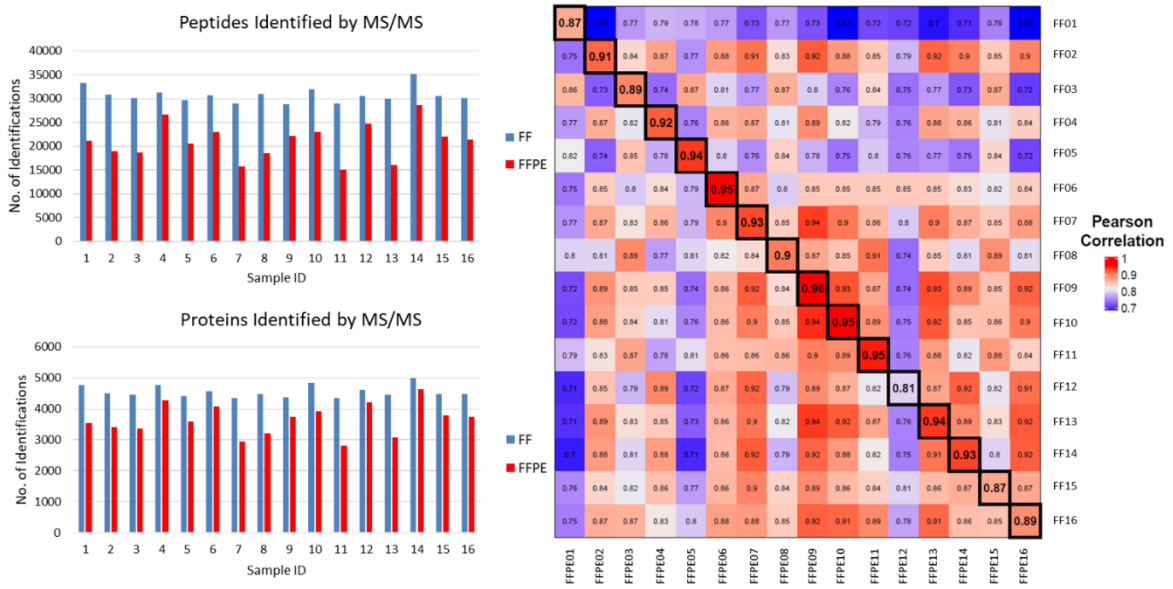


**Figure 7: Comparison of identification rates and quantification between FF and FFPE sample pairs. Left:** Peptide (top bar graph) and protein (bottom bar graph) identifications of FF (blue) and FFPE (red) sample pairs. Number of identifications is based on MS/MS sequencing**.** Sample ID between the different bar plots resembles sample-pair assignment. Identified peptide sequences (top) are later assigned to their corresponding protein group (bottom). Mean number of identifications for peptides was 30719 (SD: 1655) in FF and 20974 (SD: 3835) in FFPE. The mean number of identifications for proteins was 4555 (SD: 188) in FF and 3649 (SD: 508) in FFPE. **Right**: Quantitative correlation between FF and FFPE sample pairs. Pearson correlations between all FF-FFPE sample pairs are marked with a black square and have a mean correlation of 0.91 (SD: 0.04).

When comparing the FF-FFPE pairs directly, proteins that have been identified in both sample types showed a very high correlation with a mean of 0.91 (SD 0.04) in terms of log2 protein abundance (Figure 7, right).

## 3.1.2 In Depth Analysis of IDH-Mutant Glioma using FF and FFPE Tissue

*LC-MS measurements were performed by Dr. Ashok Kumar Jayavelu, at the Max Planck Institute of Biochemistry in Martinsried, Germany. Text passages from subchapters of 3.1.2.1 to 3.1.2.4 have been taken from Felix et al. "HIP1R and VIM immunohistochemistry predict 1p/19q status in IDH-mutant glioma" (manuscript in preparation).*

With access to a timsTOF Pro coupled with an EASY-nLC™ 1200 HPLC and having a successfully working FFPE sample preparation protocol, a remeasurement of FF samples and an expanded FFPE group was undertaken. In total 35 FF samples and 72 FFPE samples were analyzed. The FF group consisted out of 15 astrocytomas, IDH-mutant and 20 oligodendrogliomas, IDH-mutant, 1p/19q-codeleted. The FFPE cohort consisted out of 36 astrocytomas, IDH-mutant and 36 oligodendrogliomas, IDH-mutant, 1p/19q-codeleted. From 12 patients FF as well as FFPE samples were available for proteomic analysis. All samples were measured in PASEF-DDA mode and bioinformatically processed as described in the subchapters of 2.2.10.

### 3.1.2.1 Proteomes from FF and FFPE tumors are comparable

Using a 2 h gradient we identified by MS/MS with a DDA-PASEF method on average 5171 (+/-322) protein groups and 26830 (+/- 2586) unique peptides per sample in FF tissue. In total we identified 8220 proteins and 82885 unique peptides in the FF cohort. In FFPE tissue, we identified by MS/MS on average 3905 (+/- 493) protein groups and 16681 (+/- 3549) peptides per sample. In total we identified 7912 proteins and 72867 unique peptides in the FFPE cohort.

To identify significantly differentially abundant proteins (DAP), astrocytoma and oligodendroglioma samples were compared for each tissue cohort. From patients with paired FF and FFPE samples only FF samples were used and their FFPE

counterparts were excluded to prevent statistical inflation. ATRX is well established to be differentially expressed in astrocytoma and oligodendroglioma due to ATRX mutations in IDH-mutant astrocytomas [1]. Indeed, ATRX was one of the most significantly differentially abundant proteins between astrocytoma and oligodendroglioma (Figure 8a, b). Overall, we found 134 DAPs in the FF cohort and 397 in the FFPE cohort. The higher DAP count in the FFPE material was attributed to the larger size of the FFPE cohort. 78 DAPs were found in both cohorts (Figure 8c). Pearson-correlations from the 12 matched pairs of FF and FFPE tissues from the same tumor ranged from 0.7 to 0.87 (Figure 8d, e).

#### 3.1.2.2 Protein Signature Allows Distinction of Astrocytoma and Oligodendroglioma

We evaluated whether the 78 consistent DAPs can be utilized as a signature for discriminating astrocytoma and oligodendroglioma. To this end, hierarchal clustering was performed on the z-scored log2 intensities of the overlapping cohort of DAPs. Oligodendroglioma and astrocytoma samples separated with only one oligodendroglioma in FF and two astrocytoma and two oligodendrogliomas in the FFPE cohort clustering aberrantly (Figure 8f).

Rarely, DNA methylation-based classification results in a classification of 1p/19q-codeleted tumors as “high grade astrocytomas, IDH-mutant”. We included 5 such FFPE oligodendroglioma samples (termed “rare oligodendroglioma”), that had been excluded from previous differential expression analysis, to determine if the similarity to astrocytomas visible on the DNA methylation level is also evident on the proteome level. Interestingly, 4/5 rare oligodendrogliomas clustered together with oligodendrogliomas, whereas one clustered into the astrocytoma cohort (Figure 8f).

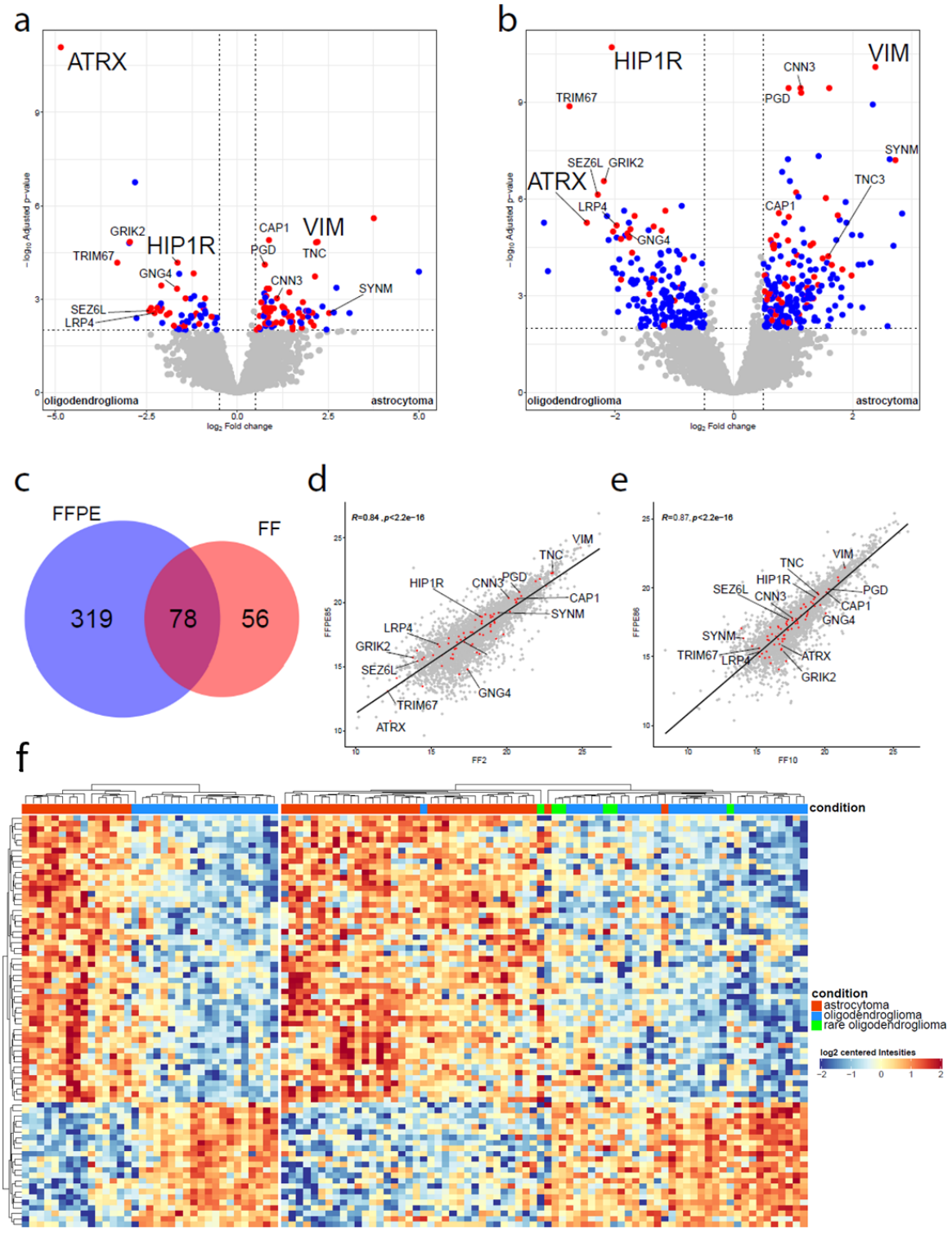


**Figure 8: Differential protein expression analysis of astrocytomas and oligodendrogliomas**. Volcano plot of significantly differentially abundant proteins (DAP) between astrocytomas and oligodendrogliomas in FF (**a**) and FFPE (**b**) tissue. Blue dots represent proteins which were found either in the FF or FFPE tissues exclusively, red dots represent proteins which were found consistently in both cohorts. Dashed lines represent threshold for significance (p-value <0.01) and fold change (1.5) **c**: Venn diagram depicting number of significantly DAP in FF and FFPE tissue. Overlap shows the number of proteins which were significantly different in both cohorts. Pearson correlation analyses between a matched pair of FF and FFPE sample of an astrocytoma (**d**) and an

oligodendroglioma (**e**). **f**: Hierarchical clustering of the FF (left) and FFPE (right) cohort using the overlapping significantly DAP between astrocytoma and oligodendroglioma. Rare oligodendrogliomas (condition green): 1p/19q-codeleted tumors with the highest brain tumor classifier score for “high grade astrocytoma, IDH-mutant”.

#### 3.1.2.3 HIP1R and VIM as Surrogate Biomarkers for 1p/19q Status

Using the most significant proteins from the volcano plots in Figure 8a and 8b, we evaluated various antibodies for immunohistochemical detection. For several proteins no suitable antibodies were available (TRIM67), immunohistochemical staining was unsatisfactory or the staining provided largely redundant information (e.g. VIM CNN3 and TNC). Further evaluation revealed HIP1R and VIM as promising candidates which were later on chosen for more detailed evaluation since they provided the most consistent stains.

Expression patterns of HIP1R and VIM were evaluated using 20 astrocytomas and 15 oligodendrogliomas which were initially investigated by LC-MS (chapter 3.1.2). For HIP1R, four different expression levels (H0-H3, Figure 9) were recognized, as well as four different expression levels for vimentin (V0-V3, Figure 10). These expression patterns in combination with an evaluation strategy were later on verified by PD Dr. David Reuss and Dr. Annekathrin Reinhardt using an independent cohort of 50 astrocytoma and 50 oligodendroglioma which were not analyzed using LC-MS.

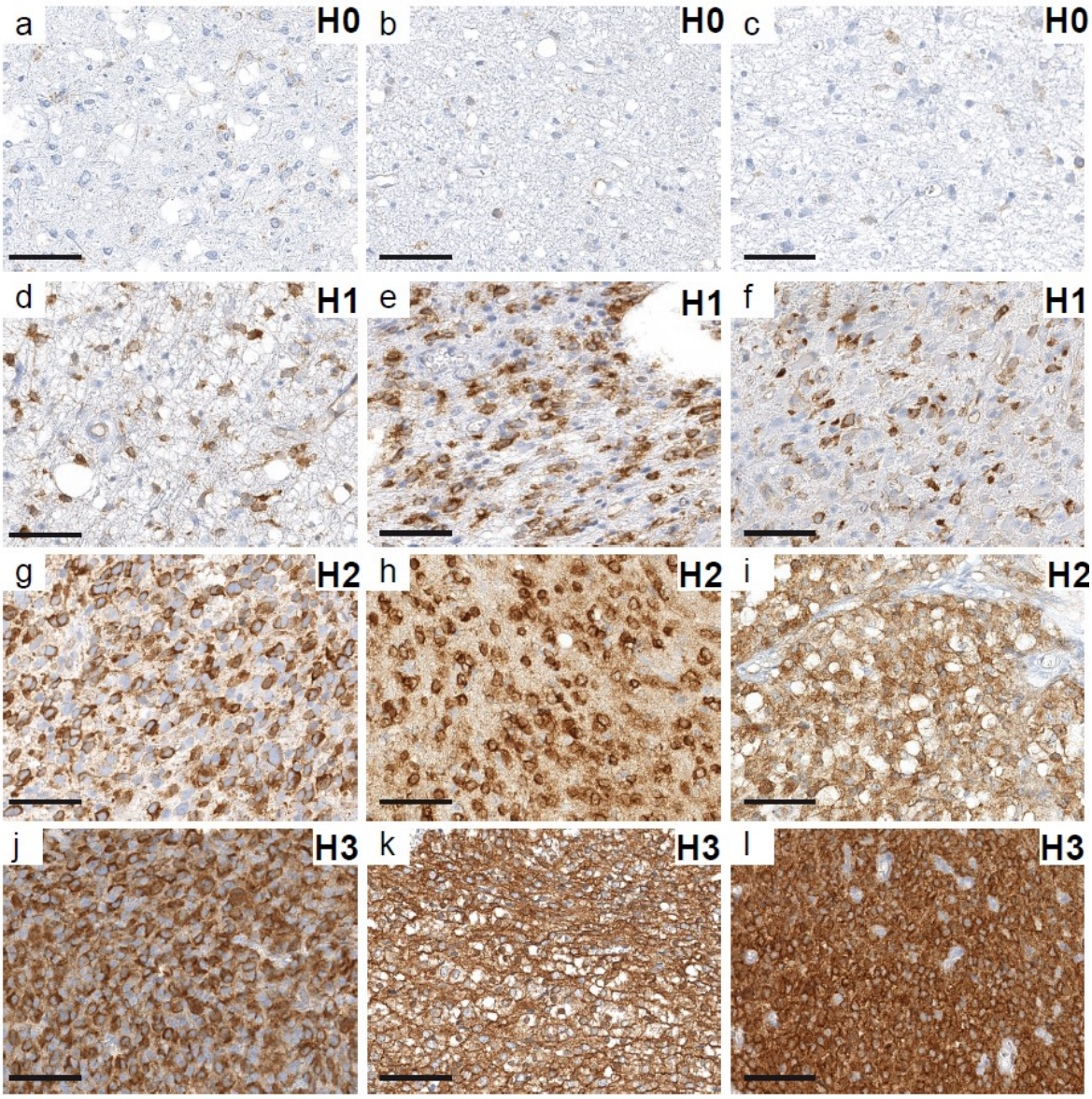


**Figure 9: HIP1R intensity scaling**. Each row shows 3 exemplary stains for each intensity. **H0:** macroscopically/low magnification negative; microscopically / high magnification single/few cells are weakly to moderately positive. **H1:** macroscopically/low magnification negative or very slightly brown; microscopically / high magnification many cells are weakly positive / few cells are moderately positive. **H2:** macroscopically/low magnification brown; microscopically / high magnification many cells are moderately / strongly positive but do not form diffusely positive areas. **H3:** macroscopically/low magnification brown areas; microscopically / high magnification all or almost all cells are moderately to strongly positive forming homogenously positive areas (positive cells are closely packed without space in between; only vessels are negative). Brown staining represents positivity of HIP1R, black line indicates a length of 60µm. Images were captured at 400-fold magnification.

