Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

M.Sc. Dennis Friedel

born in: Heidelberg, Germany

Oral examination: 18.03.2026

Proteomic analysis of isocitrate dehydrogenase mutant

Astrocytoma reveals three novel subtypes linked with

malignant progression

Referees:
Prof.Dr. Stefan Wiemann
Dr. Mikhail Savitski

Acknowledgement

I want to express my deepest thank to everyone who supported to this work and made my PhD journey an unforgettable experience. Firstly, I would like to thank Prof. Dr. Andreas von Deimling for giving me the chance to work in such an exciting field.

Furthermore, I would like to thank Dr David Reuss for being such a great supervisor. Thank you for guiding and teaching me, as well as the many motivating discussions we had about brain tumours, their diagnostics, and clinical meanings.

I would also like to thank Prof.Dr Wiemann and Dr. Mikhail Savitski for being my first and second referees and for participating in my thesis advisory committee. Thank you very much for your patience and every valuable discussion we had as well as very constructive advice that helped me to find my own way on this journey.

In addition, I would like to thank Prof. Dr. Stefan Wölfl for being the chairman in the upcoming defence and Dr. Juyan Lu for his participation in my PhD examination committee.

I would like to express my thanks to my colleague and friend, Dr. Daniel Schrimpf. Not only has he provided invaluable support and mentoring in analysing methylation data and managing information within our database, but he has also reminded me of the importance of taking breaks, like going outside for a coffee, to stay focussed.

I would like to express my deepest gratitude to Christian Hagenlocher and Dr. Marius Felix for the preparation and analysis of the tissues using LC-MS. Additionally, I want to acknowledge Dr. Ginaluca Sigismondo for setting up the new analysis pipeline for proteomic data analysis in such a short time that it was possible to continue with the data acquisition.

Moreover, I want to express my thanks to Dr. Ahmed Sadik who was always there when I had questions regarding any bioinformatic problems. Ahmed you are a great Mentor and I will definitely never forget your help during my PhD Journey and looking forward for the next joint projects.

Many thanks to all the people working at the Neuropathology for creating such a wonderful working atmosphere and for making my PhD journey a great time. In this context, I would like to thank Dr. Isabell Bludau and Dr. Armin Hadzic for their support and guidance in the last months of my PhD. Special thanks to my Office-buddy and friend Felix Hinz for all the great times and scientific and non-scientific conversations we shared during our PhD-Journeys. I really miss it that we are now not in the same office anymore.

I would like to thank my family and friends for their support along this long way and for motivating me during my studies and to never give up until I reach my goal.

Last but not least, I would like to thank my wife Ina for her endless support and love as well as her endless patience with my forgetfulness over the past four months. I would have never made it without you.

Abstract

IDH-mutant Astrocytoma are malignant brain tumours with highly variable biological behaviour and clinical prognosis. The prognostic significance of current histopathological grading for lower-grade IDH-mutated astrocytoma remains controversial. This thesis investigates the prognostic potential of the IDH-mutant Astrocytoma proteome by linking protein abundance with clinical outcomes, characterising associated biological processes and evaluating its relevance for risk stratification of lower grade IDH-mutated astrocytoma. DIA-based whole proteome profiling of 80 treatment naive IDH-mutant Astrocytoma (31 WHO grade 2; 30 WHO grade 3; 19 WHO grade 4) identified 11 623 different proteins. Unsupervised analysis revealed three biologically and clinically distinct subtypes: a neuronal subtype linked with favourable overall survival, a metabolic with intermediate outcome ($p < 0.001$; HR=8.53) and a mesenchymal linked with the worst outcome ($p < 0.001$; HR=16.46). In-depth analysis of the intermediate group showed striking upregulation of proteins linked to mitochondrial biogenesis as well as amino acid and lipid metabolism with simultaneous dysregulation of subunits of the electron transport chain (ETC) complexes I. Enrichment of MYC target proteins supported the presence of a MYC driven metabolic reprogramming. A decision-tree prediction model trained on these findings demonstrated robust performance (BACC = 1) and demonstrated biological and clinical reproducibility across two in-house and two publicly available datasets. This thesis provides novel molecular and clinical insight into IDH-mutant astrocytoma through comprehensive proteomic characterisation. The discovered subtypes exhibit distinct clinical trajectories linked with distinct biological phenotypes offering potential for an improved risk stratification. Moreover, the unexpected discovery of distinct metabolic alterations in the intermediate group provides potential explanations on how these tumours gain partial independence from the IDH mutation. Finally, the introduced decision-tree model may serve as a potential framework for enhancing clinical stratification of lower grade IDH mutant astrocytoma patients.

Zusammenfassung

IDH-mutierte Astrozytome sind maligne Hirntumore, die mit sehr unterschiedlichen Prognosen und biologischen Verhalten behaftet sind. Die Gradierung für niedriggradige IDH-mutierte Astrozytome ist derzeit noch sehr umstritten da sie lediglich auf histopathologischer Begutachtung basiert. Das Ziel dieser Arbeit, ist es die prognostische Relevanz des Proteoms IDH mutierter Astrozytome zu ermitteln, indem der Zusammenhang zwischen Protein Expression und klinischen Verläufen untersucht wird. Zusätzlich charakterisiere ich assoziierte biologischen Prozesse und nehme Bezug auf die Risko-Stratifizierung dieser Tumore. Durch DIA-basierter Proteom-Analyse von FFPE-Proben von 80 Patienten mit primären IDH-mutierten Astrozytomen (31 WHO-Grad 2, 30 WHO-Grad 3, 19 WHO-Grad 4) wurden 11 623 Proteine identifiziert. Durch agnostische Analyse der Daten konnten 3 biologisch und klinisch unterschiedliche Untergruppen identifiziert werden, die wie folgt definiert sind: Eine neuronale Gruppe mit günstigem klinischem Verlauf, eine metabolische mit intermediärem Verlauf ($p<0.001$; HR=8.53) und eine mesenchymale Gruppe mit schlechtem Verlauf ($p <0.001$; HR=16.46). Durch detaillierte Analyse der intermediären Gruppe wurde eine mitochondrielle Veränderung gefunden bei den Proteinen welche zur mitochondriellen Biogenese, Aminosäure sowie Lipid Metabolismus beitragen stark überexprimiert sind und im Gegenzug die Untereinheiten des ersten Komplexes der Elektronentransportkette fehlen. Eine weitere Anreicherung von MYC assoziierten Proteinen in der intermediären Gruppe unterstützt die These einer MYC-getriebenen metabolischen Re-programmierung. Ein Eintscheidungsbaum-Modell, welches auf den vorher identifizierten unterschiedlich exprimierten Proteinen der Untergruppen basiert, zeigte gute Leistung (BACC =1) bei deren Bestimmung. Zudem bewies es biologische und klinische Reproduzierbarkeit der vorherigen Ergebnisse in zwei Abteilungsinternen und zwei öffentlich Zugänglichen Datensätzen. Die hier zugrundeliegende Arbeit liefert neue klinischen und molekularen Erkenntnisse zu IDH mutierter Astrozytome durch die umfassende Analyse ihres Proteoms. Die identifizierten klinisch unterschiedlichen Untergruppen zeichnen sich durch verschieden biologische funktionale Phänotypen aus. Diese können mögliche neue Ansatzpunkte zur klinischen Risiko-Stratifizierung bieten. Die unerwartete Entdeckung einer metabolischen Veränderung in der intermediären Gruppe deutet zudem auf mögliche Mechanismen hin, durch die IDH mutierte Astrozytome teilweise unabhängig von der IDH-Mutation werden können. Schließlich bietet das hier vorgestellte Entscheidungsbaum-Modell einen vielversprechenden Rahmen für die Verbesserung der klinischen Stratifizierung von Patienten mit niedriggradigen IDH-mutierten Astrozytomen.

Table of Contents

Table of Figures

1. Introduction

1.1. Isocitrate dehydrogenase mutant Astrocytoma

Astrocytoma, isocitrate dehydrogenase mutant (AIDH) are a malignant type of tumour which can develop in the central nervous system (CNS) of adults and are characterised by the occurrence of a hotspot mutation in either Isocitrate dehydrogenase (IDH)1 or 2. [1]. AIDH show an infiltrative growth which impedes with the complete surgical resection of the tumour tissue and consequently leads to tumour recurrence and progression [2]. While AIDH tumours are frequently initially diagnosed as slow growing lower malignancies, they possess the potential to eventually progress into more aggressive tumours over time resulting in a significant worsening of patient prognosis [3, 4]. In this context, the clinical behaviour of AIDH has been shown to be highly variable with some patients dying within a few years to the development of aggressive tumours while others survive over many years or even decades [5].

1.1.1. WHO classification

AIDH are classified and graded according to the World Health Organisation Classification System of Tumours of the Central Nervous System (CNS WHO) which represents the international standard classification for brain tumours (Figure 1). AIDH belongs to the larger group of diffuse gliomas, which are considered to originate from glial progenitor cells [6]. Diffuse gliomas are the most prevalent primary central nervous system (CNS) tumours [7] and are further classified into adult type and paediatric type gliomas. Other types of adult type diffuse gliomas are glioblastoma, IDH wild-type and oligodendroglioma, IDH mutant (OIDH) [1] which have been shown to differ in their prognosis and the age of onset from AIDH [7]. In this regard, AIDH depict a moderate incidence [7] where these tumours are often diagnosed as lower malignancies during the third and fourth decade of a patient lifetime [4]. While glioblastomas may share many histologically features with AIDH, they have been demonstrated to depict a significantly poorer prognosis [8, 9]. Furthermore, molecular alterations defined to distinguish AIDH from OIDH include the concomitant loss of the chromosomal arms 1p and 19q in OIDH. Additional markers for AIDH are loss of nuclear expression or pathogenic mutations in the alpha thalassemia/mental retardation syndrome (ATRX), X-linked ATRX, and tumour protein 53 (p53) mutations [1].