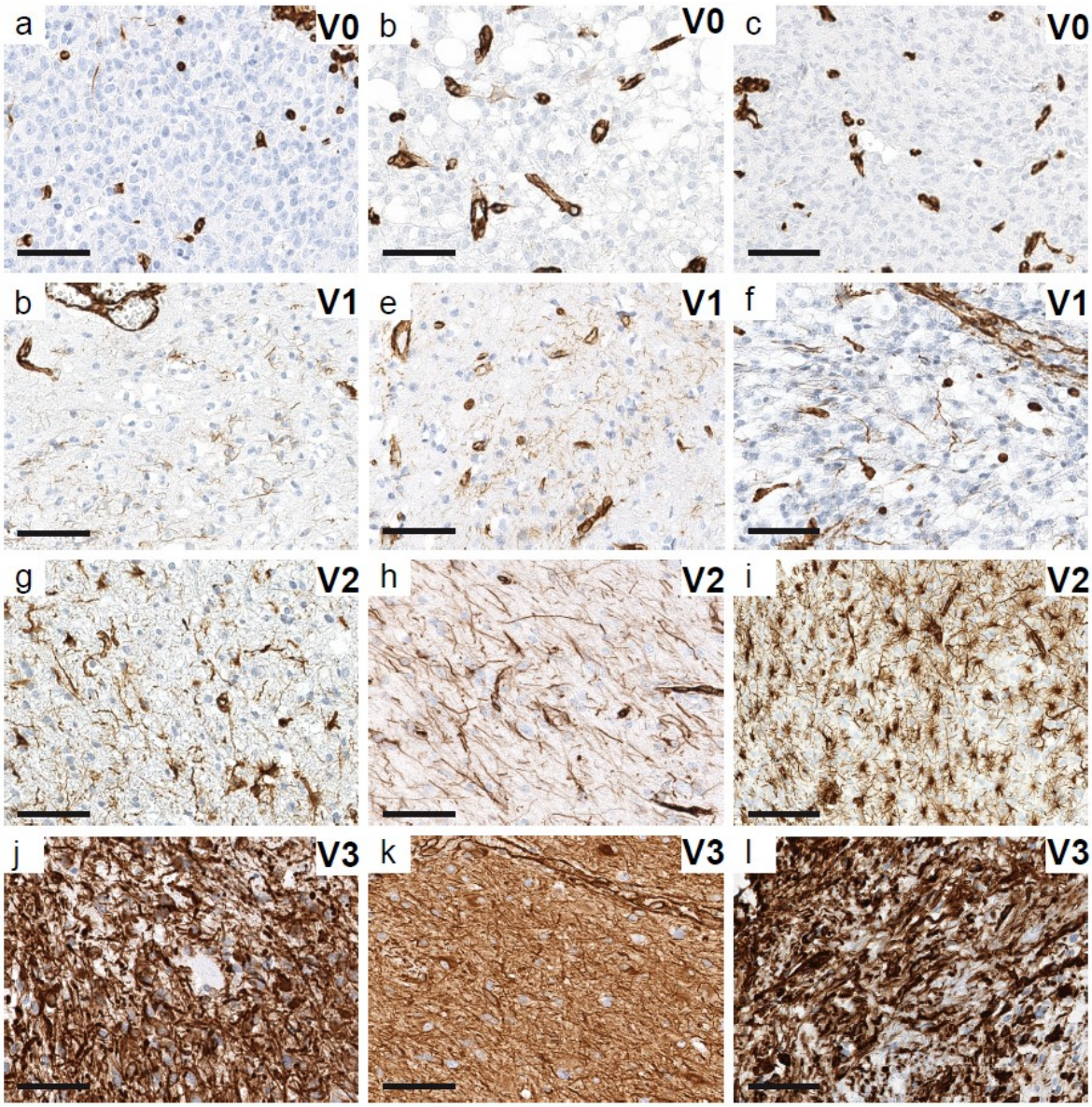


**Figure 10: VIM intensity scaling.** Each row shows 3 exemplary stains for each intensity. **V0:** no tumor cells are labelled, only vessels or other non-neoplastic cells are positive. **V1** light staining intensity with typical fibrillary pattern around the nucleus. **V2:** moderate to high staining intensity of fibrillary patterns in a large fractions of tumor cells with occasional star-like stain pattern (**i**). **V3:** high intensity of all or ≥70% tumor cells. Brown staining represents positivity of VIM, black line indicates a length of 60µm. Images were captured at 400-fold magnification.

#### 3.1.2.4 Entity-Specific Expression Patterns of HIP1R/VIM and Rules for Interpretation

Most commonly, oligodendrogliomas showed high abundance levels of HIP1R (H2, H3) and absent (V0) or low (V1) levels of VIM. In contrast astrocytomas commonly showed low (H0) to moderate (H1) levels of HIP1R but high levels of VIM (V3), results being fully consistent with those from the proteomic analyses. Frequently areas with different expression levels for both HIP1R and VIM occurred within in a given tumor. In some samples there was an intratumoral inverse expression pattern of HIP1R and VIM observable with areas of high-level HIP1R expression and low-level VIM expression abruptly alternating with areas with low-level HIP1R and high-level VIM expression. Positivity for IDH1-R132H in both areas confirmed this as true intratumoral heterogeneity in a fraction of oligodendrogliomas. In several other cases the VIM positive areas could be identified as being non-neoplastic by IDH1-R132H-negativity in IDH1-R132H-mutant tumors (Figure 11, middle vs right column). Stains of HIP1R and IDH1-R132H revealed identical distribution patterns of positive cells strongly suggesting that HIP1R was expressed in tumor cells only (Figure 11, left vs middle column). In contrast, VIM was present in various cell types. It always showed positivity in vessels making it suitable as an internal control. Additionally, reactive astrocytic proliferations with IDH1-R132H negativity and leptomeningeal structures showed VIM expression as well. Normal parenchyma showed similar expression patterns for VIM as many oligodendrogliomas with positive vessels and surrounding cells with no expression. Since infiltration zones presented a variable picture, only solid tumor areas were considered suitable for evaluation. In specimens with relevant HIP1R expression (at least H1) only the highest scoring area was used for concomitant VIM evaluation. In Summary, using these two markers and the according scoring strategy, we identified 4 patterns of high entity-specificity.

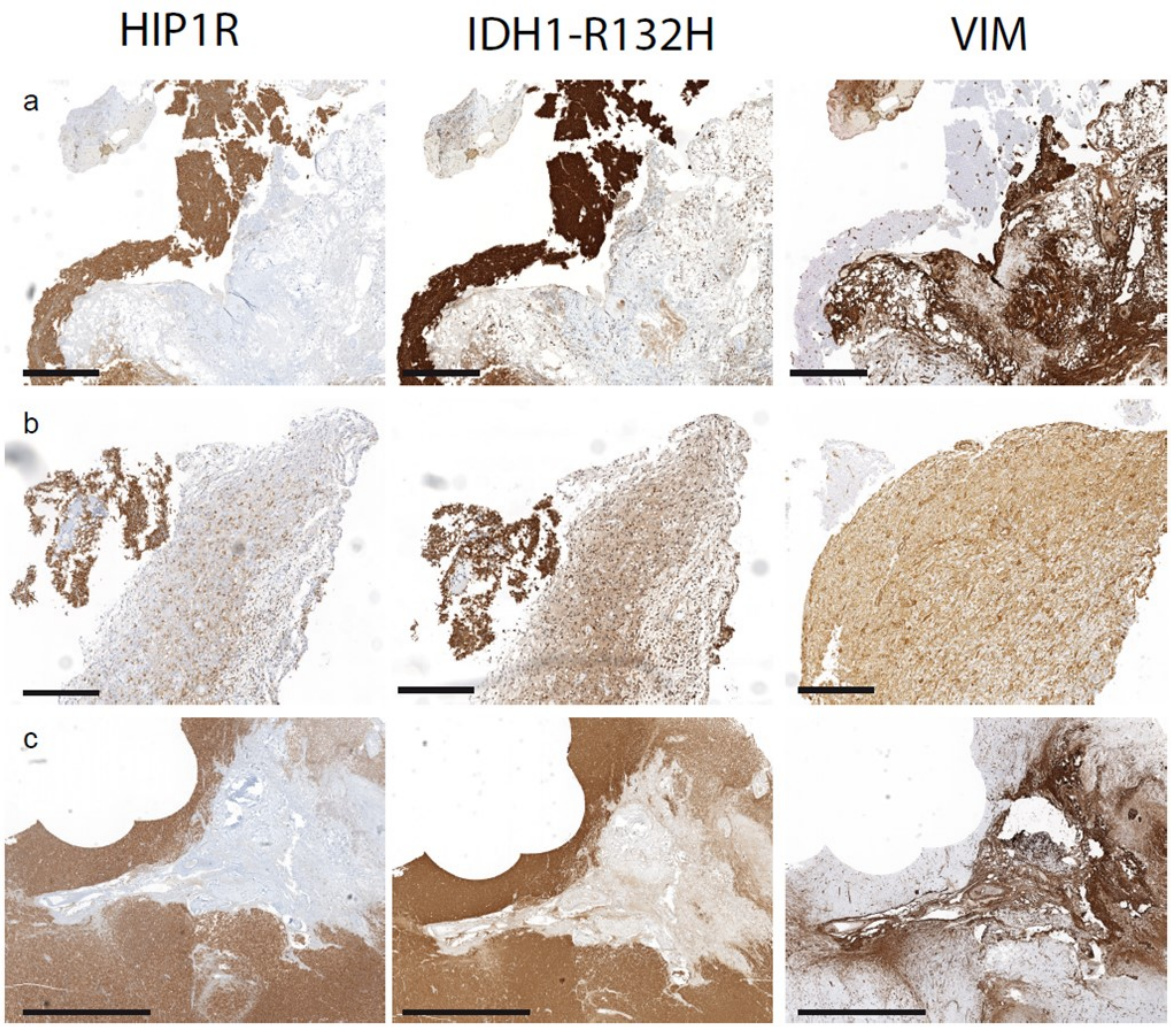


**Figure 11: Staining of HIP1R (left column), IDH1-R132H (middle column) and VIM (right column).** Examples of HIP1R mimicking staining pattern of IDH1-R132H. High positivity of VIM in non-neoplastic areas of low IDH1-R132H expression. Black lines indicate 600 mm (**a**, **b**, **c**), 300µm (**d**, **e**, **f**), 2mm (**g**, **h**, **i**).

In the majority of cases, the assessment is straightforward and the interobserver variation is minimal, not affecting the overall interpretation. In cases where both markers ranged in the moderate or strong expression level, interobserver variance in evaluation was increased. A comparable immunoreactivity of both HIP1R and VIM expression of was observed in a minority of both astrocytomas and oligodendrogliomas and thus was considered non-specific for either entity. In cases were the scoring results were ambiguous, sensitivity could be increased and interobserver variance reduced by directly comparing the HIP1R and VIM immunoreactivity in a given case. If the HIP1R staining was clearly stronger than VIM, the tumor was classified as an oligodendroglioma and if the VIM staining was

clearly stronger than HIP1R the tumor was classified as an astrocytoma. If the immunoreactivities still appeared approximately similar in intensity, the case was called non-determinable (N.D.). Following this practice, a strong (H3) expression of HIP1R in combination with a moderate (V1) expression of VIM was seen in oligodendrogliomas only and a strong expression of VIM (V3) in combination with a moderate expression of HIP1R (H1) was restricted to astrocytomas, except one case of oligodendroglioma. However, in contrast to astrocytomas there were also large areas in this tumor showing the typical sparse VIM expression of oligodendrogliomas. Therefore, strong expression of VIM should not result in a calling of astrocytoma if there are parts of the tumor with sparse VIM expression (V0 or V1) as well. Rarely, astrocytomas may be present with sparse VIM expression similar to oligodendrogliomas. However, HIP1R is not expressed higher than moderate in these cases. Therefore, determination of a tumor as oligodendroglioma required a strong expression of HIP1R (H2 or H3). Especially in grade III oligodendrogliomas VIM expression may be strong, which is however typically mixed with areas of sparse expression. In most of these cases HIP1R is expressed highest in areas with low-level VIM expression, retaining the typical oligodendroglioma pattern in focal areas only. Therefore, in tumors with (very) strong expression of HIP1R, VIM should be evaluated in these high-HIP1R areas only, even if these areas are small. A simplified summary of the evaluation algorithms for VIM and HIP1R in IDH-mutant gliomas is given in Figure 12.