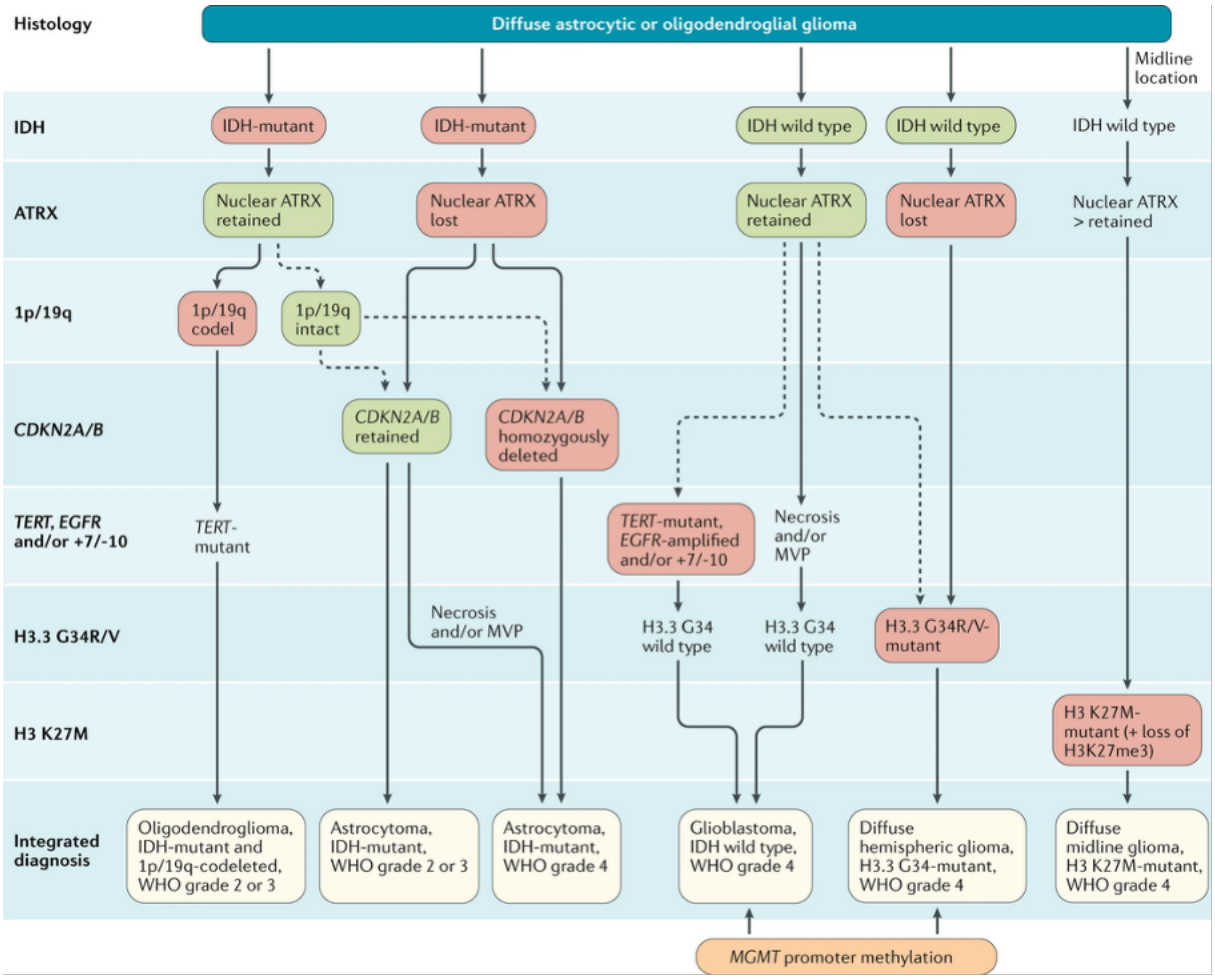


Figure 1: CNS5 WHO Classification guidelines for the integrated diagnosis of AIDH in adults. Adapted from [2].
AIDH are histologically identified by exhibiting diffuse astrocytic or oligodendroglioma features in. Further molecularly stratification is used to distinguish AIDH from GBM and OIDH through the presence of the IDH mutation as well as retainment of chromosomal arms 1p/19q and loss of ATRX.

1.1.2. WHO grading

AIDH are graded as WHO grade 2, 3 or 4. Low(er) grade AIDH (WHO grades 2 and 3) have a more favourable prognosis compared to high-grade AIDH (WHO grade 4), which represents a highly malignant group with the most dismal prognosis.

According to WHO CNS5, low(er)- and high-grade AIDH are distinguished by morphologic features and by the assessment of the copy number status of the gene locus cyclin-dependent kinase inhibitor 2A/B (CDKN2A/B) [1]. CNS WHO grade 4 AIDH are characterised histologically by the presence of necrosis and/or vascular proliferation [1]. Furthermore, several studies have shown that homozygous loss of CDKN2A/B is associated with a poor prognosis in AIDH [10-13], therefore it has been introduced in CNS5 WHO as a molecular indicator of the WHO grade 4, irrespective of the histopathologic appearance [1].

Despite the advancements in CNS5 WHO grading of high grade AIDH, the distinction of low-grade AIDH into WHO grades 2 and 3 remains a matter of histological identification [14]. According to CNS5 WHO, AIDH WHO grade 2 have the lowest malignancy and are characterised by exhibiting well-developed differentiation, low to moderate cell density and mild nuclear irregularities. In contrast, WHO grade 3 CNS AIDH is characterised by focal or dispersed anaplasia, a "significant mitotic activity" and the absence of a homozygous loss of CDKN2A/B. A threshold for the mitotic activity is however not defined [15].

The prognostic relevance of the grading of low-grade AIDH is currently strongly debated, as the overall survival of patients with grade 2 and 3 tumours is similar and grading is often subject of substantial inter-observer variability [16-20]. Furthermore, the rational for distinguishing lower-grade AIDH originates from historical series that included both patients with IDH wild-type and IDH-mutant tumours [14]. Following the introduction of the WHO 2016 classification, it cannot be assumed to remain valid for determining prognosis for IDH-mutant glioma patients [8]. In this context, it became imperative for up-to-date studies to investigate clinical, radiological, and molecular factors that influence the outcome of patients with AIDH with the aim to improve the current risk stratification.

1.1.3. Treatment

Despite its debated role, the CNS5 WHO grading system remains a substantial resource for treatment decisions for AIDH patients. Current treatment guidelines, which include radio- and chemotherapy or a "watch and wait" treatment depending whether an AIDH patient is categorised as high or low risk (Figure 2). In high-risk patients, chemotherapy is administered employing alkylating agents, such as Temozolomide (TMZ) [21]. Conversely, the "watch and wait "-strategy entails active monitoring of the patient without direct adjuvant treatment following surgery for low-risk AIDH patients. The categorization of low and high risk is determined based on the CNS5 WHO grade, the size of the tumours, the patient's age, and the presence or absence of neurological deficits [2].

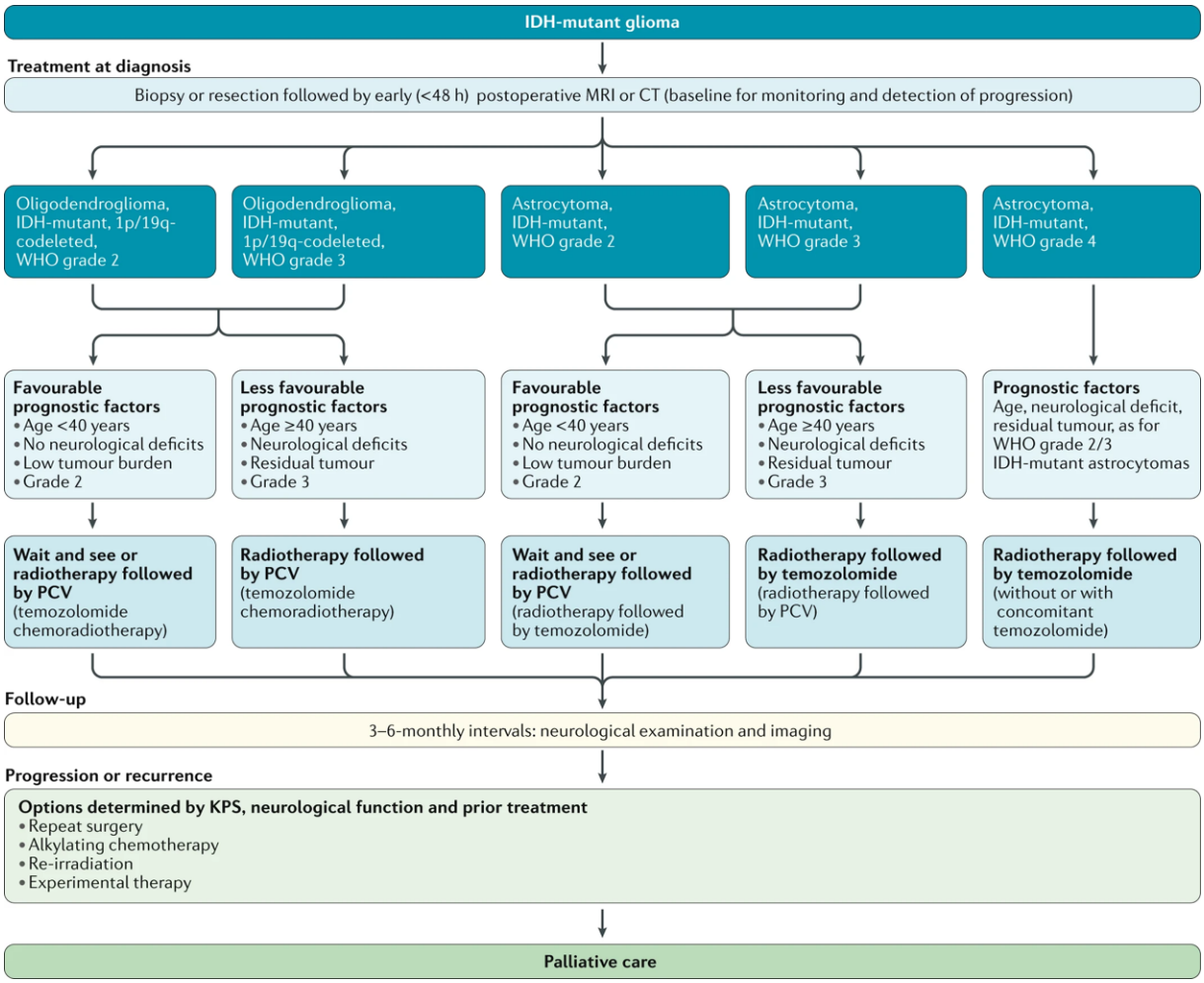


Figure 2: Treatment guidelines for IDH-mutant glioma. Adapted from [2].
Treatment decision for AIDH patients is based in several prognostic factors including patient age, exhibition of neurological deficits, tumour burden, and WHO grade. For patients indicating favourable prognostic factors a 'Wait and see' approach is used while patients with less favourable prognostic factors are treated with temozolomide and radiotherapy.

1.1.4. Genetic and epigenetic alterations associated with malignant progression of AIDH

Given recent advances in molecular diagnostics, it became paramount to elucidate whether molecular data might offer a better risk stratification for AIDH patients. Recent studies are aiming to elucidate this question through genetic, epigenetic and transcriptional interrogation.

1.1.4.1. Genetic alterations

Next-generation sequencing has emerged as an important tool for investigating potential genomic events, including mutations, chromosomal instability and gene-level somatic copy number alterations (SCNA). Including CDKN2A/B there are several studies which identified alterations which can occur during malignant tumour progression including SCNA like homozygous deletion of RB1 [13], amplification of CDK4, MET, MYCN, mutations in PIC3CA or PIK3R1 [22-27] as well as global chromosomal alterations such as deletions of 14q

[17] or an overall increase of the mutational load [13]. However, identifying alterations impacting patient prognosis remains complex, as many of them occur simultaneously. Indeed, currently no other molecular alteration beside CDKN2A/B homozygous deletion is relevant for WHO grading [15].