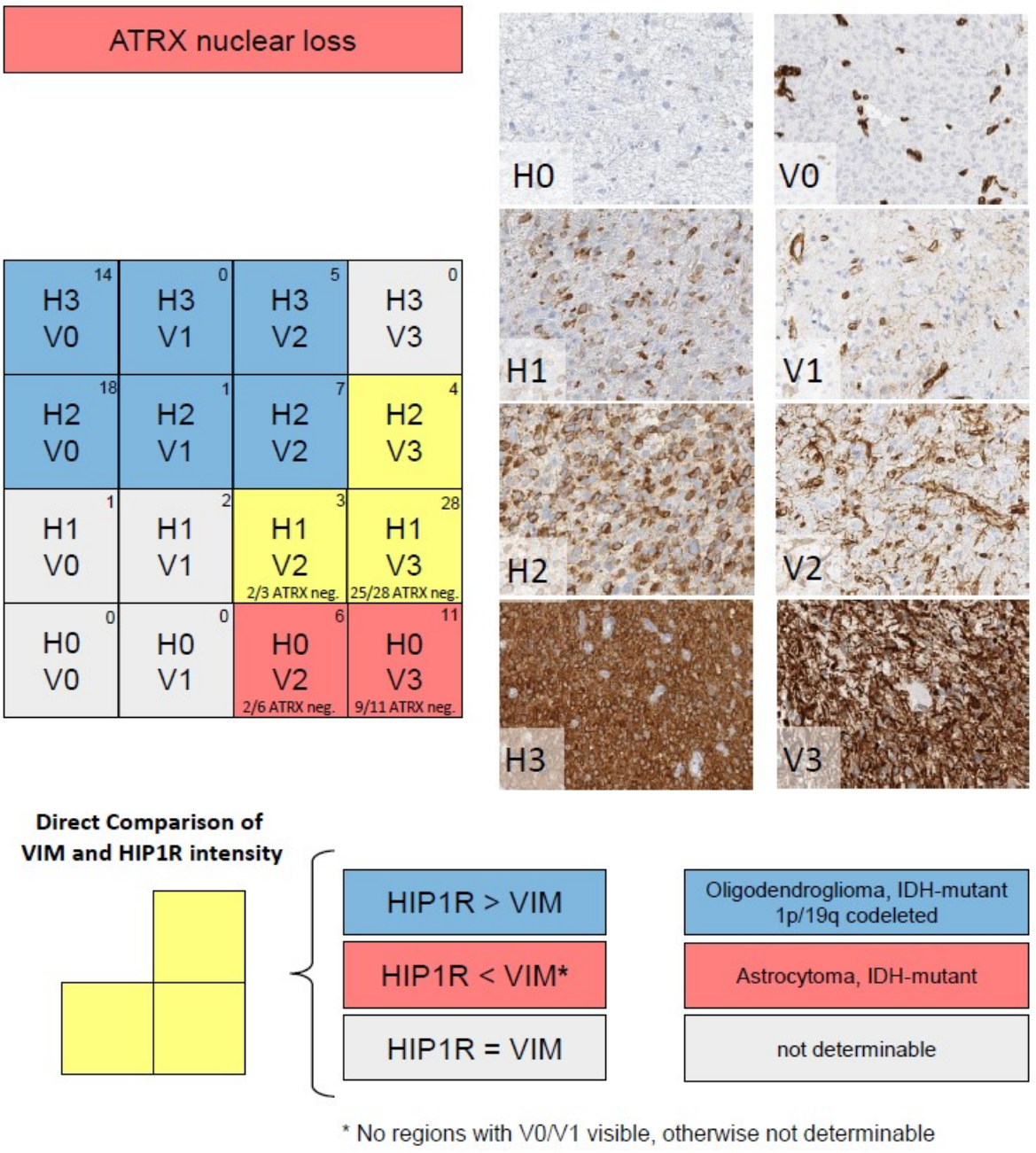


**Figure 12: Algorithm for the classification of IDH-mutant glioma according to HIP1R/VIM immunohistochemistry.** The HIP1R level should be evaluated first. If sparse expression is present (H0) and the VIM level is moderate (V1) to high (V3) the tumor is identified as astrocytoma (red labeled boxes). If the HIP1R level is high (H2, H3) and the VIM level is no more than V2, the tumor is identified as oligodendroglioma (blue labeled boxes). If both the HIP1R level and the VIM level are ambiguous (yellow labeled boxes), the stains should be compared directly. If HIP1R is stronger, the tumor is identified as oligodendroglioma. If VIM is stronger, the tumor is identified as astrocytoma, as long as there are no areas in the tumor with sparse VIM expression (V0, V1) as well. All other potential results including a similar expression level of HIP1R and VIM in the tumor are non-determinable (N.D.). The distribution of HIP1R and VIM expression from the validation cohort is shown in the top right corner and the fraction of samples with ATRX loss are noted on the bottom of every combination box. Distribution numbers are based on the evaluation from DR.

#### 3.1.2.5 Validation of HIP1R/ATRX as a Surrogate for 1p/19q status in IDH-Mutant Glioma

Examples of combinatory results are given in Figure 13. Following these rules 14/15 oligodendrogliomas and 18/20 astrocytomas of the initial cohort could be correctly identified with only 2/35 cases being non-determinable. Including loss of ATRX into the assessment as a specific marker for astrocytomas correctly identified the two astrocytomas which were previously undeterminable by HIP1R/VIM without challenging the interpretation of HIP1R/VIM results. Using this immunohistochemical HIP1R/VIM/ATRX approach, 34/35 tumors of the initial series could be correctly classified as oligodendroglioma or astrocytoma (97% sensitivity, 100% specificity).

Next, we evaluated the specificity and sensitivity as well as the interobserver concordance of immunohistochemical HIP1R/VIM expression patterns in an independent cohort of 100 IDH-mutant gliomas (50 astrocytomas and 50 oligodendrogliomas). For all tumors 450k/850k based CNV profiles for valid 1p/19q determination were available. HIP1R and VIM immunohistochemical stains were scored in combination case-wise from two trained neuropathologists, PD Dr. David Reuss (DR) and Dr Annekathrin Reinhardt (AKR) independent from each other and blinded for the histological diagnosis, the 1p/19q-and ATRX-status.

Based on the HIP1R/VIM immunohistochemistry only DR called 46/50 (sensitivity 92%) oligodendrogliomas and 45/50 (90%) astrocytomas with 100% specificity and called 9/100 cases non-determinably. Including loss of ATRX as marker for astrocytomas correctly identified 4 further cases as astrocytomas (49/50) increasing the sensitivity of a HIP1R/VIM/ATRX triple-marker approach to 98% for astrocytomas and 95% for the whole series.

AKR called 47/50 (sensitivity 94%) oligodendrogliomas and 47/50 (sensitivity 94%) astrocytomas with 100% specificity and called 6/100 cases non-determinable. Including loss of ATRX as marker for astrocytomas correctly identified 2 further cases as astrocytomas (49/50) increasing the sensitivity of a HIP1R/VIM/ATRX triple-marker approach to 98% for astrocytomas and 96% for the whole series.

The overall interobserver concordance was high. 91/100 calls were identical between both observers. DR called 3 tumors correctly which were undeterminable for AKR and AKR called 6 tumors correctly which were undeterminable for DR. As

all calls were correct, the interobserver variance occurred only between correct entity calls and calls as non-determinable.

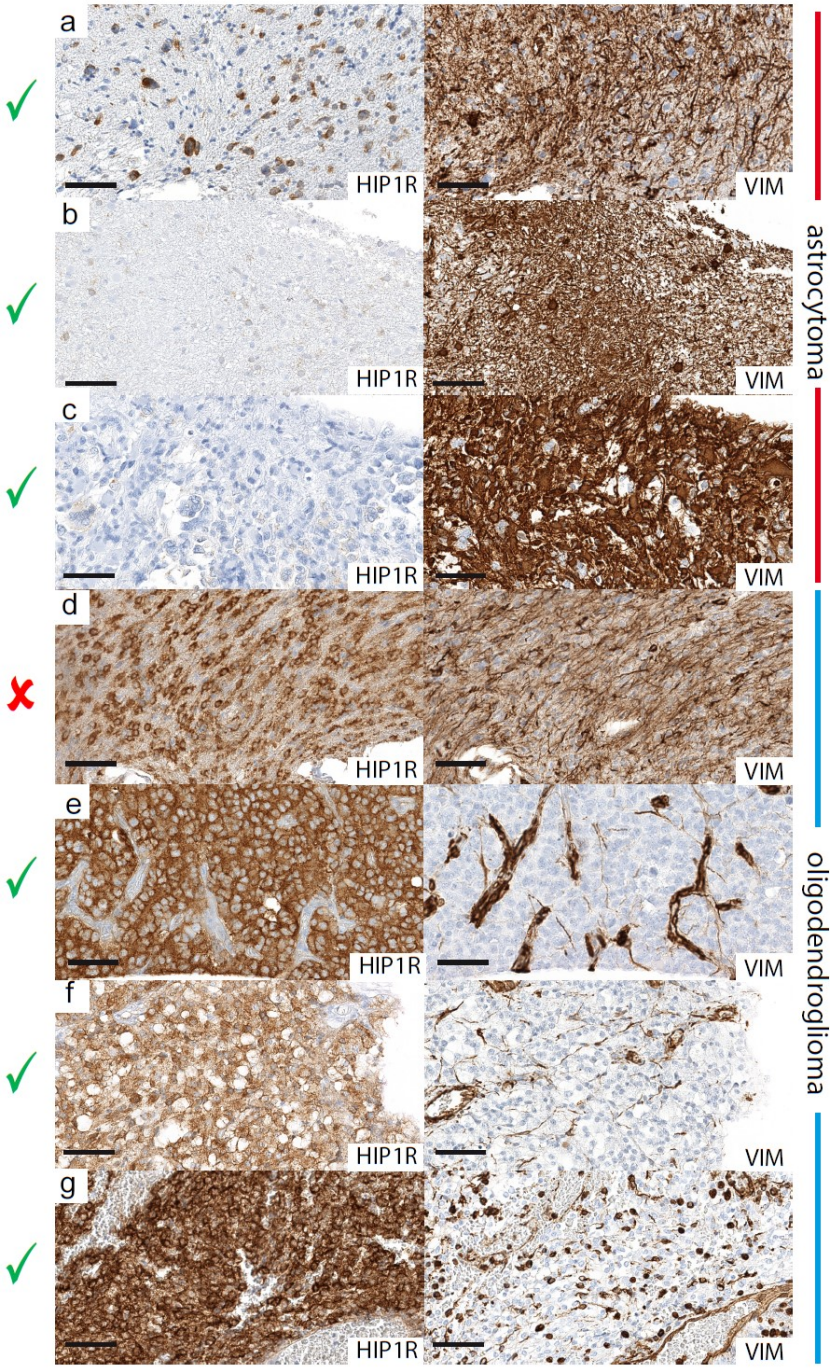


**Figure 13**: **Pairwise comparison of HIP1R (left side) and VIM (right side) in astrocytomas (top 3) and oligodendrogliomas (bottom 3). a**, **b**, **c**: astrocytomas with low-level HIP1R expression and high-level VIM expression. **d**: a non-determinable oligodendroglioma with similar HIP1R and VIM staining intensities. **e**, **f**, **g**: oligodendrogliomas with high-level HIP1R expression and low-level VIM expression. Green check marks indicate determinable cases, red cross indicates not-determinability. Black line indicates a length of 60µm. Images were captured at 400-fold magnification.

### 3.1.2.6 1p/19q-Coded Proteins in Oligodendrogliomas

*LC-MS measurements were performed by Dr. Ashok Kumar Jayavelu, at the Max Planck Institute of Biochemistry in Martinsried, Germany. Text passages in the subchapters of 3.1.2.6. to 3.1.2.8 have been taken from Felix et al. "Copy number alterations profoundly impact the proteome of IDH-mutant Gliomas" (manuscript in preparation).*

Given the utmost importance of 1p/19q co-deletion for the classification of IDH-mutant gliomas, raises the question whether overall abundances of proteins which genes are located on the respective chromosomal arms, are altered. Indeed 29 of 78 overlapping DAPs were encoded on chromosomal arms 1p or 19q (chapter 3.1.2.1). Therefore, all 1p (n=247) and 19q (n=101) encoded proteins in the more numerous FFPE cohort were examined.

Barcode plots were used to visualize the enrichment of these proteins according to the ranked t-statistic obtained from the previous DAP analysis. For both 1p and 19q, an enrichment of proteins with lower ranks was visible indicating a downregulation in both gene sets (Figure 14). In line with these findings, gene set testing showed that 58% of the proteins from 1p were significantly ($p<0.01$) downregulated in oligodendroglioma. However, only 35% of the proteins in 19q showed significant ($p<0.01$) downregulation in oligodendrogliomas. The gene set enrichment analysis indicated that there is a clear impact of 1p/19q co-deletion on the proteome and that many proteins are downregulated. However, it also revealed that there is a fraction of proteins not significantly regulated or even upregulated in oligodendrogliomas.

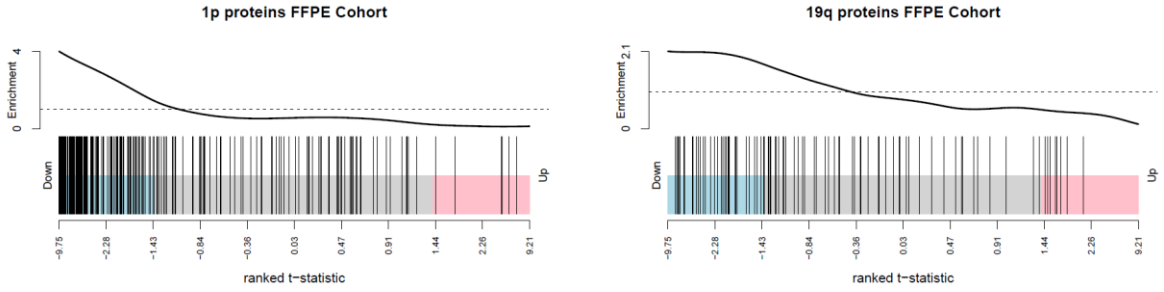


**Figure 14: Analyses of 1p and 19q coded proteins in astrocytoma and oligodendroglioma**. Bar plots of gene set enrichment analysis using all identified proteins found on 1p (left) and 19q (right). Proteins are represented as bars and placed according to their DAP ranks. On both 1p and 19q an enrichment of downregulated proteins (blue region) is visible.

#### 3.1.2.7 Determination of Chromosomal Wide Protein Abundances

Based on the profound impact of 1p/19q co-deletion on the proteome of IDH-mutant gliomas, the possibility of emulating a chromosomal copy number profile based on proteomic data was explored.

Considering the relative nature of the protein quantification and an unavoidable variability of clinical samples, a “Chromosomal Protein Ratio” (CPR) as a specific type of normalization was calculated. For this, log2 protein abundances deriving from all chromosomal arms that where determined as unaltered by DNA methylome data (median absolute deviation, MAD), were selected to calculate a mean background abundance for each protein. Importantly, chromosomal arms with less than 50 quantified proteins were not considered to avoid the impact of outliers and missing values. The log2 abundances of each sample were then subtracted by the background to calculate CPRs. Next, we aligned the CPR scores for each protein along the chromosomal position of the coding gene. A moving average (window size =10 CPRs) was applied to smooth-out outliers. To reduce individual sample variance, CPRs were centered by the mean CPR of its sample.

To reduce the complexity of all protein abundances coded by a chromosomal arm to a single dimension, the mean of all smoothed CPRs for each chromosome arm were calculated. An overview of the resulting CPR-means for chromosomes 1-22 as a heatmap obtained from the FFPE cohort is shown in Figure 15a. 1p/19q co-deletions in oligodendroglioma were readily visible. Additionally, hierarchal clustering using only the CPR-means of 1p and 19q was performed (Figure 15b) to check if those are discriminative enough to achieve a clear separation between astrocytomas and oligodendrogliomas. Here, a clear distinction of astrocytomas and oligodendrogliomas except four astrocytomas samples clustering together with oligodendrogliomas was possible. This approach showed a similar discriminative power as the specific protein signature (Figure 8f). By considering partial losses of the 1p and 19q arms, it was found that one of the misclustered astrocytomas had a partial loss on 19q and an additional partial loss on 1p (indicated by “X” in Figure 15b). Interestingly, low CPR-means in 19q were also observed in a subgroup of astrocytomas harboring partial or complete losses of 19q.

Because the CPR-means of oligodendrogliomas consistently reflected the status of

1p and 19q and extended analyses were performed on 4q. Here, a loss of 4q is present in about 30% of oligodendrogliomas and may also occur in IDH-mutant astrocytomas, especially in higher grade tumors [7]. In order to assess an association to the chromosomal status the CRP-means for 4q ordered from high to low were plotted. The association of CRP-mean and chromosomal status is shown in Figure 15c

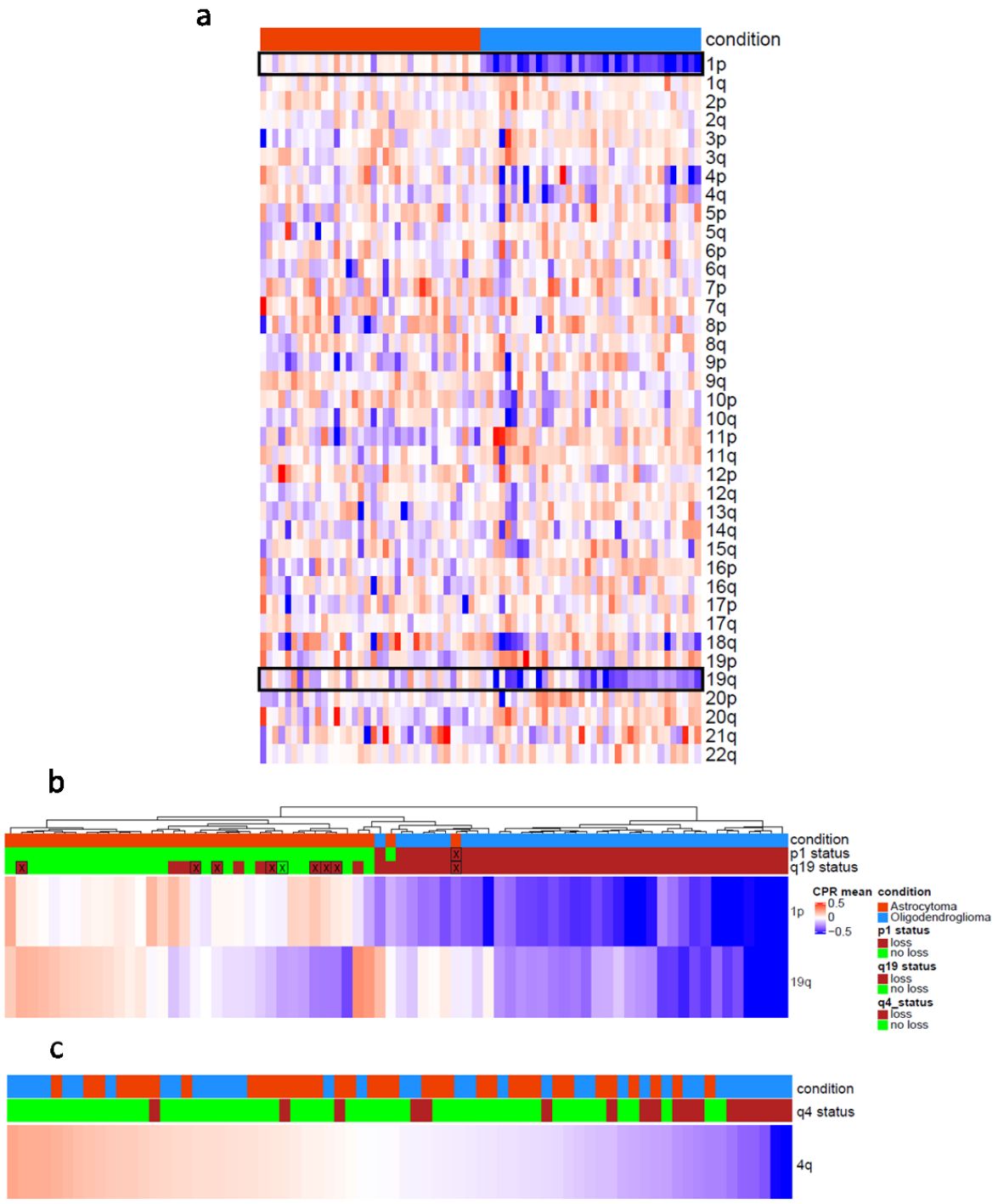


**Figure 15: Analyses of chromosomal wide protein abundances. a**: Global heatmap showing CPR-means across all chromosome arms. Conditions were sorted manually in astrocytoma (left) and oligodendroglioma (right); FFPE cohort. **b**: Hierarchical clustering of the FFPE tissue cohort using CPR-means of chromosome arms 1p and 19q. "X" indicates partial loss in the given chromosome arm. **c**: Hierarchical clustering of the FFPE tissue cohort using CPR-means of chromosome arm 4q. Ratios are sorted from high to low from left to right. Likelihood of a q4 loss status rises with lower CPRs and is independent from tumor diagnosis.

### 3.1.2.8 Prediction of Chromosomal Copy Number Variations

The CPRs were used to generate virtual copy number variation plots. Using the CPRs of unaltered chromosomal arms as control distribution, chromosomal alterations were inferred where CPRs were lower or higher than the 1.2-fold standard-deviation from the control distribution. Direct comparison with segment CNV plots derived from the DNA methylation profile of the same tumor, showed a high correlation between shifts predicted from the proteome and actual DNA copy number variations (Figure 16).

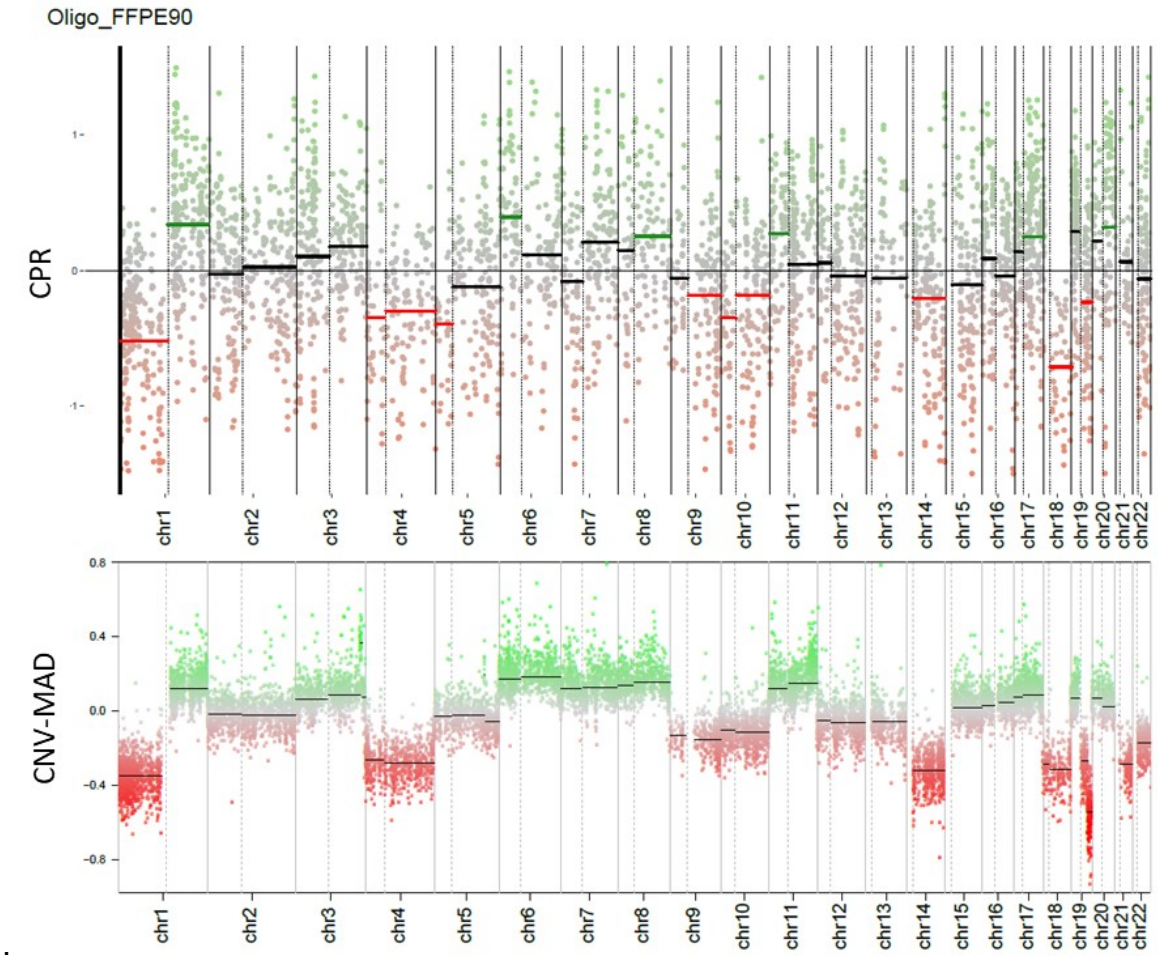


**Figure 16: Proteomic prediction of chromosomal copy number alterations**. Segment plots from an oligodendroglioma (FFPE tissue): **Top**: Segment plot of CPRs derived from the proteomic data. Horizontal lines on each chromosome arm represent CPR-means for proteins of the according arm. Gains and losses exceeding the chromosome arm specific threshold are marked through green (gains) and red (loss) lines. **Bottom**: CNV-Segment plot derived from DNA-methylation data. Chromosome arms in both segment plots are numerally ordered from left to right, chromosome arms p and q are divided by a dotted line for each chromosome.

Finally, circos plots termed “Chromosomal Protein Ratio Plot” (CPRP) were generated and compared with circos CNV plots (Figure 17). For oligodendrogliomas the consistency in FFPE samples between CPRP and CNV was 97% for 1p and 75% for 19q. Other chromosomal alterations had accuracies of 67-97%, (mean 84%, SD 6.7%). For astrocytomas the accuracies spanned for all chromosome arms from 61-92%, (mean 80% SD 7,8%).

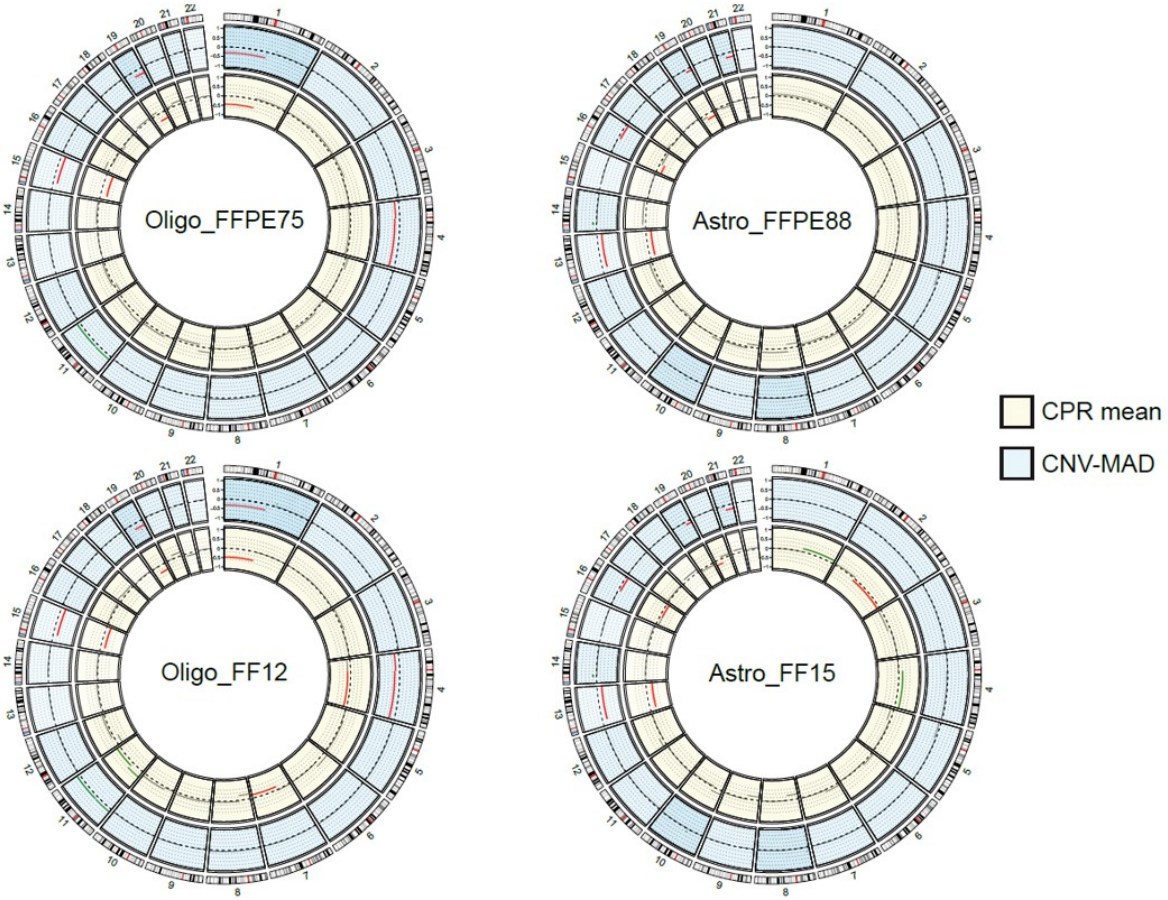


**Figure 17: Chromosomal Protein Ratio Plot-CPRP.** Circos-plots of FF/FFPE sample pairs of oligodendroglioma (**left**) and astrocytoma (**right**). Outer ring depicts chromosome orientation with their according numbers, red lines mark the transition from p and q arms. Middle blue ring represents CNV-status. Inner yellow ring represents CPR-means. Gains and losses exceeding the chromosome arm specific thresholds on CNV and CPR-mean level are marked through green (gain) and red (loss) lines.

### 3.1.2.9 Detecting Abnormalities in Methylation Data with the Help of CPRPs

During the process of generating the CPRPs, a sample was identified that showed multiple mismatches between the CNV and CPR pattern. The methylation classifier

classified the tumor as an astrocytoma and the CNV plots showed no loss of 1p and 19q. In contrast, the CPR means showed a clear loss of 1p and slight loss of 19q. The clustering analysis, using the proteomic signature generated in chapter 3.1.2.2, clustered the sample clearly in to the oligodendroglioma cohort. Further inspection of the proteome data revealed that a high abundance of ATRX was detected which was subsequently confirmed by immunohistochemistry. The combination of all these findings led to a re-extraction of the tissue for DNA methylation sequencing. The presumption that the first extraction was conducted on a different tumor due to a mix up of samples was confirmed. Indeed, the second extraction resulted in a major improvement of matching hits between the CNV and CPR means as can be seen in Figure 18. The new methylation data revealed a CNV plot with a clear 1p/19q co-deletion and the classifier categorized it as an oligodendroglioma as well. Eventually this changed the integrated diagnosis of the tumor from an astrocytoma to an oligodendroglioma.