1.1.4.2. Epigenetic profiling

Two major methylation subtypes of AIDH have been described from DNA methylation analysis characterised and termed in accordance to their global genome wide methylation levels into Glioma-CpG Island Methylator Phenotype (G-CIMP)-low and G-CIMP-high [26]. Here, patients with an AIDH G-CIMP-low subtype exhibited alterations in cell cycle pathways as well as homozygous CDKN2A/B deletions and MET amplifications. In addition, these patients showed a significantly worse overall survival than patients with the G-CIMP-high subtype [26, 28-30]. Furthermore, long-term studies analysing paired patient samples demonstrated that some G-CMIP-high tumours turned into G-CIMP-low during tumour recurrence highlighting genome-wide DNA-demethylation is associated with malignant tumour progression of AIDH [31, 32]. In parallel to this studies, a DNA-methylation-based random-forest classifier termed „Heidelberg methylation classifier“ has been developed to enhance robustness and precision in brain tumour diagnosis compared to conventional diagnostic methods [33, 34]. According to this methylation classifier AIDH are distinguished into two classes, with distinct survival outcome termed “Astrocytoma, IDH-mutant; high-grade” (A IDH HG) and “Astrocytoma, IDH-mutant; low-grade” (A IDH LG). The prognostic significance of these groups was further substantiated in the context of only WHO grade 3 and the WHO grades 2-4 [28, 35].

1.2. Metabolic phenotype of IDH mutant glioma

Tumour cells have been demonstrated to adapt their metabolism to meet the substantial energy demands for sustained proliferation. This remodelling was initially observed by Otto Warburg and later termed as the Warburg effect. It describes the metabolic transition from oxidative phosphorylation (OXPHOS) to glycolysis for primary energy generation even in the presence of oxygen, resulting in the production of lactic acid [60].

IDH mutant tumours, however, exhibit a distinct metabolic shift in comparison to the conventional Warburg effect resulting from the neomorphic function acquired through a hotspot missense mutation either in isocitrate dehydrogenase 1 (IDH1) or IDH2. The mutation is regarded as an early event and driver of tumorigenesis of IDH mutant glioma as it is homogeneously distributed throughout in the tumour [36] and preserved in recurrent tumours

[37, 38]. Though, it has been described that the mutation loses significance during tumour progression when more aggressive mutations are acquired, which enable IDH mutant glioma to proliferate independently of IDH [39]. In addition to driving tumorigenesis, the metabolic shift further leads to disruption and alteration of numerous metabolic pathways, which challenges the tumour to sustain continuous proliferation. In this context, IDH mutant gliomas have been demonstrated to exhibit a substantially slower proliferation rate compared to tumour cells lacking the IDH mutation [40, 41]. Although these processes present intriguing vulnerabilities, it remains unclear how these dependencies evolve during the progression of AIDH towards more malignant tumours that are less dependent on the mutant IDH-protein for growth.

1.2.1. IDH mutation

IDH is the third enzyme of the Krebs cycle which performs the oxidative decarboxylation of isocitrate into α-ketoglutarate (α-KG) while reducing Nicotine Amid Dinucleotide (NAD)+ to NADH or of Nicotine Amid Dinucleotide Phosphate (NADP)+ to NADPH [42]. Furthermore, IDH plays a pivotal role in cellular metabolism, as it catalyses the rate-limiting step in the tricarboxylic acid (TCA) cycle by converting isocitrate to α-KG, which subsequently serves as precursors for the biosynthesis of numerous other metabolites [43]. Mutations often occur in cytosolic IDH1 and mitochondrial IDH2 at codons R132 and R172 [44] introducing a neomorphic enzymatic function on IDH, enabling it to convert α-KG, the product of wild-type IDH, into D-2-hydroxyglutarate (2-HG). As depicted in Figure 3, isocitrate is enzymatically transformed into α-KG in IDH-WT cells by a catalytic homodimer of two wildtype IDH1 monomers. In IDH1-mutated cells, a wildtype monomer and one monomer carrying the R132H mutant from a catalytic active heterodimer which performs the conversion of α-KG into 2-HG under oxidation of NADPH+ into NADP+ [63].

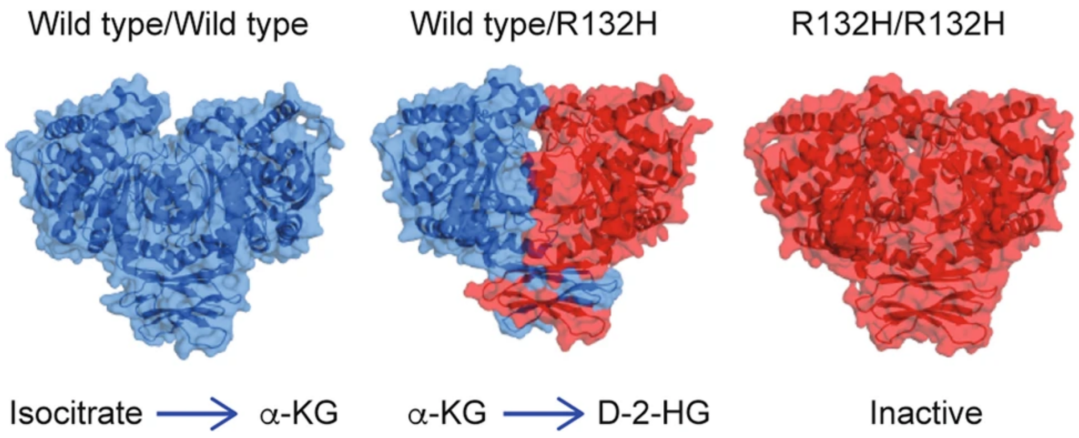


Figure 3: Conversion of α-KG in IDH-WT and IDH-WT/mutant complexes. ***Adapted from [63].***

1.2.2. IDH mutation specific inhibitors

Due to its important role as early driver of tumorigenesis in IDH mutant glioma [36], recent studies and trials have evaluated the potential role of mutation-specific inhibitors as a targeted treatment option. Here, the oral brain-penetrating inhibitor of mutant IDH1 and IDH2 Vorasidenib, has demonstrated substantial improvement in progression-free survival among AIDH patients with CNS WHO grade 2 which were considered for "watch and wait". Patients treated with Vorasidenib showed an improvement of median progression-free survival (PFS) from 11.1 months to 27.7 months when compared to patients treated with placebo [45].

1.2.3. D-2-hydroxyglutarate

2-HG is considered as an oncometabolite [46] which acts as an antagonist of α-KG by inhibiting α-KG-dependent dioxygenases, such as prolyl-4-hydroxylase (PHD) and further the histone lysine demethylases (KDM) as well as DNA demethylases from the ten-eleven translocation (TET) enzyme family [47, 48]. Consequently, IDH mutations causes hypermethylation of DNA and histone proteins, which are widely anticipated as significant early mechanisms of the underlying tumorigenesis of IDH mutant glioma [49, 50].

In parallel to its oncogenic function, it has also been described that a high concentration of 2-HG, like α-KG, impair with mitochondrial respiration by inhibiting ATP synthase and subsequently inhibiting mammalian target of rapamycin (mTOR) signalling. [51]. In addition, 2-HG production has further been described to inhibit 2-oxyglutarate-dependent dioxygenases [47], including prolyl-4-hydroxylase (C-P4H) catalyses the maturation of collagen. Therefore, endoplasmatic reticulum (ER) stress is induced due to accumulation of immature collagen [52] which further leads to initiation of ER autophagy to survive the stress response [53].

1.2.4. Deficiencies in lipid metabolism

Neomorphic function of the IDH mutation causes a significant increase in turnover of NADPH into NAD^+. NADPH is not only crucial for maintenance of cellular redox balance but also generating energy through the respiratory chain and *de novo* lipogenesis [54-56]. Consequently, IDH mutant cells become dependent on exogenous lipid sources to sustain proliferation [57] causing a diminished membrane integrity of both the endoplasmic reticulum (ER) and the Golgi apparatus [58].

1.2.5. Glutamine metabolism

Glutamine metabolism is frequently altered in IDH mutant glioma as its upregulation sustains the growth of IDH mutant glioma [59, 60] by serving as a carbon and nitrogen source as well as for the synthesis of larger molecules including proteins, purines, pyrimidines and amine

sugars [61]. Glutamine is further converted into Glutamate by the glutaminase (GLS) enzyme, which is crucial metabolite for various processes, including de novo lipogenesis, energy production through the TCA cycle through conversion into α-KG or as an antioxidant by conversion it into glutathione [62]. In this context, GLS-mediated production of α-KG was discovered to be crucial for sustaining tumour growth as the single allele production of wildtype IDH in IDH mutant glioma alone seems to be insufficient to generate enough α-KG. This dependence was substantiated by the inhibition of GLS, which demonstrated a reduction in tumour growth. However, this effect was only observed when glucose was deprived [63]. In parallel, it was demonstrated that IDH mutant gliomas overcome the α-KG sufficiency from single allele wildtype IDH production by overexpression of Glutamate Dehydrogenase (GLUD) 1 and GLUD2, both enzymes responsible for converting glutamate to α-KG. Notably, it was shown that the effect of tumour growth could be inhibited in murine glioma progenitor cells *in vitro* and *in vivo* by increase of IDH mutant expression [64]. Given the high abundance of glutamate as a neurotransmitter in the brain, it is widely hypothesised that IDH mutant gliomas develop a unique metabolic phenotype, becoming „glutamate-suckers “. Here, overexpression of GLUD serves as a crucial mechanism in these cells, enabling the production of sufficient α-KG and facilitating the biosynthesis of other essential molecules. These processes appear to be essential for the growth and survival of IDH-mutant gliomas. [65].

1.2.6. Mitochondrial metabolism and oxidative stress

The IDH mutation leads to a heightened reliance of the tumour metabolism on mitochondrial functions. In this context, it has been demonstrated that IDH-mutant tumours exhibit a greater utilisation of OXPHOS compared to IDH wild-type (WT) tumours making them vulnerable to mitochondrial stress or inhibition of complexes of the electron transport chain (ETC) [40]. Additionally, it has been demonstrated that due to the inhibition of cytochrome c oxidase enzymes by 2-HG [51], IDH mutant cells exhibit heightened sensitivity to Bcl-2 inhibitors. This heightened sensitivity is attributed to the low mitochondrial threshold for apoptosis [66]. On the other hand, it was also demonstrated that IDH-mutant tumours aim to reduce elevated oxidative stress caused by the decreased NADPH levels by increasing NADPH production via the oxidative pentose phosphate pathway (OXPPP) [57, 67]. Despite the overregulation of OXPPP, oxidative stress in IDH mutant glioma has been consistently demonstrated to be significantly elevated. This heightened oxidative stress may contribute to the observed heightened sensitivity of these cells to radiation therapy [67].

1.2.7. Hypoxia

Inhibition of prolyl-4-hydroxylase (PHD) enzymes due to α-KG depletion and 2-HG increase [68] in IDH mutant glioma has further been shown to provide stabilization of tumour growth under hypoxic conditions. PHD enzymes require α-KG to hydroxylate and subsequently initiate the degradation of hypoxia-inducible factor 1α (HIF-1α), a transcription factor which is essential to sustain cell survival and proliferation under hypoxic conditions. HIF-1α stabilization leads to the expression of target genes such as the vascular epidermal growth factor (VEGF) [69] leading to an increased cell and invasiveness under hypoxic condition [70].