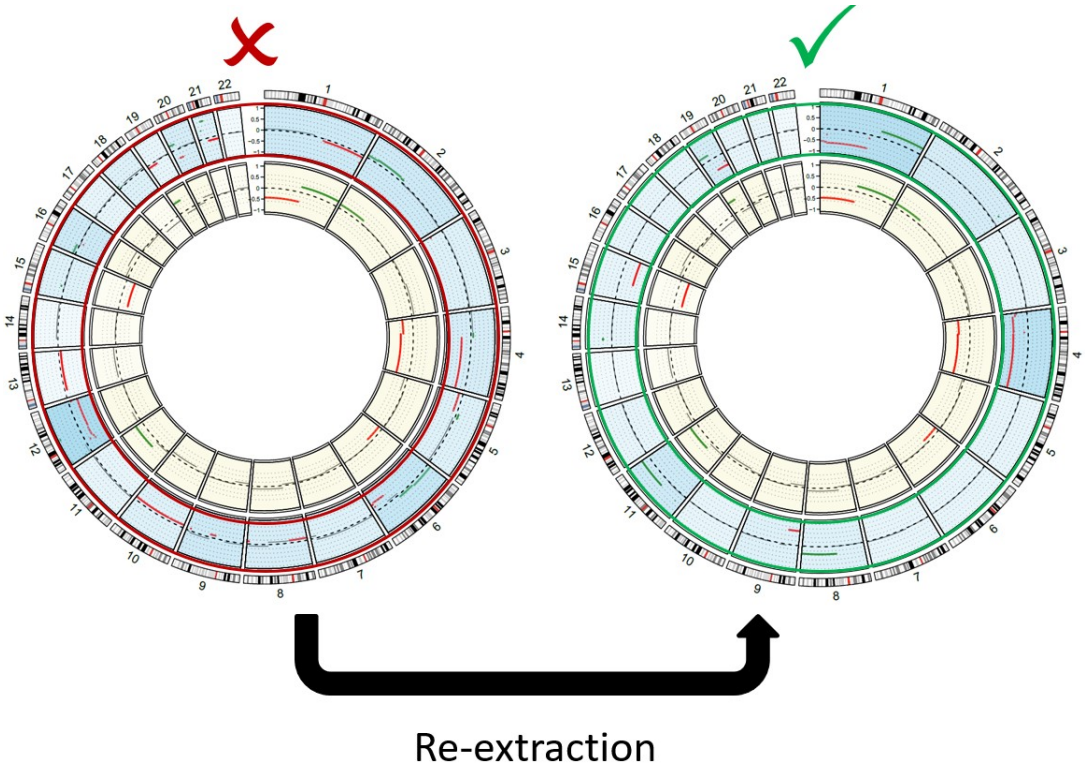


**Figure 18: Detecting abnormalities in methylation sequencing using CPRs**. **Left:** CPRP of an alleged astrocytoma since no 1p/19q loss was detected (red outlined blue ring). Barely any matches can be seen in any of the chromosome arms. Presence of a 1p and a slight 19q loss in the CPR mean data give hints for an oligodendroglioma (inner yellow ring). **Right:** CPRP of the same tissue with methylation data from the re-extraction (green outlined blue ring) which shows a clear improvement of matching to the CPR data (yellow ring)

## 3.2 Identification of a Potential Proteome Signature for Prognosis and Grading of IDH-mutant Astrocytoma

### 3.2.1 Experiment Overview

Since diffuse and anaplastic astrocytoma have no major differences in terms of overall survival (chapter 1.1.3), raises the question if a redefinition of the current grading system is possible by solely looking into protein abundances between different survival cohorts.

For this, FFPE peptide-extractions from the previous astrocytoma cohort (chapter 3.1.2) consisting of 18 IDH-mutant diffuse astrocytomas, 12 IDH-mutant anaplastic astrocytomas and 20 IDH-mutant glioblastomas were measured on the Exploris 480 platform as described in the subchapters of 2.2.9 and are further described as the “discovery cohort”. All samples had documented patient survival data starting from the beginning of treatment until the end of the observation (censor=0) or death (censor=1). Additionally, a validation cohort composed of 18 WHO grade III astrocytomas samples derived from the NOA04 clinical trial was included [77]. These samples were analyzed using our in-house proteomics facility using a Bruker timsTOF Pro coupled with a Bruker nanoElute HPLC. In short, after FFPE sample preparation, 400 ng of sample were loaded onto the nanoElute HPLC using a 2h gradient. The timsTOF Pro programming and downstream analysis were performed as in subchapters of 2.2.10. The quantitative protein data was extracted from the dataset and evaluated identically to the initial discovery cohort.

### 3.2.2 Optimization of the Grading for IDH-Mutant Astrocytomas

An initial Kaplan Meier plot shows the survival rates of the discovery cohort using current WHO grading systematic. As depicted in Figure 19a (left), WHO grade IV tumors show lower survival rates than WHO grades II and III. Among the grade II and III tumors no significant difference was observed (Figure 19,right). In order to find novel biomarker candidates, differential expression analysis was performed as described in chapter 3.1.1.1. Since a clear separation between WHO grades II/III from grade IV was visible, t-tests between WHO grade II vs. IV and grade III vs. IV were performed. Resulting p-values were adjusted via Benjamini-Hochberg procedure and proteins below a p-value of 0.01 and a fold change higher than 1.5 were considered significantly deregulated. The 71 overlapping proteins from both differential expression analyses, were then used for k-means clustering (k=2). The resulting two survival clusters were then used as conditions in the survival analysis. As can be seen in Figure 19b (left) a tendency of separation is noticeable between short- and long-term survivors that is similar to the existing WHO grading system. In order to assess how well the differentiation between WHO grade II and III was, WHO grade IV tumors were left out of the Kaplan Meier plot as depicted in Figure 19b (right). Here we see an almost identical separation in terms of survival as in the existing WHO grading system between WHO grade II and III (Figure 19a, right). However, a more significant separation of the survival cohorts in the validation cohort could be seen (Figure 19a, bottom).

In summary both the discovery and validation cohort show tendencies of identifying short- and long-term survival cohorts using the overlapping differential protein expression signature, but no significant change in the separation efficiency could be seen compared to the WHO grading.

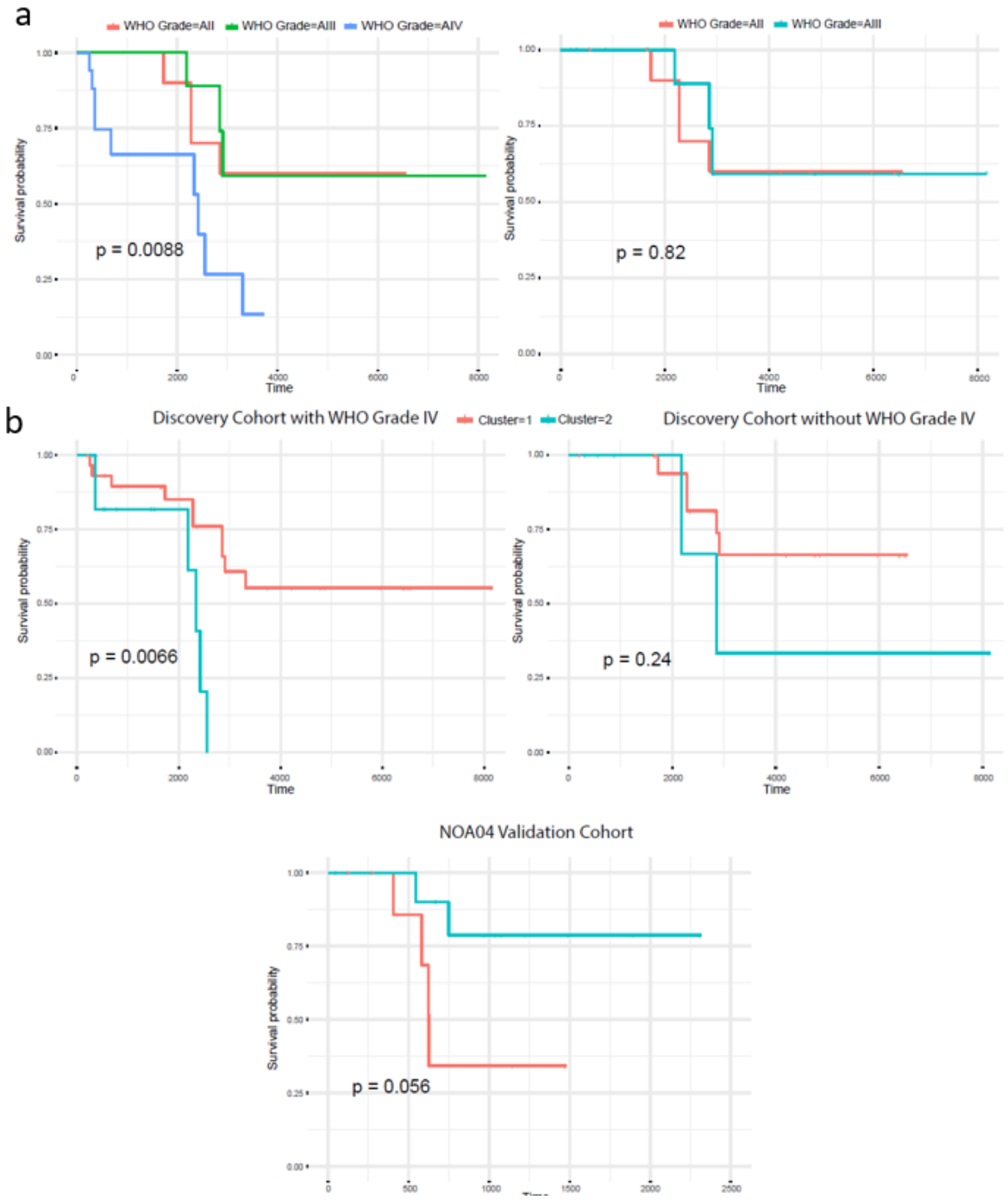


**Figure 19: Cluster analysis using significantly deregulated proteins between WHO grades II, III and IV tumors. a:** Survival rates of the astrocytoma discovery cohort using the current WHO grading. **Left:** Kaplan-Meier plot of the complete cohort with WHO grade IV tumors included. **Right:** Kaplan-Meier plot only using WHO grade II and III of the discovery cohort. **b:** After differential expression analysis between WHO grade II vs. IV and III vs. IV from the discovery cohort, overlapping proteins reaching the defined threshold of p-value <0.01 and foldchange >1.5 were used for k means clustering (k=2) of z-scored abundances in both the discovery cohort as well as the NOA04 validation cohort. Resulting clusters were used as condition in the Kaplan Meier analysis (**Left:** complete discovery cohort. **Right:** Discovery cohort without WHO grade IV tumors. **Bottom**: NOA04 validation cohort).

Since in comparison to the WHO grading a similar separation behavior was observed, the scope in defining tumor cohorts was redefined. Ignoring the WHO grading of these tumors completely, a search for proteins with abundance levels correlating the best with the survival rates all patients was carried out. For this, each protein was assigned an abundance cut point, which defined when a protein abundance was low or high. In more detail, the samples were ordered from low to high abundance and a cut point defined that separated the samples into two cohorts where the survival rate separation of the resulting cohorts was the most significant. The procedures for this were carried out using the R-Package “maxstat” with the command “surv_cutpoint”.

The discovery dataset revealed 317 proteins with a p-value <0.01. Although many proteins were found, survival plots of the respective abundance cohorts revealed only a small group of proteins which divided cohorts in a properly fashion. Especially a subunit of the protein complex ATP synthase (ATP5S) showed a distinct separation of long and short-term survivor cohorts. Next to ATP5S, another 11 out of the 16 ATP synthase subunits were identified in the data set as well which showed similar tendencies of separation. Next, k-means clustering (k=2) of z-scored protein abundances of all 12 subunits found was carried out in order to define two survival clusters. WHO grade IV tumors were left out of the clustering analysis in the discovery cohort, since it was more attractive to see how well WHO grade II and III could be separated. The resulting survival clusters with the 12 subunits found, are depicted in Figure 20. The NOA04 validation cohort was clustered separately using the same ATP synthase subunit compilation. The resulting clustering analyses and Kaplan-Meier plots are shown in Figure 20 revealing an improved separation in the Kaplan Meier plots in the discovery cohort in comparison to Kaplan Meier plots based on the WHO Grading (Figure 19a, right). The Kaplan Meier plot of the NOA04 validation cohort showed similar results where only astrocytoma WHO grade III were present.

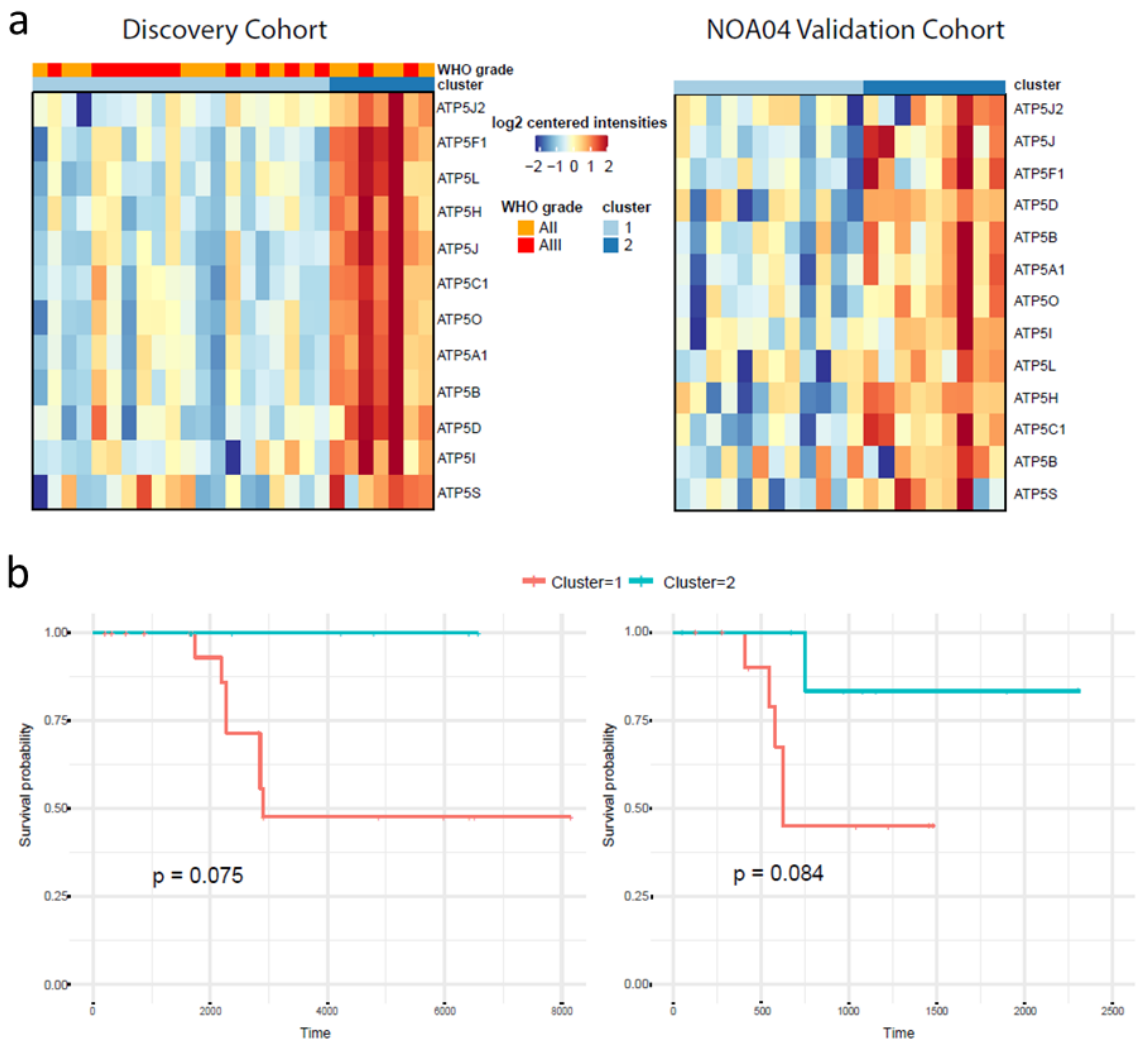


**Figure 20: Cluster analysis using ATP-synthase subunits as a prognostic biomarker for IDH-mutant astrocytomas. a:** k-means clustering analysis using z-scored abundances of 12 of the of 16 subunits of the ATP synthase complex from the discovery cohort (left) and NOA04 validation cohort (right). **b:** Kaplan-Meier plots of the discovery cohort (left) and NOA04 validation cohort (right) using the assigned cluster cohorts from **a.**