1.4. Proteomics

The proteome comprehends all proteins that can be produced or modified by a biological system, including living organisms and other living systems [71-74]. Unlike the transcriptome, which lists all RNA molecules that can be expressed in a biological system, the proteome shows dynamic changes due to internal and external stimuli. Translation rate of messenger RNA (mRNA) controls protein levels, while protein degradation rate influences protein flux. Factors affecting protein degradation include location, conformation, post-translational modifications, and the ability to form stable complexes [75]. Additionally, various mechanisms regulating protein biosynthesis, the existence of distinct proteoforms derived from different splicing variants, and post-translational modifications result in a substantial increase in the complexity of the proteome [76]. To address this complexity, extensive large-scale interrogations have been conducted to study the proteome and development instrumentation for studying protein levels, isoforms, localization, interactions, post-translational modifications (PTMs), and turnover [77]. Consequently, contemporary proteomic technologies possess the potential to facilitate comprehensive analysis of the proteome in great depth and could elucidate knowledge gaps that have yet been inaccessible by genomic and transcriptomic research [78]. In this context, this technology possesses the potential to establish a foundation for the identification of novel protein biomarkers with diagnostic applications, as well as novel drug discoveries in the field of cancer research [79-81].

1.4.1. Mass-Spectrometry based proteomics

Mass spectrometry-based proteomics has become the dominant technology in proteomic research, enabling the comprehensive characterisation of protein structures and functions, and the identification and quantification of entire protein profiles from biological samples [78, 82]. Two approaches (Figure 4) are distinguished within the MS-based proteomics field: bottom-up (also known as shotgun proteomics) and top-down proteomics.

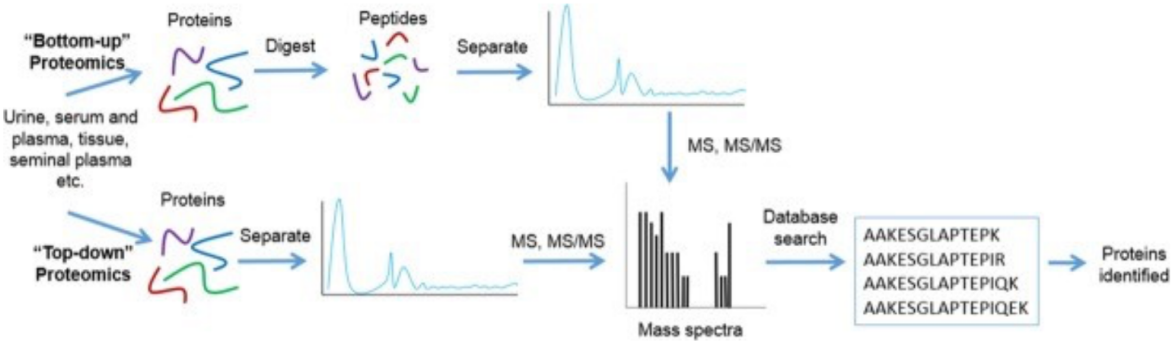


Figure 4.: Workflow for top-down and bottom-up ("shotgun") proteomics. Adapted from [83]**.**

In top-down proteomics, intact proteins are separated through liquid chromatography (LC), followed by ionisation and injection into a mass spectrometer (MS) enabling identification of the mass spectra of the intact protein as well as its fragments [84]. The advantage of this approach is that it enables analysis of the different protein isoforms that can arise from a single gene as well as post-transcriptional modifications (PTMs). However, top-down proteomics analysis only provides limited insight into complex samples, as analytical and computational challenges rise due to the increasing number of charge state distributions and the substantial number of product ions [85]. In contrast, bottom-up proteomics involves the initial fragmentation of proteins through enzymatic or chemical digestion, followed by subjection to liquid chromatography, ionisation, and analysis via tandem mass spectrometry (MS/MS). Next, a protein database is utilised to characterise the open-reading frame of the protein from which the peptide originated [86]. The primary advantage of this approach lies in the ease of peptide separation through reversed-phase liquid chromatography (RPLC). Peptides have enhanced ionisation capabilities [87] and exhibit a more predictable fragmentation pattern in comparison to intact proteins. Therefore, bottom-up proteomics is frequently employed in biological and clinical studies with the aim to investigate whole proteomes, as it facilitates high-throughput analysis and the identification and quantification of thousands of proteins.

1.4.2. Acquisition modes in bottom-up proteomics

Numerous bottom-up proteomics protocols have been developed, published, and reviewed over the past years [86, 88-91]. In this context, protein digestion is frequently performed using trypsin, followed by separation using reverse phase liquid chromatography (RPLC). Subsequently, the separated proteins are ionised via electron spray ionisation (ESI), followed by MS data acquisition. Different acquisition approaches have been developed including Data-dependent acquisition (DDA), Selected Reaction Monitoring (SRM), Parallel Reaction Monitoring (PRM) and Data-independent acquisition (DIA) (Figure 5) [92].

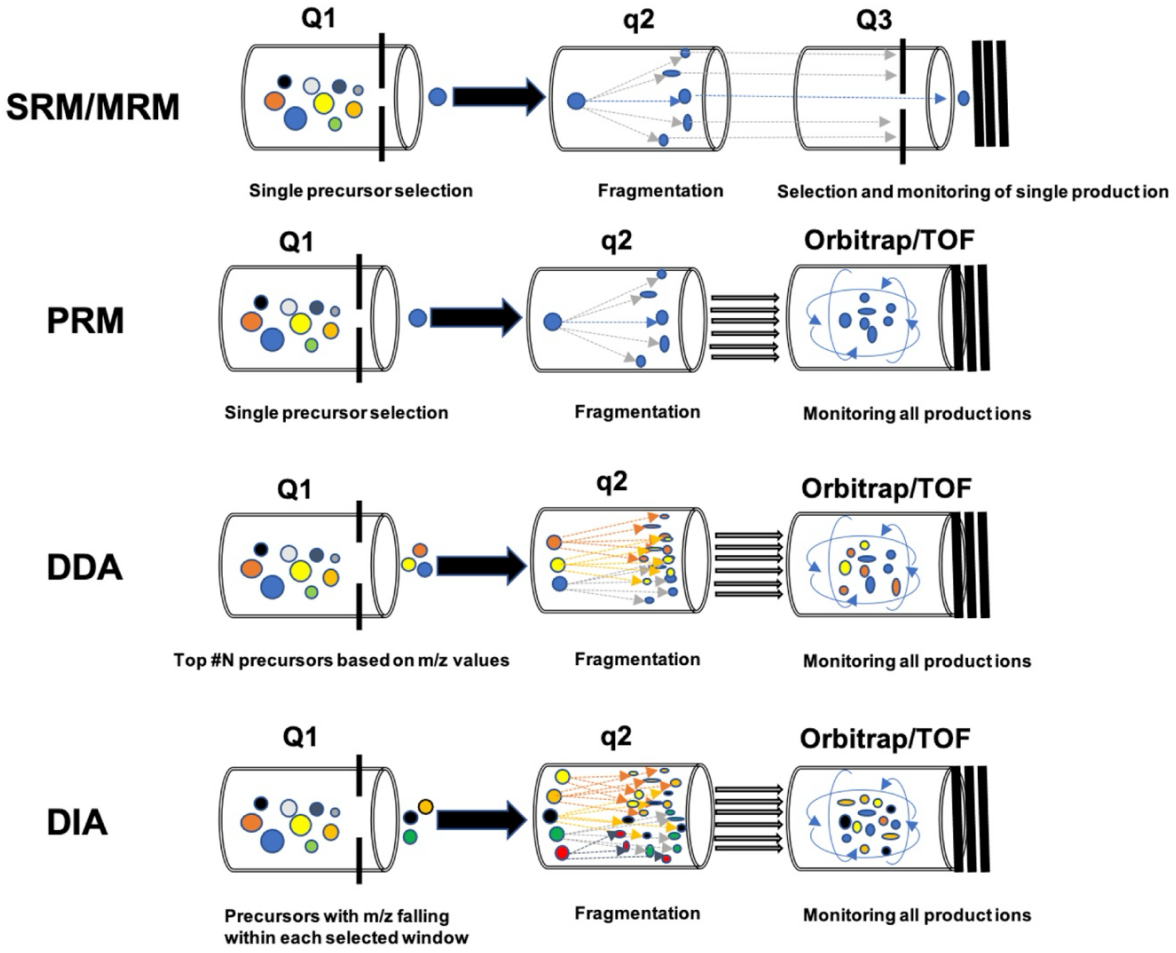


Figure 5.: Overview of bottom-up proteomics methods for MS data acquisition. Adapted from [92].

1.4.2.1. Data-dependent acquisition

DDA aims to acquire MS/MS spectra of peptides as they elute from the LC column. It determines the exact scan sequence for each analysis based on the data observed by the mass spectrometer [93, 94]. During DDA, precursor scans (MS1) are performed iteratively in quadruple 1 (Q1) until charged peptides (precursor ions) are detected. These ions are then automatically selected based on simple heuristics (e.g. N precursors with the highest MS1 intensity) for further ion fragmentation and MS/MS (MS2) acquisition in quadruple 2 (Q2). Subsequently, fragment ion spectra are compared with protein sequence databases like UniProt (https://www.uniprot.org/) to discover peptide spectrum matches (PSMs) [95, 96]. However, due to the heuristic nature of precursors selection, repetition of an DDA experiment with the same sample will lead to a different set of identifications, making it to the major drawback of DDA experiments. In this context, a strategy called „match between runs “(MBR) has been developed to reduce the heuristic selection bias by enhancing the proteome coverage though transferring identification information between samples during multi-sample analysis [97, 98].

1.4.2.2. Selected Reaction Monitoring & Parallel Reaction Monitoring

SRM and PRM are targeted approaches where prior information about the protein of interest, such as its MS-observable precursors, expected retention times, frequently observed fragment ions, and relative abundance, are necessary for acquisition [99]. As prior information indicate the specific peptide precursors for fragmentation in MS2, PRM allows precise quantification of corresponding signals across their anticipated chromatographic retention time range [100]. Subsequently, quantification of the signals is conducted via peak detection and signal integration over chromatographic dimensions. Consequently, SRM and PRM provide high sensitivity and selectivity about peptide detection and quantification. Although, a drawback of this strategy is that the proteome coverage in SRM and PRM are limited to a small set of proteins due to limitations in machine speed which prevents their use in whole proteome studies [92].

1.4.2.4. Data independent acquisition

DIA has become the primary method for clinical-oriented bottom-up proteomics studies for as it provides broad protein identification, comparable to that attainable with DDA paired with consistency and quantitative accuracy inherent in SRM/PRM measurements [91, 101-103]. In this context, DIA has been shown to provide an overall favourable performance in terms of sensitivity, selectivity, reproducibility, and proteome depth coverage compared to DDA [104, 105] and offers significant cost savings compared to targeted methods. During DIA, precursors are acquired by separation into successive mass-to-charge (m/z) isolation windows (5-25 Da) in Q1 followed by their fragmentation in Q2. Finally, fragmented ions within a particular ionisation window are then analysed in a systematic and unbiased manner by a high-resolution-accurate-mass (HRAM) analyser. [106, 107]. As the co-fragmentation of numerous peptides in the same precursor window leads to highly complex MS2 spectra a peptide spectral library is required for linking precursors to a fragment [105, 108]. Although such a library was obtainable from an DDA experiment, recent advancements in DIA bioinformatics processing also facilitated direct analysis of DIA data in library-free mode without the necessity of a DDA-based spectral library [109, 110].

1.4.3. Raw data analysis

Following data acquisition, the identification and quantification of peptides (pre-processing, respectively) of the acquired raw spectral data is carried out. Pre-processed data can then subsequently be interrogated in the final downstream analysis step.