# 4. Discussion

## 4.1 Linking Proteomics with Current Knowledge of IDH-Mutant Glioma

With the introduction of molecular parameters for the diagnosis of IDH-mutant gliomas, neuropathologists are obligated to acquire genetic information of the patient's tumor in order to find a final integrated diagnosis (chapter 1.1.2). Although the precision of diagnostics has dramatically improved with these additional parameters, the dependency of using complex diagnostical infrastructures has also increased the effort to obtain the correct diagnosis. Strives in trying to find antibodies as diagnostic candidates to substitute genetic analysis have partially helped in being less independent from these workflows, such as the discovery of antibodies for ATRX, IDH1-R132H and BRAF-V600E [1, 20, 78]. Proteomics shows a great opportunity to find more biomarker candidates to distinguish between various tumor entities by histochemistry since protein abundance on the proteomic level can be semiquantitative linked to staining intensities of antibodies [79]. So far, comparative multi-omic investigations on IDH-WT and IDH mutant glioma were carried [80, 81]. However, elaborative investigations using larger cohorts of IDH-mutant glioma are so far missing. The presented study shows how proteomic investigations usefully combined with genetic profiling can reveal potential biomarkers for diagnostics and grading of IDH-mutant glioma, that were so far not detected as significant or further investigated using DNA und mRNA sequencing [82, 83].

### 4.1.1 Identification Rates between FFPE and FF Samples

As previously mentioned in chapter 1.2.1.1, FFPE samples have a highly diagnostic and medical-scientific value, since they preserve biological information of the sample over long periods of time. In addition, these sample are well documented and have valuable clinical information such as overall survival. FF samples maintain over time a high quality of biological information as well, but due to space-limiting preservation requirements in liquid nitrogen or -80°C freezers, sample numbers are limited. It was therefore of high interest to establish an FFPE sample preparation protocol suitable for LC-MS proteomics in order to have access to these sample types for medical-scientific investigations.

First, a fresh frozen protocol was established since it is generally simpler and better established in the proteomic field than FFPE sample preparations. The resulting data from the initial fresh frozen cohort showed accordance to previous studies from astrocytoma and oligodendroglioma, such as the detection of the diagnostically relevant biomarker ATRX as a significantly deregulated protein (chapter 3.1.1.1) [1]. We therefore proceeded in establishing a FFPE protocol which would be compatible for tissue punches to harmonize extraction practices which are commonly used for genomic investigations in the neuropathology. The established FFPE protocol was compared with the FF protocol by evaluating the protein identifications and quantities originating from LC-MS measurements. Identification rates on the peptide and protein level were compared and reported LFQ quantities of overlapping protein IDs in FF-FFPE sample pairs evaluated, using FF and FFPE tumor tissues which derived from the same patient.

As described in chapter 3.1.1.2, higher identification rates with lower fluctuations between the samples could be seen on the peptide and protein level in FF tissue compared to FFPE tissue. Similar identification rates and fluctuations have been described in the literature [74, 84], likely attributed due to the different fixation times between FFPE blocks that occur during general working procedures in the routine lab and result in different degrees of cross linking. Occasionally, certain FFPE samples show fixation artefacts that can be seen when carrying out immunohistochemistry [85]. In this case, proper staining only occurs on the boarders of the tumor and the center show less to no positivity. This results from an insufficient fixation where only the surface of the tumor was penetrated with formalin and the core of the tumor left prone to degradation. This could lead to uncontrolled peptide cleavages from various endogenous proteases which result not only inadequate immunohistochemistry staining, but also possible lower identification rates on the proteome level. Another possibility is that the sample preparation technique for FFPE tissue has more variability concerning peptide extraction from the tissue. Compared to the sample preparation for FF tissue, the main differences are the initial deparaffinization steps using the bead mill and 1h lysis in the Barocycler (chapters 2.2.5 and 2.2.7.2). Although possible, it is still rather unlikely that these steps are the reason for this kind of fluctuation. This is supported by the fact that the identification fluctuation of the protocol was barely visible when comparing the

neuropathology protocol to a commercially available protocol using the same tumor block. In addition, a separate study comparing FFPE protocols showed similar fluctuations where three different tumor areas were extracted with the same protocol. In contrast, similar identification rates could be seen when comparing different protocols using the same tumor region [74]. Therefore, the variability of tissue quality and fixation is more likely to be the reason for identification fluctuations.

Although lower identification rates were seen in FFPE tissue compared to FF tissue, the label free quantification of proteins was very similar and also comparable to other protocols showing similar correlations [75, 86]. Normally, in order for a protein quantification to be precise, at least two peptides should be identified for the given protein [53]. In some cases, this could mean that a certain protein in the fresh frozen cohort was possibly quantified using a higher number of peptides and the same protein in the FFPE cohort with only two peptides. Although a different number of peptides were then used for quantification, both the FF and FFPE could have reported similar LFQ intensities for the protein.

### 4.1.2 Protein Signature for IDH-Mutant Gliomas

By comparing the DAP in FF and FFPE tissue of IDH-mutant gliomas two sets of protein signatures to distinguish astrocytomas and oligodendrogliomas were acquired (chapter 3.1.2.1 and 3.1.2.2). Since the number of FFPE samples was much higher than those of the FF cohort, more proteins passed the significance threshold in the FFPE cohort. Although there were some significant proteins in the FFPE cohort which were just below the significance threshold in the FF cohort, a rather conservative filtering was maintained in order to have a robust protein signature. The correct clustering of 4/5 “rare oligodendrogliomas” (Figure 8f) which were not part of the differential expression analysis showed that the signature represents more the 1p/19q status of the tumor, rather than the DNA methylation profile.

Although a so-called “rectangular strategy” [87] was applied in finding this signature, an additional verification cohort with larger dimensions is necessary to solidify our results. Later on, combined with targeted approaches such as parallel reaction

monitoring, a high throughput diagnostic pipeline could be established, where only a subset of proteins from the discovered signature would be used to distinguish between astrocytomas and oligodendrogliomas. However specific thresholds for protein abundances need to be adapted in order to define when a certain protein amount is characteristic for a certain entity. In general, this type of discovery approach will most likely be applied more often in the future, when looking into other brain tumor types where diagnostic biomarkers are scarce. Even beyond, classification algorithms such as the methylation classifier developed in the Heidelberg neuropathology [88] are foreseeable approaches in the proteomic field, which will additionally help diagnosing tumors and most possibly be a fundament in discovering new tumor entities.

### 4.1.3 DAP as Biomarker Candidates for IDH-Mutant Gliomas using Immunohistochemistry

*Chapter 4.1.3 contains text passages that have been taken from Felix et al. "HIP1R and VIM immunohistochemistry predict 1p/19q status in IDH-mutant glioma" (manuscript in preparation).*

Quantitative proteome analyses of both the FF and FFPE astrocytoma and oligodendroglioma cohorts in chapter 3.1.2.1 (Figure 8a, b) revealed a subset of differentially expressed proteins which were further investigated by immunohistochemistry. HIP1R which was found upregulated in oligodendrogliomas is a relative of huntingtin-interacting protein 1 (HIP1), which is highly present in several different tumor types and which has transforming activity in vitro [89, 90]. While this is not reported for HIP1R there seems to be at least some functional overlap with both proteins being involved in receptor tyrosine kinase stabilization and endocytosis [91]. In contrast, newer findings indicate that HIP1R plays a major role in lysosomal degradation of PD-L1 which could hinder the evasion from T cell–mediated immune surveillance [92]. If this correlates with the general higher survivability of oligodendrogliomas compared to astrocytomas is questionable, since PD-L1 expression is rather low in these tumors due to the increased methylation status of its promoter [93]. Consistent with our results, HIP1R was recently found downregulated in astrocytomas compared to oligodendrogliomas on the mRNA level in a gene expression meta-analysis of histology-defined low-grade gliomas

[82].

VIM is well known to be expressed in astrocytic brain tumors which functions as an intermediate filament and has been already discovered in earlier proteomics studies to be downregulated in oligodendroglioma [94]. A very recent study revisited VIM expression patterns with molecularly defined tumors and suggested it as a biomarker for distinguishing oligodendroglioma from astrocytoma [95].

Another intermediate filament named SNYM was also upregulated in astrocytomas which could also interact with VIM, alters the dynamics of the actin cytoskeleton and therefor changes the motility of astrocytic tumor cells [96]. Newer findings show that SYNM modulates radiation sensitivity and DNA double strand break repair in carcinoma cells [97]. Another highly upregulated protein was TNC, a extracellular matrix protein which is known for its many roles in various diseases of the central nervous system and in tumorigenic properties in gliomas [98].

### 4.1.4 VIM/HIP1R Expression as a Dual Surrogate Biomarker for 1p/19q Status

*Chapter 4.1.4 contains text passages that have been taken from Felix et al. "HIP1R and VIM immunohistochemistry predict 1p/19q status in IDH-mutant glioma" (manuscript in preparation).*

In order to determine 1p/19q statuses, array-based techniques such as 850K or SNP arrays are needed to obtain information of multiple positions across the chromosomal arms. However, these techniques are not available everywhere, are expensive and time consuming. FISH-based 1p/19q determination is therefore most widely used but has its limitations regarding probe hybridization efficiency, aneuploidy and the inability to differentiate between segmental and whole arm deletion [99]. Therefore, there is a high motivation in finding immunohistochemical surrogate biomarkers for 1p/19q.

The loss of function biomarker ATRX closely associated with the mutation or deletion of the *ATRX* gene(s) became an established marker for 1p/19 retained IDH-mutant gliomas. However, retained ATRX is insufficient to classify an IDH-mutant glioma as an oligodendroglioma since a significant fraction of astrocytomas have a retained ATRX expression as well. Alpha-internexin (INA) has have been previously described as a potential biomarker for oligodendroglioma but multiple studies

showed limited specificity [100-103]. LC-MS results in this study revealed an increased abundance of INA in oligodendroglioma which however did not reach statistical significance. Reduced amounts of H3K27me3 in a vast majority of oligodendroglioma was been reported in immunohistochemical and proteomic studies [104, 105]. However other studies have shown that its usefulness as a surrogate marker for 1p/19q co-deletion is limited due to a fraction of astrocytomas which show a lack of H3K27me3, a reported selectivity for IDH1-R132H mutant oligodendrogliomas and intra-laboratory variability of results [106, 107].

The proposed dual surrogate Biomarker using HIP1R/VIM for 1p19q determination aims at recognizing tissue patterns which are more robust and reproducible than purely quantifying a single antibody staining. Evaluating a single antibody stain can be influenced by unknown and variable factors such as fixation time or variability in perception of intensity between different observers. Since a tissue punch was used for tumor extraction and further downstream proteomic analysis, staining intensities of HIP1R and VIM should only be evaluated in the same tumor area which further reduces factors of variability in staining interpretation.

In Summary, the data in chapter 3.1.2.5 shows that a correct classification of astrocytomas and oligodendrogliomas is possible using the three immunohistochemical markers ATRX/HIP1R/VIM reducing the need for molecular analyses for all IDH-mutant gliomas up to 95%.

### 4.1.5 Chromosome Status and its Effect on the Proteome

During the investigation in finding a protein signature to distinguish between astrocytomas and oligodendrogliomas, a protein candidate list was acquired that had an unusual high percentage of significantly deregulated proteins coded on the 1p and 19q arm (chapter 3.1.2.6). Almost all of these proteins were significantly down regulated in oligodendrogliomas and were consequently an observation that was affected by the 1p deletion in oligodendrogliomas. The effect was also seen for proteins coded on 19q, but to a less extent. This was most likely due to a fraction of astrocytoma samples in the series harboring partial or complete losses of 19q, which is commonly seen in these tumors [108]. By generating a baseline abundance for each protein, abundance ratios for each protein were calculated. In summary, the

ratios combined in form of a mean, reflected the genetic chromosome status for each tumor with a high accordance. Still, some cases were observed where the chromosome status on the protein level did not correlate well with the genetic status. These observations were mostly seen on small chromosome arms where only a few candidates were found to calculate the CPRs. Additionally, if these arms were more often than usual deregulated on the genetic level, less proteins abundances were available to calculate a mean abundance for an unregulated chromosome arm. This greatly increased the error for each mean and therefore affected the actual outcome of the CPRs. Tumor heterogeneity could also have been a reason for a part of the observed discrepancies since proteins and DNA were not necessarily extracted from the same tumor region. In the future, larger dataset will be acquired in order to refine the method and extend the approach to other tumor entities.

Considering the data output of the CPRs greatly synergizes with already given data repositories in the neuropathology, there is a great motivation in integrating this type of data output into already existing diagnostic infrastructures. Chromosomal alterations are not only used as a signature biomarker in oligodendrogliomas but also show characteristic patterns in other entities such as the gain of chromosome 7 and loss of chromosome 10 in IDH-WT glioblastomas which we could detect in an independent experiment [109, 110]. Another study investigated copy number variations and the effect on the proteome in cell cultures. Instead of using a mean proteome abundances for ratio calculation, fold changes of relative protein levels between normal and cancer cells were calculated and compared with the fold changes of genetic copy numbers of their respective gene [111].

Although the diagnostic pipeline for methylation sequencing is well established, there are rare cases where the output does not fit well other known information of the tumor. CPR-plots could later on be used as a quality control to make sure that the genetic data is truly originating from the tumor it is linked to in the database (chapter 3.1.2.9). This can potentially improve the annotation of samples with correct diagnoses and prevent systematical errors in statistical analyses due to heterogenous tumor cohorts.