A single peptide can depict various PTM in bottom-up proteomics as it can be recorded with varying modifications, charges, and amino acid chain lengths depending on the effectiveness of the digestion enzyme used [86]. Peptides with their modifications and charge state recorded in the MS are commonly termed precursor ions, these are further fragmented, and the peptide sequence is estimated through matching the fragmented spectra with a reference database (e.g. Universal Protein Resource (UniProt) [111]). Next, the quantity of the precursor is estimated and further combined to obtain the peptide quantity. Furthermore, quantity of the intact protein can be inferred by using multiple peptide quantities which belong to the open reading frame of the corresponding protein [112, 113].

As the structure of raw data differs between DDA and DIA measurements, software solutions exist that can be applied for the preprocessing step (see Table 1). As several of these tools are exhibit varying performance due to the implementation of distinct algorithms, they come with their own advantages and disadvantages that a user must carefully consider in relation to their instrumentation and study objectives.

Table 1.: List of frequently used freeware software for pre-processing raw proteomic data.

Name	Method	Publication
MaxQuant	DDA	[98, 114-117]
MSFragger	DDA	[118]
Mascot	DDA	[119]
MS-GF+	DDA	[120]
X! Tandem	DDA	[121, 122]
MaxDIA	DIA	[123]
Skyline	DIA/Target	[124]
DIA-NN	DIA	[110]

In this context, a recent survey on the current state of the art DIA analysis software indicated a high degree of consistency in the results of the tools despite the variations in implementation and adjustable parameters [125]. In comparison, it has been shown that the software DIA-NN [110] performs on a consistent high level across numerous tested categories. DIA-NN demonstrated best performance across different instrumentations in categories of identification, computational efficiency and in regard of compatibility with different raw DIA formats, especially when used in library-free mode [125].

1.4.4. Downstream analysis of proteomic data

Independent of the selected preprocessing approach, raw data analysis provides a comprehensive list of identified proteins or peptides along with their respective quantities. This data can be further utilised in the downstream analysis for biological interpretation. In this regard key analysis steps proposed for downstream analysis of proteomic data analysis include a series of steps distinguished into the data preparation and the subsequent significance and biological analysis [126].

The data preparation comprehends a series of pre-processing steps to prepare the proteomic data for statistical and biological analysis and include data filtering, normalization, batch correction and missing values imputation [126]. Assessment of proteins that are significantly altered upon different groups can be obtained by the *limma* R package [127]. This package uses a linear regression model to estimate an Empirical Bayesian adjusted moderated t-statistic, which supports comparisons of conditions with an unbalanced design or a small sample size per group. Originally this package has been developed for microarray data but it has been shown to be efficient for shotgun proteomic data as well as it overcomes the paucity of the measurement matrix by sharing variance between proteins [127]. Pathway analysis has become a pivotal tool for the interpretation of the potential biological meaning of lists of proteins identified through DEA. In this regard, Gene set enrichment analysis (GSEA) provides biological interpretation by enabling significance testing for the enrichment of a gene signature in the context of a compared group [128].

1.4.5. Proteomic landscape of AIDH

Liquid-chromatography coupled to mass spectrometry (LC-MS) has become as a promising technology to study the biology and clinical significance of the proteomes of brain tumours [129-132]. In this context, our group demonstrated the clinical utilisation of LC-MS proteomics in IDH mutant glioma by discovering the proteins VIM and HIP1R as novel immunohistochemistry (IHC) markers for a cost-effective and reliable diagnosis of differentiating AIDH from Oligodendroglioma with IDH mutation (OIDH) [130, 133]. Furthermore, recent proteomic investigations discovered novel potential prognostic and targetable processes in GBM [134, 135], suggesting utilization of proteomics for clinical stratification and identification of targets for therapeutic intervention. Recent studies conducted interrogating the proteomic landscape of AIDH were designed to elucidate differences of the proteomes of IDH wildtypes (IDHwt) and IDH mutant (IDHmut) diffuse glioma together [136, 137]. However, the insight into proteome of the lower grade AIDH and its prognostic potential remains poorly understood. In this regard, it is of high interest whether lower grade AIDH can

be stratified based on their protein expression and whether the resulting subtypes align with the current histology depended WHO grading. Furthermore, it is possible that these results could represents a more effective method for their clinical stratification and uncover proteins and biological processes associated which may provide novel insights into the biology of malignant progression of these tumours.

1.5. Aims of work

Motivated by the preceding background, which depicted MS-based proteomics as a novel tool to interrogate the proteomic landscape of AIDH and its contribution to disease biology. The objectives of this thesis were the following:

1. Assessment and analysis of the proteomic profiles of a clinical and molecular well characterised cohort of AIDH cases to evaluate contribution to tumour malignancy.
2. Comprehensive characterization of proteins in context of clinical outcomes of AIDH to discover potential candidates and biological mechanisms contributing to tumour malignancy
3. Evaluation and testing of a prediction model based on biological and clinical findings to test for reproducibility and possible utilization in a clinical context

2. Methods

2.1. Data Acquisition

2.2.1. Tumour Series form internal sources

Step 2.2.1. Cohort Curation and a retrospective examination of tumour diagnosis and grades of IDH mutant Astrocytoma according to the CNS5 WHO criteria were conducted by Dr. David E. Reuss (DER), Heidelberg Neuropathology.

Formalin-fixed and paraffin-embedded Tumour tissues from 80 patients were ascertained for this study to analyse the whole proteome of AIHD patients by DIA (DIA Study Cohort, respectively, Table 2). Samples were acquired from the archives of the Departments of Neuropathology and Neurosurgery of the University Hospital Heidelberg as well as the Departments of Neuropathology Mannheim (in-house data, respectively). Samples were included by the following criteria: (1) all tumours had a IDH1 mutation, (2) all tumours had a WHO2021 CNS diagnosis of Astrocytoma grade 2,3 or 4, (3) FFPE and fresh frozen (FF) material for further molecular analysis was available. (4) both chromosomes 1p and 19q were retained. (5) Status of markers like CDKN2A/B was available (6) Long term monitoring of progression-free (PFS) and overall survival (OS) was available.

The curation of tissues and data was conducted in accordance with local ethics regulations and received approval. In addition, it was aimed to curate cases where Illumina Infinium HumanMethylation450 (450k) or Infinium Methylation EPIC (850k) BeadChip-based data was available for further epigenetic data analysis, Methylation class prediction and copy number analysis. Methylation class prediction of the cases was acquired from the "Heidelberg Methylation Classifier" version 12.8 [138]. All Tumours diagnosis were re-examined according to the integrated diagnoses standards by the CNS5 WHO 2021 and HE-stained sections were marked for tissue extraction by DER.

Older data records of proteomes from AIDH patient cohort analysed at the Neuropathology Heidelberg were included in this study for independent validation and benchmarking of findings from the DIA study cohort. The first cohort included, 24 patients (9 WHO grade 2,11 WHO grade 3, 4 WHO grade 4) from our recent study investigating proteomic markers for AIDH and OIDH (HIP1R_VIM, respectively, Table 2) [130]. The second study comprehends proteomic profiles from archived tumour materials of 66 low-grade AIDH patients (26 WHO grade 2, 40 WHO grade 3) enrolled in the GGN [139] and NOA study trails [140] (GGN_NOA, respectively, Table 2).

Table 2: Inhouse proteomic datasets.
Tabular representation of the in-house proteomic datasets used in this study.

Dataset	Material	Acquisition Mode	Pre-processing	Groups
DIA Study Cohort (n=80)	FFPE	DIA-PASEF	DIA-NN version 1.8.1	n(grade 2) = 29 n(grade 3) = 31 n(grade 4) = 20
HIP1R_VIM (n=24)	FFPE/FF	DDA-PASEF	MaxQuant version 2.4.2.0	n(grade 2) = 9 n(grade 3) = 11 n (grade 4) = 4
GGN_NOA (n=76)	FFPE	DDA-PASEF	MaxQuant version 2.4.2.0	n(grade 2) = 26 n(grade 3) = 40

2.2.1.1. Formalin-fixed and paraffin embedding of tumour tissue

Step 2.2.1.1. was performed by the routine diagnostic laboratory of the Heidelberg Neuropathology.

Formalin-fixed and paraffin embedding (FFPE) of tumour material performed by immersing tumour material into formalin after surgical resection for a minimum of six hours to enable complete fixation. Next, the tumour underwent a dehydration process, in which it was incubated in progressively increasing concentrations of ethanol and xylene. Subsequently, the tumour was put into heated liquid paraffin and moulded into a solid FFPE block.

2.2.2.2. Fresh-Frozen preserved tumour tissue

Step 2.2.1.2. was performed by the routine diagnostic laboratory of the Heidelberg Neuropathology.

Tumour material was put into cryopreservation after surgical removal from the patient followed by archiving at the department of Neurosurgical Research led by Prof. Christel Herold Mende.

2.1.2. Publicly available sources

2.1.2.1. Datasets

A collection of publicly available datasets was curated for independent validation and benchmarking of findings from the DIA study cohort. Datasets used in this study were acquired from the publications of *Bader et al.* [131] and *Wong et al.* [132], which both studied proteomes of different types of IDH mutant glioma.

Table 3.: Publicly available proteomic datasets.
Tabular representation of the publicly available datasets containing IDH mutant astrocytoma samples and the sources where the data was retrieved from.

Author	Data	Instrument	Labelling	Preprocessing	Groups
Bader et al. [131]	Whole Proteome + Phosphoproteome (DDA)	EASY-nLC 1200 coupled to a Q Exactive HF and a Q Exactive HF-X Orbitrap mass spectrometer (Thermo Fisher Scientific)	Label-free	MaxQuant + Normalized and Imputed	10 Astrocytoma 11 Oligodendroglioma 11 Glioblastoma 10 CNS Tissue
Wong et al. [132]	Whole Proteome (DDA)	Orbitrap Fusion Tribrid MS platform (Thermo Scientific)	Label-free	MaxQuant unprepared	10 Astrocytoma 11 Oligodendroglioma 11 Glioblastoma 10 CNS Tissue

2.2.1.2. Biological Knowledge Databases

The contents from several Biological Databases were acquired to evaluate activity of biological pathways among the identified subtypes, presence of Protein-Protein Interactions and Peptide modifications. Gene sets for functional characterisation were queried from the molecular signature database (MSIGDB) [141] by using the R-package msigdbr [142]. Gene sets from all categories of the gene ontology database [143], as well as gene sets from the MSIGDB HALLMARK category were acquired. Furthermore, gene sets of the Mitochondrial 'MitoCoP' Database published by [144] were accessed to investigating role of mitochondrial proteins and to evaluate association of processes with the proteomic subtypes of AIDH. In addition, information about Protein-Protein interactions (PPI) as well as Kinase Substrate-Relationships (KRR) were obtained from the databases Signor accessed through the R-package Omnipath [145].

2.2. Sample preparation for LC-MS data analysis

2.2.1. Sample preparation for LC-MS/MS analysis by DIA-PASEF Mode

Step 2.2.1. was performed by Christian Hagenlocher, Technical Assistant in the Department of Neuropathology, Heidelberg. The highlighted text passages below describes the methods used by him and have been taken from the supplemental methods from Aras, F.K., et al., Expansion of the spectrum of tumors diagnosed as myxopapillary ependymomas. Acta Neuropathol [146]

"For proteomic analysis, 1.5 mm in diameter biopsy punches of FFPE blocks were subjected to protein extraction protocol with the BeatBox device (Preomics, Planegg/Martinsried, Germany). In brief, four cycles of BeatBox-mediated tissue shearing in SDS 4% lysis buffer were performed, followed by SP3-based protein clean-up [3], and on-bead tryptic digestion (Promega, Fitchburg, WI, USA) in 20uL of 50mM ammonium bicarbonate at 37 Celcius degrees for 16 hours, 800 rpm. Resulting peptides were acidified to TFA 1% and stored at -20 celcius degrees until the mass spectrometry injection."

2.2.2. Sample preparation for LC-MS/MS analysis by DDA-PASEF Mode

Step 2.2.2. was performed by Dr. Marius Felix Department of Neuropathology, Heidelberg

Sample preparation of in-house cohorts HIP1R_VIM and GGN_NOA subsequently analysed by DDA-PASEF was conducted as described by our group in [130]. In brief, FF tissue samples underwent cycling pressure for efficient tissue lysis and rapid protein digestion [147, 148]. Frozen sections were processed by microscopy to ensure a minimum tumour cell content of 70%, and then 2 mm^3 of tumour tissue was processed. FFPE tissues were punched with 1.5 mm disposable biopsy punches on suitable regions and put into a bead tube (Bertin Technologies SAS, Montigny Le Bretonneux, France) for the next processing steps.

2.3. LC-MS/MS analysis

2.3.1. LC-MS/MS analysis DIA-PASEF mode

Step 2.3.1. was performed by Christian Hagenlocher, Technical Assistant in the Department of Neuropathology, Heidelberg. The highlighted text passages below describe the methods used by him and have been taken from the supplemental methods from Aras, F.K., et al., Expansion of the spectrum of tumors diagnosed as myxopapillary ependymomas. Acta Neuropathol [146].

"A total of 200ng of peptides were separated on a 25 cm analytical column (75µm x 250mm, C18, 1.7µm, 120 Å, Aurora 3, IonOpticks) using the nanoElute2 liquid chromatography system (Bruker Daltonics, Billerica, MA, USA) coupled to a Trapped Ion Mobility Spectrometry Time-of-Flight HT Mass Spectrometer (Bruker Daltonics, Billerica, MA, USA) operated in positive mode (+1.6kV). Solvent A was water with 0.1% formic acid and solvent B was 80% acetonitrile, 0.1% formic acid. Peptides were separated over 120min at 50C, the percentage of

solvent B increased in a linear fashion from 3% to 8% in 3 min, then increased to 10% at 15 min, to 40% at 100 min, to 60% in 105 min and finally to 90% at 110 min, stayed 5min at 90%, and went down to 3% for the last 5 min. Peptides were acquired in data-independent acquisition mode with parallel accumulation - serial fragmentation (dia-PASEF), in a 100 – 2000 m/z range, and a 0.65 – 1.4 1/k0 ion mobility space, with a total cycle time of 1s."

2.3.2. LC-MS/MS analysis DDA-PASEF mode

Step 2.3.2. was performed by Dr. Marius Felix Department of Neuropathology, Heidelberg. The highlighted text passages below describe the methods used by him and were obtained from our joint paper Felix, M., et al., HIP1R and vimentin immunohistochemistry predict 1p/19q status in IDH-mutant glioma. Neuro Oncol, [130].

"An ultra-high pressure nanoflow chromatography, Easy nLC 1200 (Thermo Fischer Scientific) system was coupled online to timsTOF pro mass spectrometer (Bruker Daltonics, Bremen, Germany). Peptides (equivalent to 500 ng) were loaded onto an in-house-packed 50 cm reversed-phase column (1.9 µm C18-coated porous silica beads, Dr. Maisch, Ammerbuch-Entringen, Germany) with 75 µm inner diameter and a pulled emitter tip. Chromatography was performed at 60°C with a 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)) were used for low pH peptide separation. Peptides were separated in 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."

2.4. Preprocessing of raw proteomic data

The bioinformatic identification and quantification of peptides and proteins from raw spectra data obtained from in-house DDA-/DIA-PASEF experiments data was done as previously described by our group in [146]. The analysis was conducted on a Linux system with 128 physical cores (AMD EPIC 75032 32-Core Processor) and 256Gb of RAM. Canonical protein sequences were downloaded from the Uniport Database on 28/03/2022 [149]. The database contained information of 20,547 proteins and 20,310 genes. These sequences were used for spectra matching in DDA experiments and for generating spectral libraries in DIA experiments, respectively.

2.4.1. Preprocessing raw DIA-PASEF data

Preprocessing of raw DIA-PASEF data was conducted by using the Linux command line version of the software DIA-NN version 1.8.1 [110]. DIA-NN was conducted in library-free mode using the UniProt canonical protein sequences and the recommended settings for TIMSTOF raw spectra and a FDR of 1% was used as a quality filter for precursors and proteins.

In short, the minimum fragmentation mass to charge ratio (m/z) was set to 200 minimum and 1800 m/z maximum. Furthermore, range for precursor m/z was set to 300 minimum and 1800 maximum. N-terminal methionine excision was enabled, and *in-silico* digest was set to involve cuts at Lysine (K) and Tyrosine (R). The maximum number of missed cleavages set to 1 and the peptide length were set between 7 and 30. Ranges for precursor charge were set to 1-4 and Cysteine carbamidomethylating was enabled as a fixed modification. In addition, the parameter '–reanalyse' was active to increase number of identifications by enabling identification of raw spectra using information between runs (Match-between-Runs (MBR), respectively).

2.4.2. Preprocessing raw DDA-PASEF data

Raw DDA-PASEF data was analysed by using the command line version of MaxQuant (version 2.4.2.0) [116, 117] and settings were used as described in [130]. In short, spectra were searched against canonical protein sequences obtained from the UniProt database (28/03/2022). Parameters such as enzyme specificity, FDR on peptide spectral match and protein level, precursor and fragment ion mass tolerance, fixed modifications were left on default settings. Amino acid deamidation (NQ) was added as additional parameter in variable modification to increase the number of identified proteins. The parameter MBR was enabled with default settings and the MaxQuant internal normalization algorithm MaxLFQ was used with a minimum ratio count of one. The R-package PTXQC was to identify samples with low quality [115].

2.5. Downstream analysis proteomic data

Downstream analysis of proteomic data was conducted in the programming language R version 4.3.1 [150]. The result files from MaxQuant and DIA-NN containing the information of protein intestines were loaded into the R-workspace and integrated together with clinical sample information was well as protein meta-information into a S4-object of the class 'Summarized Experiment'(SE) [151]. For the DIA data analysed by DIA-NN, the long formatted result.tsv was used and a quality filtering step was performed during integration by filtering out precursors and proteins with a q-value above 0.01 followed by log2 transformation of the normalized protein intensities of the genes identified. For in-house DDA data analysed by MaxQuant, the 'protein groups'-files were loaded input. Proteins which were reverse, contaminants or only identified by site according to MaxQuant meta information were removed followed by log2 transformation of the protein intensities. Protein tables from external datasets already were filtered and log2 transformed and could directly be loaded into a SE for further analysis.

2.5.1. Data preparation

2.5.1.1 Data preparation DIA cohort

For the DIA data cohort, samples, with less than 5000 identified proteins were considered as low quality and were subsequently removed from further downstream analysis. Two different protein filtering strategies were applied for further analysing the DIA study cohort in context of the specific questions addressed. In this regard, protein filtering for subsequent unsupervised analysis was performed under a strict complete case analysis where proteins with missing values were excluded for subsequent analysis. For characterization of the identified subtypes a permissive filtering rule was applied in the context of the proteomic subtypes to account for group specific missingness. In this context, proteins were retained for analysis when they were present in at least 80% of the samples within one of the three groups. Subsequently, distribution of the remaining missing values across the groups was inspected by using a binary Heatmap showing available and missing values.

After filtering, protein intensities were normalized using median Centering normalization implementation from the R-package 'proBatch' [152] and controlled using a bar-graph distribution. Principal component analysis (PCA) was conducted to identify potential confounding effects before and during data preparation. In addition, a PCA-Anova-test was conducted on the principal components (PC) to detect batch effects and biological grouping on the given variables 'time-batch' (June 2024, March 2025), 'WHO grade', 'identified precursors', 'identified proteins'. Discovered batch-effects of samples analysed in March and June were corrected by applying the *combat()*-function implemented in the 'proBatch' [152] on the normalized protein intensities.

Remaining missing values were imputed by a mixed imputation strategy was applied using the functions implemented in the R-package impute [153]. Proteins with more than 80% missingness in a group were imputed by a left-censored imputation strategy performing random draws from a Gaussian distribution centred in a minimal value (minProb) while K-Nearest neighbours (KNN) imputation [154] from the was used for the others.

2.5.1.2 Data preparation DDA data

Inhouse DDA data as well as DDA data obtained from *Wong et al.* [132] was prepared for further analysis by removing samples with less than 3000 identified proteins from further downstream analysis. For the evaluation of differential expressed proteins and MYC expression, permissive filtering was applied in the context of predicted proteomic subtypes to account for group specific missingness. Proteins were retained for analysis when they were

present in at least 80% of the samples within one of the three groups. As there was a strong association of proteins with low intensities and missing values, missing values were imputed using left censored minProb imputation.

2.5.2. Unsupervised analysis and proteomic subtyping

Unsupervised analysis of AIDH in the DIA study cohort was conducted by performing a dimensional reduction in a two-step procedure. The log2 median normalized and batch corrected intensities of 3042 Proteins were first analysed by PCA using the function *pcromp()* from the R-package stats [150] with scaling and centring of variables enabled. Next, the 'screeplot' function from the stats-package [150] was applied to determine the PC's which contribute most to variability using the elbow approach [155]. Relevant components were used as input for Uniform manifold approximation and projection (UMAP) to investigate the grouping pattern of samples by their biological groups in a two-dimensional space. Subsequently, the *umap()*-function from the equal named R-package [156] was used to calculate a UMAP using the initial 8 PCs, a minimum distance of 0.1 and 15 neighbours and the naïve UMAP R implementation as metrics.

Grouping of the AIDH according to their protein expression profiles was done performing a random walk-based clustering (RWC). The log2 median normalized and batch corrected intensities of 3042 Proteins were used for constructing a nearest neighbour graph using the *nng()*-function (k(neighbours)=15, algorithm= "kd_tree") from the R-package cccd [157] and the function *cluster_walktrap()* from the R-package igraph [158] was used to discover expression communities in the graph via random walks.

2.5.3. Differential protein expression analysis

Differential protein expression analysis (DEA) was performed by using the functions provided by the R package limma [127]. Two distinct analysis strategies were applied to ascertain marker proteins for each subtype as well as shared in proteomic subtypes. Subtype marker proteins were ascertained by testing each proteomic subtypes against the average expression of the other two subtypes using the following formula: *Subtype A vs (Subtype B + Subtype C)/2*. Common protein expression was evaluated by performing all possible group wise comparisons. Proteins of interest were defined when the absolute Log_2 fold-change (logFC) was higher than 1.5 and the adjusted P-value was smaller than 0.05.