### 4.1.6 Identification of Prognostic Biomarkers in IDH-Mutant Astrocytoma

As previously mentioned, the WHO grading of astrocytoma remains a challenging aspect in brain tumor diagnostics due to its low correlation with prognosis (chapter 1.1.3). Although a clear distinction in survival can be seen in WHO grade IV astrocytoma compared to grade II and III, WHO grade II and III show only small differences [24]. Therefore, an investigation to find novel prognostic biomarkers using the previous FFPE astrocytoma cohort from chapter 3.1.2 was carried out and results verified using a separate cohort of WHO grade III astrocytoma derived from the NOA04 clinical study [77]. In an attempt to find prognostic biomarkers for IDH-mutant astrocytomas, differential expression analyses of WHO grade II, III and IV tumor proteomes were carried out (chapter 3.2.2). By comparing grade II with grade IV and grade III with grade IV tumors a list of overlapping significantly deregulated proteins was identified which was later used for defining survival cohorts by k-means clustering. Since survival rates for WHO Grade II and III are rather similar compared to Grade IV tumors, benchmarking of newly defined cohorts was only based on WHO grade II and III tumors. The resulting Kaplan Meier plots revealed no true improvement in comparison to the plots based on current WHO grading. This was most likely due to the criteria for how the proteins were selected for defining survival cohorts. Instead of searching for proteins which were significantly deregulated based on the current WHO grading of astrocytoma, a redefinition of survival cohorts was necessary.

By defining for each protein an optimal abundance cutoff for long- and short-term survivor cohorts, a set of biomarker candidates reaching high significance was revealed. Some candidates showed association with tumor progression such as components of the minichromosome maintenance complex, TNC and myristoylated alanine-rich C-kinase substrate [112-114]. Still, manual inspection of the Kaplan Meier plots revealed only a subset of proteins such as the subunit ATP5S from the ATP synthase which showed a proper separation of the two survival cohorts. This was most likely due to a small number of events in the survival data which made significance testing more challenging. Eleven other subunits of the ATP synthase were identified as well by LC-MS, with similar abundances in each patient. This

indicated that not only the one subunit alone is deregulated but probably the whole complex, which could impact the function of the protein complex to a higher degree [115]. Although the protein abundance of a single subunit and its calculated cut point could be used for a Kaplan Meier analysis, a more robust separation technique is favorable. This resulted in the definition of survival cohorts based on k-means clustering of z-scores of all 12 subunits found, revealing a potential biomarker for improved grading of WHO grade II and III tumors.

Although the results were intriguing, both the discovery and validation cohort had low number of events of death (censor= 1) which made the overall analysis lack in statistical power. Additionally, a more systematic approach for defining a threshold for high and low abundance of the complex should be carried out in the future since both clustering analyses are based z-scoring and not defined protein abundance cut-offs. Next to additional proteomic analyses, IHC screenings will be also carried out in the future in order to validate these findings. For this another validation cohort is currently being acquired.

However, studies on carcinomas have already indicated that an impaired bioenergetic functioning of mitochondria is linked with tumor progression and prognosis [116-118]. This corresponds with the presented data and makes this finding an attractive approach in further understanding tumor metabolism in IDH-mutant astrocytoma.

### 4.1.7 Energy Metabolism and its Relevance for Tumor Prognosis

Identifying the downregulation of subunits from the ATP synthase in astrocytomas with lower survival sheds light in the area of tumor metabolism that has often been discussed in cancer biology of carcinoma termed the Warburg Effect [116, 119]. Otto Warburg discovered that tumors have an increased rate of aerobic glycolysis resulting from an impaired bioenergetic activity of mitochondria that is independent from the activation of oncogenes and inhibition of tumor suppressors [120-122]. Nonproliferating cells tend to have a low consumption of glucose since the oxidative phosphorylation provides a high yield of ATP and NADH. During cell proliferation there is a higher demand for metabolic intermediates and NADPH that are obtained by glycolysis and the pentose phosphate pathway [123]. Furthermore, the resulting

pyruvate from the glycolysis is oxidized to lactate in order to regenerate NADH to NAD+ to maintain the glycolytic flux. The excess lactate results in an acidification of the cellular microenvironment and promotes metastatic processes in carcinoma [124].

There are multiple ways of how the ATP synthase can be downregulated in order to possibly acquire the lower oxidative phosphorylation state. The translation of ATP synthase can be lowered through RNA silencing by bonding with other proteins such as Ras GTPase-activating protein-binding protein 1 [116]. Indeed, a proportion of samples with low abundance of ATP synthase subunits did have higher amounts of G3BP1, showing that there might be a possibility of RNA silencing. Additionally, hypermethylation of the promotor for ATP5 subunits could have led to lower transcription rates. However, next to the inhibition of the DNA demethylase by 2-hydroxygluturate which results from the IDH mutation, an additional methylation pathway would be needed, otherwise we would not see a difference in IDH-mutant glioma [158]. Another study has shown that 2-hydroxyglutarate inhibits the ATP synthase directly which supposedly decreases tumor cell growth and viability [125]. However, our findings reveal that the downregulation of the ATP synthase are associated with a shorter survival. In this case, the downregulation of the ATP synthase could provoke tumor cells to create ATP more through glycolysis and less through oxidative phosphorylation leading to increased cell proliferation. Additionally, since less protons are transferred back through the ATP-synthase into the matrix, the intermembrane space could be saturated. This could result in a higher resistance to transport protons against the concentration gradient leaving more NADH in the matrix of the mitochondria. NADH could then be used to regenerate glutathione and therefore result in a higher tolerance towards increased ROS levels which are typical for IDH mutant glioma [27, 126].

If an impaired oxidative phosphorylation metabolism in astrocytoma leads to a worse prognosis, it could later on influence the grading of astrocytoma. This would improve the accuracy for astrocytoma grading, leading to a better correlation with prognosis compared to current diagnostic criteria where WHO grade II and grade III are rather similar [21, 24].

## 4.2 Conclusion and Outlook

Within the course of this thesis a sensitive, fast and robust quantitative proteomic mass spectrometry workflow was established and applied on diagnostically relevant questions concerning IDH-mutant glioma. A proteomic signature for distinguishing between astrocytoma and oligodendroglioma was discovered that followed the diagnostically relevant chromosomal copy number status of 1p and 19q. Further extensive immunohistochemistry screenings of proteins from this signature revealed VIM and HIP1R as a possible dual surrogate biomarker for 1p/19q co-deletions in IDH-mutant glioma. These biomarkers and in combination with ATRX could later on make diagnostics of IDH-mutant glioma less dependent on technical complex genetic analyses. Further investigations of the proteomic signature revealed an enrichment of 1p- and 19q-coded downregulated proteins which highlighted the impact of gene copy numbers on the proteome. This finding led to the development of the CPR-Plots showing high correlations with genetic CNVs. These plots could be used in the future as a quality control for the genetic databases in terms of sample annotation and therefore confirm tumor cohort homogeneity. In addition, investigations of astrocytomas with clinical data revealed possible prognostic biomarkers consisting of 12 subunits from the ATP synthase which could be implemented in to the grading of astrocytoma and further improve the correlation of tumor grading with prognosis. In summary, this thesis shows that proteomics can be a powerful approach in identifying and quantifying protein signatures in brain tumors and help in revealing diagnostic and prognostic relevant patterns, but also complement genetic investigations that are practiced in the neuropathology.

In the future more, proteomic investigations beyond IDH-mutant glioma will be carried out where diagnostics remains challenging. Looking forward, a possible proteome classification system could be developed which may reveal new tumor entities and improve diagnostics for brain tumors. Additionally, enrichment techniques such as for phosphorylated peptides will be implemented which could help in tumor therapy decision making for certain tumor entities such as glioblastoma exhibiting an unmethylated O6 methylguanine-DNA-methyltransferase promotor [127].

# List of Figures

# List of Tables

# References


1. Reuss, D.E., et al., *ATRX and IDH1-R132H immunohistochemistry with subsequent copy number analysis and IDH sequencing as a basis for an "integrated" diagnostic approach for adult astrocytoma, oligodendroglioma and glioblastoma.* Acta Neuropathol, 2015. **129**(1): p. 133-46.

2. Larance, M. and A.I. Lamond, *Multidimensional proteomics for cell biology.* Nat Rev Mol Cell Biol, 2015. **16**(5): p. 269-80.

3. DeAngelis, L.M. and I.K. Mellinghoff, *Virchow 2011 or how to ID(H) human glioblastoma.* J Clin Oncol, 2011. **29**(34): p. 4473-4.

4. Ferguson, S. and M.S. Lesniak, *Percival Bailey and the classification of brain tumors.* Neurosurg Focus, 2005. **18**(4): p. e7.

5. Louis, D.N., et al., *The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary.* Acta Neuropathol, 2016. **131**(6): p. 803-20.

6. Weller, M., et al., *European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas.* Lancet Oncol, 2017. **18**(6): p. e315-e329.

7. Balss, J., et al., *Analysis of the IDH1 codon 132 mutation in brain tumors.* Acta Neuropathol, 2008. **116**(6): p. 597-602.

8. Parsons, D.W., et al., *An integrated genomic analysis of human glioblastoma multiforme.* Science, 2008. **321**(5897): p. 1807-12.

9. Hartmann, C., et al., *Patients with IDH1 wild type anaplastic astrocytomas exhibit worse prognosis than IDH1-mutated glioblastomas, and IDH1 mutation status accounts for the unfavorable prognostic effect of higher age: implications for classification of gliomas.* Acta Neuropathol, 2010. **120**(6): p. 707-18.

10. Cairncross, J.G., et al., *Specific genetic predictors of chemotherapeutic response and survival in patients with anaplastic oligodendrogliomas.* J Natl Cancer Inst, 1998. **90**(19): p. 1473-9.

11. Reifenberger, J., et al., *Molecular genetic analysis of oligodendroglial tumors shows preferential allelic deletions on 19q and 1p.* Am J Pathol, 1994. **145**(5): p. 1175-90.

12. Griffin, C.A., et al., *Identification of der(1;19)(q10;p10) in five oligodendrogliomas suggests mechanism of concurrent 1p and 19q loss.* J Neuropathol Exp Neurol, 2006. **65**(10): p. 988-94.

13. Jenkins, R.B., et al., *A t(1;19)(q10;p10) mediates the combined deletions of 1p and 19q and predicts a better prognosis of patients with oligodendroglioma.* Cancer Res, 2006. **66**(20): p. 9852-61.

14. Arita, H., et al., *Upregulating mutations in the TERT promoter commonly occur in adult malignant gliomas and are strongly associated with total 1p19q loss.* Acta Neuropathol, 2013. **126**(2): p. 267-76.

15. Koelsche, C., et al., *Distribution of TERT promoter mutations in pediatric and adult tumors of the nervous system.* Acta Neuropathol, 2013. **126**(6): p. 907-15.

16. Brosnan-Cashman, J.A., et al., *ATRX loss induces multiple hallmarks of the alternative lengthening of telomeres (ALT) phenotype in human glioma cell lines in a cell line-specific manner.* PLoS One, 2018. **13**(9): p. e0204159.

17. Cesare, A.J. and R.R. Reddel, *Alternative lengthening of telomeres: models, mechanisms and implications.* Nat Rev Genet, 2010. **11**(5): p. 319-30.

18. Sahm, F., et al., *Farewell to oligoastrocytoma: in situ molecular genetics favor classification as either oligodendroglioma or astrocytoma.* Acta Neuropathol, 2014. **128**(4): p. 551-9.

19. Capper, D., et al., *Characterization of R132H mutation-specific IDH1 antibody binding in brain tumors.* Brain Pathol, 2010. **20**(1): p. 245-54.

20. Capper, D., et al., *Monoclonal antibody specific for IDH1 R132H mutation.* Acta Neuropathol, 2009. **118**(5): p. 599-601.

21. Brat, D.J., et al., *cIMPACT-NOW update 5: recommended grading criteria and terminologies for IDH-mutant astrocytomas.* Acta Neuropathol, 2020. **139**(3): p. 603-608.

22. Olar, A., et al., *IDH mutation status and role of WHO grade and mitotic index in overall survival in grade II-III diffuse gliomas.* Acta Neuropathol, 2015. **129**(4): p. 585-96.

23. Reuss, D.E., et al., *IDH mutant diffuse and anaplastic astrocytomas have similar age at presentation and little difference in survival: a grading problem for WHO.* Acta Neuropathol, 2015. **129**(6): p. 867-73.

24. Shirahata, M., et al., *Novel, improved grading system(s) for IDH-mutant astrocytic gliomas.* Acta Neuropathol, 2018. **136**(1): p. 153-166.

25. Cimino, P.J. and E.C. Holland, *Targeted copy number analysis outperforms histologic grading in predicting patient survival for WHO grades II/III IDH-mutant astrocytomas.* Neuro Oncol, 2019. **21**(6): p. 819-821.

26. Phillips, J.J., et al., *PDGFRA amplification is common in pediatric and adult high-grade astrocytomas and identifies a poor prognostic group in IDH1 mutant glioblastoma.* Brain Pathol, 2013. **23**(5): p. 565-73.

27. Han, S., et al., *IDH mutation in glioma: molecular mechanisms and potential therapeutic targets.* Br J Cancer, 2020. **122**(11): p. 1580-1589.

28. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-3.

29. Smith, L.M., N.L. Kelleher, and P. Consortium for Top Down, *Proteoform: a single term describing protein complexity.* Nat Methods, 2013. **10**(3): p. 186-7.

30. Aebersold, R. and M. Mann, *Mass-spectrometric exploration of proteome structure and function.* Nature, 2016. **537**(7620): p. 347-55.

31. Olsen, J.V., S.E. Ong, and M. Mann, *Trypsin cleaves exclusively C-terminal to arginine and lysine residues.* Mol Cell Proteomics, 2004. **3**(6): p. 608-14.

32. Switzar, L., M. Giera, and W.M. Niessen, *Protein digestion: an overview of the available techniques and recent developments.* J Proteome Res, 2013. **12**(3): p. 1067-77.

33. Jarnuczak, A.F., et al., *Analysis of Intrinsic Peptide Detectability via Integrated Label-Free and SRM-Based Absolute Quantitative Proteomics.* J Proteome Res, 2016. **15**(9): p. 2945-59.

34. Giansanti, P., et al., *Six alternative proteases for mass spectrometry-based proteomics beyond trypsin.* Nat Protoc, 2016. **11**(5): p. 993-1006.

35. Y.HeinKirtiSharmaJürgenCoxMatthiasMann, M., *Handbook of Systems Biology, Chapter 1 - Proteomic Analysis of Cellular Systems*, ed. M.V.a.J.D. A.J. Marian Walhout. 2013: Academic Press.

36. Fox, C.H., et al., *Formaldehyde fixation.* Journal of Histochemistry & Cytochemistry, 1985. **33**(8): p. 845-853.

37. Schweiger, M.R., et al., *Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis.* PLoS One, 2009. **4**(5): p. e5548.

38. Kerick, M., et al., *Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity.* BMC Med Genomics, 2011. **4**: p. 68.

39. O'Rourke, M.B. and M.P. Padula, *Analysis of formalin-fixed, paraffin-embedded (FFPE) tissue via proteomic techniques and misconceptions of antigen retrieval.* Biotechniques, 2016. **60**(5): p. 229-38.

40. Donczo, B. and A. Guttman, *Biomedical analysis of formalin-fixed, paraffin-embedded tissue samples: The Holy Grail for molecular diagnostics.* J Pharm Biomed Anal, 2018. **155**: p. 125-134.

41. Fowler, C.B., T.J. O'Leary, and J.T. Mason, *Improving the Proteomic Analysis of Archival Tissue by Using Pressure-Assisted Protein Extraction: A Mechanistic Approach.* J Proteomics Bioinform, 2014. **7**(6): p. 151-157.

42. Guo, T., et al., *Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps.* Nat Med, 2015. **21**(4): p. 407-13.

43. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.

44. Fowler, C.B., et al., *Elevated hydrostatic pressure promotes protein recovery from formalin-fixed, paraffin-embedded tissue surrogates.* Lab Invest, 2008. **88**(2): p. 185-95.