2.5.4. Pathway analysis

Gene set enrichment analysis (GSEA) was used to evaluate the biological meaning of lists of proteins obtained by DEA. GSEA was performed using the *fgsea()*-function from the R-

package fgsea [159]. The logFC, p-value and t-test statistics from the limma analysis were used as input and tested against Genesets from Biological Knowledge Databases (Section 2.2.1.2.) using 10000 permutations, with a minimum gene set size of 5 and a maximum of 200. Pathways of interest were selected when the adjusted P-value was smaller than 0.05. Semantic similarity analysis was performed to reduce the redundancy in the identified GO pathways of interest to identify major biological contribution in context of the identified groups. This was done by using the functions implemented in the R-package simplifiyEnrichment [160]. Semantic similarity of GO-Terms of interest was estimated using the Lin algorithm and clustered according to the walktrap method in the *simplifyGO*()-function.

Furthermore, single sample gene set enrichment (ssGSEA) was used to investigate enrichment of gene sets of interest between proteomic subtypes in a per-sample manner. Per sample ssGSEA scores were determined by using the *gsva()*-function from the R-package GSVA [161] with a minimum gene set size of 5. Furthermore, overrepresentation of proteins with cluster-membership were in GO-Terms was evaluated using the functions implemented in the R-package clusterProfiler [162]. Proteins of each cluster were tested against GO Gene sets form the biological process category, while the entire set of proteins in the dataset was used as background. A p-value of 0.05 was used as cutoff for processes of interest.

2.5.5. Development and validation of a proteomic stratification model

The proteomic prediction framework was developed and tested by using machine learning functions and algorithms from the R-packages mlr3, mlr3learners and mlr3verse [163]. Sample contributions to specific subtypes were determined by estimating a per-sample scores based on subtype-specific marker protein signatures. Single sample protein signature scores were calculated using the *viper*-function, which provides more context sensitive activity estimates and more robust normalization compared to ssGSEA [164]. Viper was applied to log2 median normalized and batch corrected intensities of 6853 proteins using a minimum set size of 10, pleiotropy correction, enrichment score normalization and the rank scoring method enabled [164].

For model development, the dataset was randomly (Seed 2905) split into 70% for training and 30% for testing. A decision-tree learner (mlr3 learner rpart) was trained using 10-fold cross validation. Hyperparameter tuning was performed in 50 tuning iterations to optimize balanced accuracy (BACC) over the following search space: minimum split = 2-10, maximum depth = 1-5 and complexity parameter = 0.01-0.05. Model performance was assessed on the held-out

test data by estimating BACC and accuracy (ACC). The final decision tree was visualized by using the *rpart*-function from the R-package rpart [165].

2.6. Genome wide DNA-methylation analysis

2.6.1. Beta value acquisition

Beta values in step 2.6.1. were obtained from the routine diagnostic analysis pipeline developed and maintained by Dr. Daniel Schrimpf Neuropathology Heidelberg.

In accordance with the manufacturer's protocols, the genome-wide DNA methylation analysis was conducted using the Infinium Methylation 450k and Illumina MethylationEPIC (EPIC) BeadChip (Illumina, San Diego, USA), as previously described by our group in [33]. Data analysis was performed using R version 4.3.1. [166]. Samples underwent an individual background and colour channel dye-bias correction for normalisation, as previously reported [33]. Background correction was achieved by shifting the 5th percentile of the negative control CpG to 0. For dye-bias correction, both colour channels had their normalisation control CpG intensities scaled to 10,000. Subsequently, a batch correction between array types (450k/EPIC) was performed on the log2-transformed intensity values using the *removeBatchEffect()* function from the R-package limma (version 3.30.11) [127]. Retransformed intensities were employed to derive beta-values, with an offset of 100, as per the recommendations provided by Illumina.

2.6.2. Unsupervised analysis of beta values

The beta values of the top 50000 CpG positions ranked by variability were used for principal component analysis (PCA). The *screeplot*()-function from the stats-package [150] was applied to determine the components which contribute most to variability. The initial 9 PCs were used as input for UMAP with a minimum distance of 0.1 and 15 neighbours and the naïve UMAP R implementation as metrics.

2.6.3. Estimation of different methylated CpGs

Differential methylated positions (DMP) were analysed in the context of proteomic subtypes using the functions implemented in the R-package 'limma' [127]. Analysis was carried out on the M-values of the top 50000 CpG form unsupervised analysis (2.6.2). Relevant CpGs were defined if they had a P-value was <0.01 and a logFC larger than 0.58. The Illumina Infinium HumanMethylationEPIC manifest [167] was used to annotate CpGs of promoter and enhancer sites. CpGs in the Volcano plots are distributed according to their -log10 P-values and log fold change.

2.6.4. Deconvolution of methylation data

Cell type deconvolution from Methylation data was performed by using the online version of MethylCIBERSORT [168] (https://cibersort.stanford.edu). The deconvolution was applied on the beta values of the study cohort and signature matrix from reference methylation profiles of flow-sorted cell populations from [169] downloaded from (https://github.com/mwsill/mcibersort_scripts/). MethylCIBERSORT was conducted using 1000 permutations using relative scores without quantile normalization.

2.6.5. Inference of copy number status and CNV load

Analysis of copy number variation on the level of genes, segments and chromosomal arms was carried out on methylation data by using the R-package conumee [170] . CNV burden was calculated by determining the total number of altered arms per sample. Summary CNV plots were generated by segmenting the chromosomal arms into bins of 1%, thereby enabling the estimation of the relative frequency of gains and losses within each group. These values were subsequently plotted alongside a genomic axis that spans from 1 to 22 using the plot()-function in R. Somatic copy number alteration (SCNA) of gene louses of interest were estimated by using the functions provided by the R-package conumee [170]. Gene SCNA status was determined for 17 gene louses tumour suppressors CDKN2A/B, RB1, NF1, NF2, PTEN, SMARCB1 and oncogenes CCND1, CCND2, CDK4, CDK6 as well as EGFR, MYC, MYCN, MET, and PDGFRA based on the thresholds defined by Shirahata et al [13].

2.7. Survival analysis

The overall survival (OS) and progression-free survival (PFS) probabilities of patients stratified by WHO criteria and proteomic subtypes was estimated by using the functions provided by the R-packages survival [171] and survminer [172]. Kaplan-Meier plots were visualized by using the *ggsurvplot()*-function provided by survminer [172]. Further analysis of the prognostic significance of groups, CNV-load and SCNA status in context of clinical confounders such as patient age, gender and therapy modality was evaluated by fitting the parameters into a cox regression model. Here, the functions implemented in the R-package finalfit [173] were used for the calculation of the model, while the function *ggforest()* from the R-package survminer [172] was used to calculate the Akaike information criterion (AIC) for evaluating the goodness of the model fit and to visualize the resulting cox-regression model in a forest plot.

2.8. Statistics and Visualization

All analysis steps of the downstream proteomic, genome wide methylation, and survival analysis were conducted in the programming language R [150] version 4.3.1. or higher, (https://cran.r-project.org/), Bioconductor [174] version 3.20 (https://bioconductor.org/) or higher. Unless stated otherwise, for all boxplots, pairwise comparisons of protein signature scores were performed using Students-t-Test and Wilcoxon signed rank tests and normality of data was assessed by using the Shapiro-Wilk test [175]. All reported p-values were adjusted using the Benjamini-Hochberg procedure [176]. Heatmaps were generated using circlize [177] and ComplexHeatmap [178]. Venn diagrams were generated using VennDiagram [12], survival plots were generated using survival [171] and forest plots using survminer [172]. All additional graphical representations were generated using ggplot2 [179], ggpubr [180] and patchwork [181]. Additional colour palettes were acquired from the R-package RcolorBrewer [182] .

2.9. Artificial intelligence assisted language tools

Artificial intelligence (AI) assisted language tools were used for proofreading, stylistic, and grammatical refinements of the thesis. These tools included the large language model-based writing assistant ChatGPT (version 5.1.) [183] and writing tools provided by Apple Intelligence on a Mac (version macOS Sequoia version 15.6.1) [184].

3. Results

This study aimed to identify new opportunities for clinical stratification of lower-grade AIDH by exploring their proteome. The following results are structured into three subsections. The initial section shows findings from the unsupervised analysis of the proteome of AIDH, detailing the clinical contribution of proteomic defined subtypes. A comprehensive comparison with current WHO grading, clinical parameters (age, gender, therapy modalities), epigenetic features, genomic alterations and cell type composition was conducted to further characterize prognostic relevance of proteomic subtypes. The second section focused on proteins associated with AIDH malignancy, highlighting protein groups and biological programs linked with the distinct clinical outcomes within the discovered proteomic subtypes. The third section presents a novel prediction model for clinical stratification of low-grade AIDH patients and demonstrates the application of the framework to in-house and publicly available data to validate findings from the previous sections.

3.1. Proteomic investigation of AIDH with different clinical outcomes

To investigate the prognostic potential of the AIDH proteome, I analysed whether cases with different clinical outcomes can be distinguished based on their protein expression pattern. Therefore, a global proteome analysis of a clinically and epigenetically well-documented cohort of AIDH cases was curated and analysed by LC-MS in DIA-PASEF mode. An unsupervised analysis was performed to ascertain the grouping of AIDH cases according to their proteomes and to further evaluate whether the grouping associates with clinical outcome.

3.1.1. AIDH cohort baseline characteristics

The study cohort curated for whole proteome profiling comprehended 80 cases with primary AIDH diagnosed according to CNS5 WHO criteria [1]. The main characteristics of the study cohort are depicted in Figure 6. The cohort included 29 cases with grade 2, 31 with grade 3 and 20 with grade 4. To assess the prognostic significance of the proteomic data, clinically relevant information including patient age at diagnosis, gender, treatment, progression-free and overall survival years were gathered. The average patient age was 37.5 years, with a range of 18 to 62 years. The cohort included 31 males, 41 females and 5 anonymized cases. Treatment information was provided for 75 cases: 49 cases received no treatment after their initial

diagnosis, 24 patients received radiotherapy and chemotherapy (RTC) while one patient received treatment with radiotherapy alone.

Epigenetic data was available for 71 cases to enable further evaluation of the findings of the proteomic investigation in the context of CpG methylation status and copy number alterations. Furthermore, epigenetic data was used to determine the methylation class according to the Heidelberg Methylation Classifier (version 12.8) [138]. The majority of low grade 2 and 3 AIDH were classified as Astrocytoma, IDH-mutant low-grade (A_IDH_LG) and only 6 were grouped into the Astrocytoma, IDH-mutant high-grade (A_IDH_HG) class. Conversely, majority of grade 4 AIDH were classified as A_IDH_HG and only 3 were classified as A_IDH_LG. Notably, there were 6 AIDH classified into other groups referring to Oligodendroglioma, IDH mutant (O_IDH; n=3) and Oligosarcoma, IDH mutant (OLIGOSARC_IDH; n=3). However, these predictions were found to be incorrect, as none of the cases exhibited co-deletion of chromosomal arms 1p/19q, which is a diagnostic marker for Oligodendroglioma [1]. Furthermore, CNV statuses of CDKN2A/B and other potential malignant progression associated gene level somatic copy number alterations were ascertained using the thresholds described by *Shirahata et al* [13].

Proteomic profiles were acquired with LC-MS in DIA-PASEF mode. After removing precursors and proteins with an FDR below 1%, the dataset still consisted of 149 698 high-confidence precursors and 11 623 proteins. After exclusion of samples with low protein identification, the proteomic dataset contained 77 samples for subsequent bioinformatic analysis.