45. Kobashigawa, Y., M. Sakurai, and K. Nitta, *Effect of hydrostatic pressure on unfolding of alpha-lactalbumin: volumetric equivalence of the molten globule and unfolded state.* Protein Sci, 1999. **8**(12): p. 2765-72.

46. Gao, H., et al., *Accelerated Lysis and Proteolytic Digestion of Biopsy-Level Fresh-Frozen and FFPE Tissue Samples Using Pressure Cycling Technology.* J Proteome Res, 2020. **19**(5): p. 1982-1990.

47. Nesvizhskii, A.I., O. Vitek, and R. Aebersold, *Analysis and validation of proteomic data generated by tandem mass spectrometry.* Nat Methods, 2007. **4**(10): p. 787-97.

48. The UniProt, C., *UniProt: the universal protein knowledgebase.* Nucleic Acids Res, 2017. **45**(D1): p. D158-D169.

49. Nesvizhskii, A.I. and R. Aebersold, *Interpretation of shotgun proteomic data: the protein inference problem.* Mol Cell Proteomics, 2005. **4**(10): p. 1419-40.

50. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment.* J Proteome Res, 2011. **10**(4): p. 1794-805.

51. Ong, S.E. and M. Mann, *Mass spectrometry-based proteomics turns quantitative.* Nat Chem Biol, 2005. **1**(5): p. 252-62.

52. Wichmann, C., et al., *MaxQuant.Live Enables Global Targeting of More Than 25,000 Peptides.* Mol Cell Proteomics, 2019. **18**(5): p. 982-994.

53. Cox, J., et al., *Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ.* Mol Cell Proteomics, 2014. **13**(9): p. 2513-26.

54. Prianichnikov, N., et al., *MaxQuant software for ion mobility enhanced shotgun proteomics.* Mol Cell Proteomics, 2020.

55. Hu, J., et al., *The importance of experimental design in proteomic mass spectrometry experiments: some cautionary tales.* Brief Funct Genomic Proteomic, 2005. **3**(4): p. 322-31.

56. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

57. Banerjee, S. and S. Mazumdar, *Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte.* Int J Anal Chem, 2012. **2012**: p. 282574.

58. Klampfl, C.W., *Review coupling of capillary electrochromatography to mass spectrometry.* J Chromatogr A, 2004. **1044**(1-2): p. 131-44.

59. Gelpi, E., *Interfaces for coupled liquid-phase separation/mass spectrometry techniques. An update on recent developments.* J Mass Spectrom, 2002. **37**(3): p. 241-53.

60. Yang, F., et al., *High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis.* Expert Rev Proteomics, 2012. **9**(2): p. 129-34.

61. Gillet, L.C., A. Leitner, and R. Aebersold, *Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing.* Annu Rev Anal Chem (Palo Alto Calif), 2016. **9**(1): p. 449-72.

62. Gillet, L.C., et al., *Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis.* Mol Cell Proteomics, 2012. **11**(6): p. O111 016717.

63. Ludwig, C., et al., *Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial.* Mol Syst Biol, 2018. **14**(8): p. e8126.

64. Zhang, F., et al., *Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020.* Proteomics, 2020. **20**(17-18): p. e1900276.

65. Doerr, A., *DIA mass spectrometry.* Nature Methods, 2015. **12**(1): p. 35-35.

66. Roepstorff, P. and J. Fohlman, *Proposal for a common nomenclature for sequence ions in mass spectra of peptides.* Biomed Mass Spectrom, 1984. **11**(11): p. 601.

67. Hunt, D.F., et al., *Protein sequencing by tandem mass spectrometry.* Proc Natl Acad Sci U S A, 1986. **83**(17): p. 6233-7.

68. Olsen, J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis.* Nat Methods, 2007. **4**(9): p. 709-12.

69. Syka, J.E., et al., *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry.* Proc Natl Acad Sci U S A, 2004. **101**(26): p. 9528-33.

70. Oh, S., et al., *Integrated pharmaco-proteogenomics defines two subgroups in isocitrate dehydrogenase wild-type glioblastoma with prognostic and therapeutic opportunities.* Nat Commun, 2020. **11**(1): p. 3288.

71. Clark, D.J., et al., *Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma.* Cell, 2019. **179**(4): p. 964-983 e31.

72. Petralia, F., et al., *Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer.* Cell, 2020. **183**(7): p. 1962-1985 e31.

73. Marchione, D.M., et al., *HYPERsol: High-Quality Data from Archival FFPE Tissue for Clinical Proteomics.* J Proteome Res, 2020. **19**(2): p. 973-983.

74. Coscia, F., et al., *A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis.* J Pathol, 2020. **251**(1): p. 100-112.

75. Zhu, Y., et al., *High-throughput proteomic analysis of FFPE tissue samples facilitates tumor stratification.* Mol Oncol, 2019. **13**(11): p. 2305-2328.

76. Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote)omics data.* Nat Methods, 2016. **13**(9): p. 731-40.

77. Wick, W., et al., *Long-term analysis of the NOA-04 randomized phase III trial of sequential radiochemotherapy of anaplastic glioma with PCV or temozolomide.* Neuro Oncol, 2016. **18**(11): p. 1529-1537.

78. Capper, D., et al., *Assessment of BRAF V600E mutation status by immunohistochemistry with a mutation-specific monoclonal antibody.* Acta Neuropathol, 2011. **122**(1): p. 11-9.

79. Carnielli, C.M., et al., *Combining discovery and targeted proteomics reveals a prognostic signature in oral cancer.* Nat Commun, 2018. **9**(1): p. 3598.

80. Djuric, U., et al., *Defining Protein Pattern Differences Among Molecular Subtypes of Diffuse Gliomas Using Mass Spectrometry.* Mol Cell Proteomics, 2019. **18**(10): p. 2029-2043.

81. Wang, L.B., et al., *Proteogenomic and metabolomic characterization of human glioblastoma.* Cancer Cell, 2021. **39**(4): p. 509-528 e20.

82. Wang, S., et al., *Gene expression meta-analysis in diffuse low-grade glioma and the corresponding histological subtypes.* Scientific Reports, 2017. **7**(1): p. 11741.

83. Weller, M., et al., *Molecular classification of diffuse cerebral WHO grade II/III gliomas using genome- and transcriptome-wide profiling improves stratification of prognostically distinct patient groups.* Acta Neuropathol, 2015. **129**(5): p. 679-93.

84. Davalieva, K., et al., *Comparative evaluation of two methods for LC-MS/MS proteomic analysis of formalin fixed and paraffin embedded tissues.* J Proteomics, 2021: p. 104117.

85. Chatterjee, S., *Artefacts in histopathology.* J Oral Maxillofac Pathol, 2014. **18**(Suppl 1): p. S111-6.

86. Marchione, D.M., et al., *HYPERsol: High-Quality Data from Archival FFPE Tissue for Clinical Proteomics.* Journal of Proteome Research, 2020. **19**(2): p. 973-983.

87. Geyer, P.E., et al., *Revisiting biomarker discovery by plasma proteomics.* Mol Syst Biol, 2017. **13**(9): p. 942.

88. Capper, D., et al., *DNA methylation-based classification of central nervous system tumours.* Nature, 2018. **555**(7697): p. 469-474.

89. Rao, D.S., et al., *Altered receptor trafficking in Huntingtin Interacting Protein 1-transformed cells.* Cancer Cell, 2003. **3**(5): p. 471-82.

90. Engqvist-Goldstein, A.E., et al., *An actin-binding protein of the Sla2/Huntingtin interacting protein 1 family is a novel component of clathrin-coated pits and vesicles.* J Cell Biol, 1999. **147**(7): p. 1503-18.

91. Hyun, T.S., et al., *HIP1 and HIP1r stabilize receptor tyrosine kinases and bind 3-phosphoinositides via epsin N-terminal homology domains.* J Biol Chem, 2004. **279**(14): p. 14294-306.

92. Wang, H., et al., *HIP1R targets PD-L1 to lysosomal degradation to alter T cell-mediated cytotoxicity.* Nat Chem Biol, 2019. **15**(1): p. 42-50.

93. Berghoff, A.S., et al., *Correlation of immune phenotype with IDH mutation in diffuse glioma.* Neuro Oncol, 2017. **19**(11): p. 1460-1468.

94. Grzendowski, M., et al., *Differential proteome analysis of human gliomas stratified for loss of heterozygosity on chromosomal arms 1p and 19q.* Neuro Oncol, 2010. **12**(3): p. 243-56.

95. Kim, S.I., et al., *Revisiting vimentin: a negative surrogate marker of molecularly defined oligodendroglioma in adult type diffuse glioma.* Brain Tumor Pathol, 2021.

96. Pan, Y., et al., *Intermediate filament protein synemin contributes to the migratory properties of astrocytoma cells by influencing the dynamics of the actin cytoskeleton.* FASEB J, 2008. **22**(9): p. 3196-206.

97. Deville, S.S., et al., *The Intermediate Filament Synemin Regulates Non-Homologous End Joining in an ATM-Dependent Manner.* Cancers (Basel), 2020. **12**(7).

98. Hanmin, C., et al., *Pleiotropic Role of Tenascin-C in Central Nervous System Diseases: From Basic to Clinical Applications.* Front Neurol, 2020. **11**: p. 576230.

99. Woehrer, A. and J.A. Hainfellner, *Molecular diagnostics: techniques and recommendations for 1p/19q assessment.* CNS Oncol, 2015. **4**(5): p. 295-306.

100. Ducray, F., et al., *alpha-Internexin expression identifies 1p19q codeleted gliomas.* Neurology, 2009. **72**(2): p. 156-61.

101. Ducray, F., et al., *Anaplastic oligodendrogliomas with 1p19q codeletion have a proneural gene expression profile.* Mol Cancer, 2008. **7**: p. 41.

102. Ducray, F., et al., *Diagnostic and prognostic value of alpha internexin expression in a series of 409 gliomas.* Eur J Cancer, 2011. **47**(5): p. 802-8.

103. Eigenbrod, S., et al., *alpha-Internexin in the diagnosis of oligodendroglial tumors and association with 1p/19q status.* J Neuropathol Exp Neurol, 2011. **70**(11): p. 970-8.

104. Feller, C., et al., *Histone epiproteomic profiling distinguishes oligodendroglioma, IDH-mutant and 1p/19q co-deleted from IDH-mutant astrocytoma and reveals less tri-methylation of H3K27 in oligodendrogliomas.* Acta Neuropathol, 2020. **139**(1): p. 211-213.

105. Filipski, K., et al., *Lack of H3K27 trimethylation is associated with 1p/19q codeletion in diffuse gliomas.* Acta Neuropathol, 2019. **138**(2): p. 331-334.

106. Habiba, U., et al., *Loss of H3K27 trimethylation is frequent in IDH1-R132H but not in non-canonical IDH1/2 mutated and 1p/19q codeleted oligodendroglioma: a Japanese cohort study.* Acta Neuropathol Commun, 2021. **9**(1): p. 95.

107. Pekmezci, M., et al., *Loss of H3K27 trimethylation by immunohistochemistry is frequent in oligodendroglioma, IDH-mutant and 1p/19q-codeleted, but is neither a sensitive nor a specific marker.* Acta Neuropathol, 2020. **139**(3): p. 597-600.

108. Capper, D., et al., *Practical implementation of DNA methylation and copy-number-based CNS tumor diagnostics: the Heidelberg experience.* Acta Neuropathol, 2018. **136**(2): p. 181-210.

109. Wiestler, B., et al., *Integrated DNA methylation and copy-number profiling identify three clinically and biologically relevant groups of anaplastic glioma.* Acta Neuropathol, 2014. **128**(4): p. 561-71.

110. Stichel, D., et al., *Distribution of EGFR amplification, combined chromosome 7 gain and chromosome 10 loss, and TERT promoter mutation in brain tumors and their potential for the reclassification of IDHwt astrocytoma to glioblastoma.* Acta Neuropathol, 2018. **136**(5): p. 793-803.

111. Geiger, T., J. Cox, and M. Mann, *Proteomic Changes Resulting from Gene Copy Number Variations in Cancer Cells.* PLOS Genetics, 2010. **6**(9): p. e1001090.

112. Xiang, W., et al., *Myristoylated alanine rich protein kinase C substrate is a potential cancer prognostic factor that regulates cell migration and invasion in glioblastoma.* Oncol Rep, 2019. **41**(4): p. 2464-2470.

113. Hua, C., et al., *Minichromosome Maintenance (MCM) Family as potential diagnostic and prognostic tumor markers for human gliomas.* BMC Cancer, 2014. **14**: p. 526.

114. Rupp, T., et al., *Tenascin-C Orchestrates Glioblastoma Angiogenesis by Modulation of Pro- and Anti-angiogenic Signaling.* Cell Rep, 2016. **17**(10): p. 2607-2619.

115. Gingras, A.C., R. Aebersold, and B. Raught, *Advances in protein complex analysis using mass spectrometry.* J Physiol, 2005. **563**(Pt 1): p. 11-21.

116. Sanchez-Arago, M., L. Formentini, and J.M. Cuezva, *Mitochondria-mediated energy adaption in cancer: the H(+)-ATP synthase-geared switch of metabolism in human tumors.* Antioxid Redox Signal, 2013. **19**(3): p. 285-98.

117. Cuezva, J.M., et al., *The bioenergetic signature of cancer: a marker of tumor progression.* Cancer Res, 2002. **62**(22): p. 6674-81.

118. Bruggemann, M., et al., *Systematic Analysis of the Expression of the Mitochondrial ATP Synthase (Complex V) Subunits in Clear Cell Renal Cell Carcinoma.* Transl Oncol, 2017. **10**(4): p. 661-668.

119. Vander Heiden, M.G., L.C. Cantley, and C.B. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation.* Science, 2009. **324**(5930): p. 1029-33.

120. Warburg, O., *On respiratory impairment in cancer cells.* Science, 1956. **124**(3215): p. 269-70.

121. Warburg, O., *On the origin of cancer cells.* Science, 1956. **123**(3191): p. 309-14.

122. Acebo, P., et al., *Cancer abolishes the tissue type-specific differences in the phenotype of energetic metabolism.* Transl Oncol, 2009. **2**(3): p. 138-45.

123. Cuezva, J.M., et al., *The tumor suppressor function of mitochondria: translation into the clinics.* Biochim Biophys Acta, 2009. **1792**(12): p. 1145-58.

124. Cardone, R.A., V. Casavola, and S.J. Reshkin, *The role of disturbed pH dynamics and the Na+/H+ exchanger in metastasis.* Nat Rev Cancer, 2005. **5**(10): p. 786-95.

125. Fu, X., et al., *2-Hydroxyglutarate Inhibits ATP Synthase and mTOR Signaling.* Cell Metab, 2015. **22**(3): p. 508-15.

126. Dong, Y., et al., *Reversibility of Age-related Oxidized Free NADH Redox States in Alzheimer's Disease Neurons by Imposed External Cys/CySS Redox Shifts.* Sci Rep, 2019. **9**(1): p. 11274.

127. Wick, W., et al., *Phase II Study of Radiotherapy and Temsirolimus versus Radiochemotherapy with Temozolomide in Patients with Newly Diagnosed Glioblastoma without MGMT Promoter Hypermethylation (EORTC 26082).* Clin Cancer Res, 2016. **22**(19): p. 4797-4806.