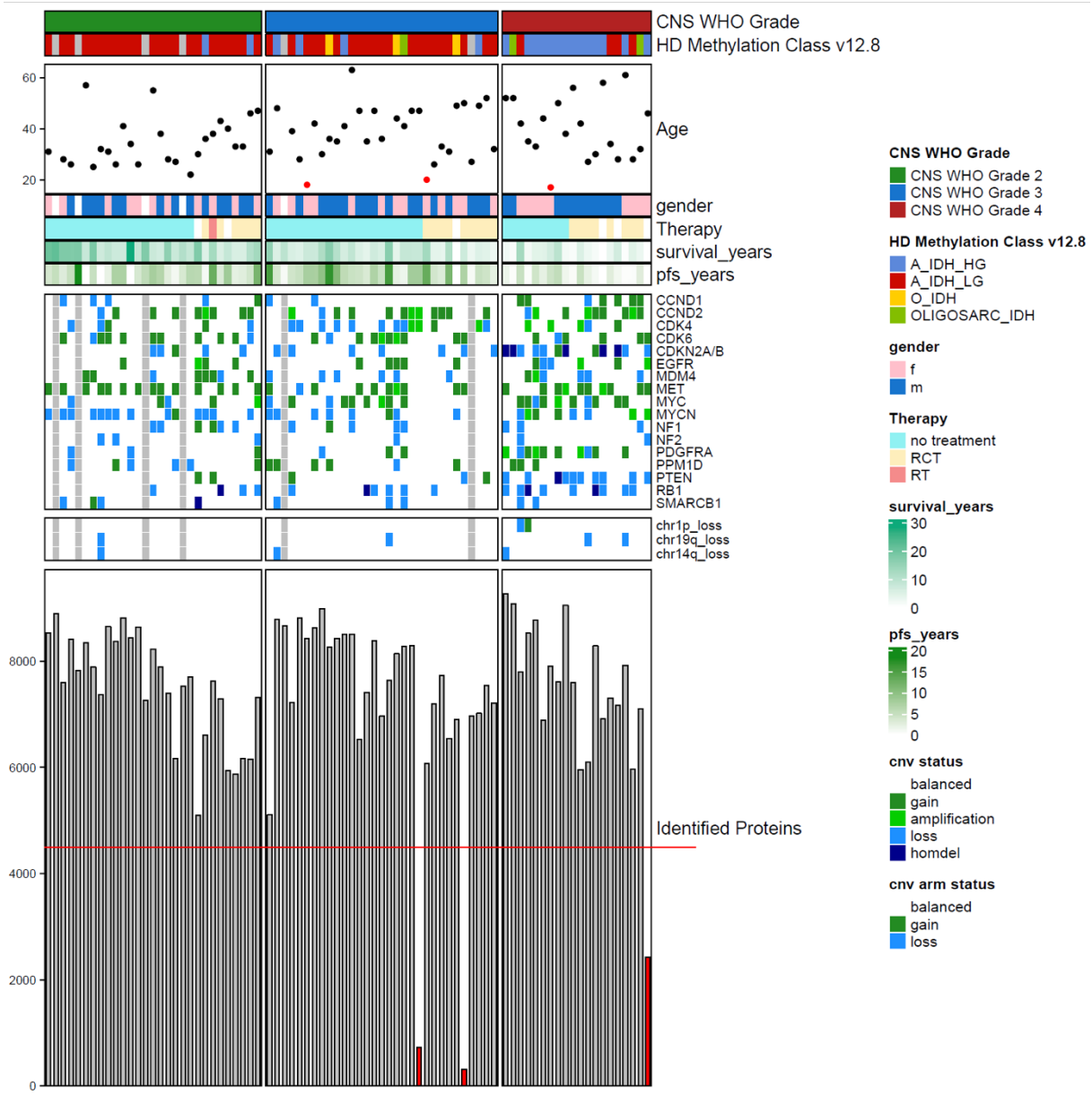


Figure 6: Baseline characteristics of the AIDH study cohort (n=80).
Summary heatmap depicting main molecular and clinical characteristics of the AIDH cases curated for proteomic analysis. Missing molecular values of samples without further epigenetic information are coloured as grey (n=6). Columns of the Heatmap were split according to the CNS5 WHO grades to visualize the contribution of the parameters according to the current gold standard of AIDH risk stratification. The first oncoprint shows the gene level somatic copy number alteration (SCNA) for the tumour suppressors CDKN2A/B, RB1, NF1, NF2, PTEN, SMARCB1 and oncogenes CCND1, CCND2, CDK4, CDK6 as well as EGFR, MYC, MYCN, MET, and PDGFRA. The second oncoprint shows the status of the chromosomal arms 1p, 19q and 14q. Colours in the oncoprint indicate copy number statuses balanced, gain, amplification, loss and homozygous deletion. A bar chart shows the number of identified proteins for each sample before preprocessing. A red line in the bar chart indicates the threshold for the minimum number of proteins required for inclusion for downstream analysis. Samples excluded from further downstream analysis are highlighted in red (n=3) while other are coloured grey in the bar chart. Abbreviations: CNS, Central Nervous System; CNV, Copy number variation; HD, Heidelberg; RCT, Radiotherapy and Chemotherapy; RT, Radiotherapy; PFS, Progression-free Survival; WHO, World Health Organization.

3.1.2. Unsupervised analysis identifies three novel proteomic subtypes associated with distinct clinical outcomes

An unsupervised analysis was carried out in two steps in the absence of proteins with missing values to account for the potential risk of artificial confounding by imputation. The Log2 median normalized and batch corrected intensities of 3042 proteins were analysed by principal component analysis to reduce data dimensionality and redundancy. The initial 8 principal components (PCs) were determined as the most relevant from a scree-plot representation by using to the elbow approach (Figure 7). Subsequently, informative PCs were submitted in a second step to Uniform manifold approximation and projection (UMAP) to investigate the grouping pattern of AIDH cases based on their proteomic profiles in a two-dimensional space.

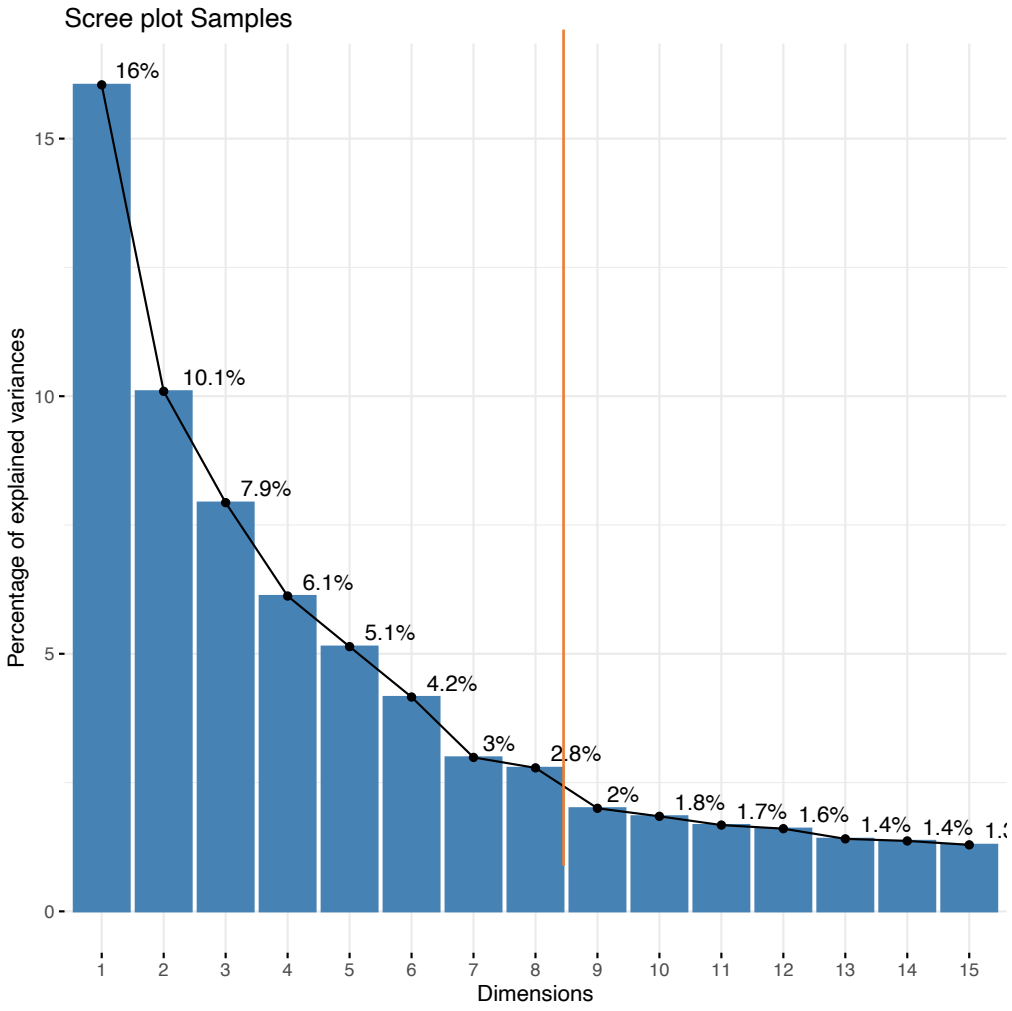


Figure 7: Scree-plot representation of the total variance in the initial 15 principal components.
Bar plot representation of the percentage of variance within each principal component (PC). A red line highlights the elbow cutoff which was used to determine the most relevant components for subsequent UMAP analysis. The plot shows a drop in percentage of explained variance at PC 8.

The UMAP showed AIDH cases as a range of connected groups (Figure 8A), which is consistent with AIDH consisting of a biological continuum from a lower to a higher-grade malignancy [3, 4]. Notably, WHO grade 4 AIDH showed substantial separation in the UMAP from low grade AIDH. In contrast, low grade 2 and 3 cases were heterogeneously distributed suggesting difference between histological grading and the AIDH proteome.

To investigate the low-grade AIDH groups indicated by UMAP, a Random Walk Clustering (RWC) was used to identify clusters based on proteomic expression profiles. RWC has been shown to be superior in identifying groups in continuous data, as suggested for AIDH [3, 4]. RWC simulates a stochastic walk through a graph of sample similarities, identifying groups of samples that tend to be more connected with each other [185]. To align with the suggested pattern of connected groups in UMAP, RWC was performed with the same number of neighbours (k=15) as used for UMAP calculation. This resulted in the identification of three novel proteomic AIDH subtypes that exhibited strong overlap with the UMAP group distribution (Figure 8B). The subtypes were named after their order in the UMAP to emphasize their potential contribution in the malignant continuum. The group with homogeneous grade 4 AIDH was termed AIDH high-grade (AIDH_HG). The middle group was called AIDH intermediate-grade (AIDH_INT), and the upper group in the UMAP was called AIDH low-grade (AIDH_LG).

Exploration of the distribution of WHO grades in the proteomic subtypes (Figure 8 C) showed that the vast majority of WHO grade 4 were distributed in AIDH_HG (77%). In contrast, more than half of WHO grade 2 (56%) and a substantial subset of WHO grade 3 AIDH (27%) were distributed in AIDH_LG. AIDH_INT showed heterogeneous distribution from all WHO grades, including 60% of grade WHO 3, 44% of WHO grade 2, and 20% of WHO grade 4, underscoring recent assumptions that it might reflect an intermediate group in the context of malignant progression.

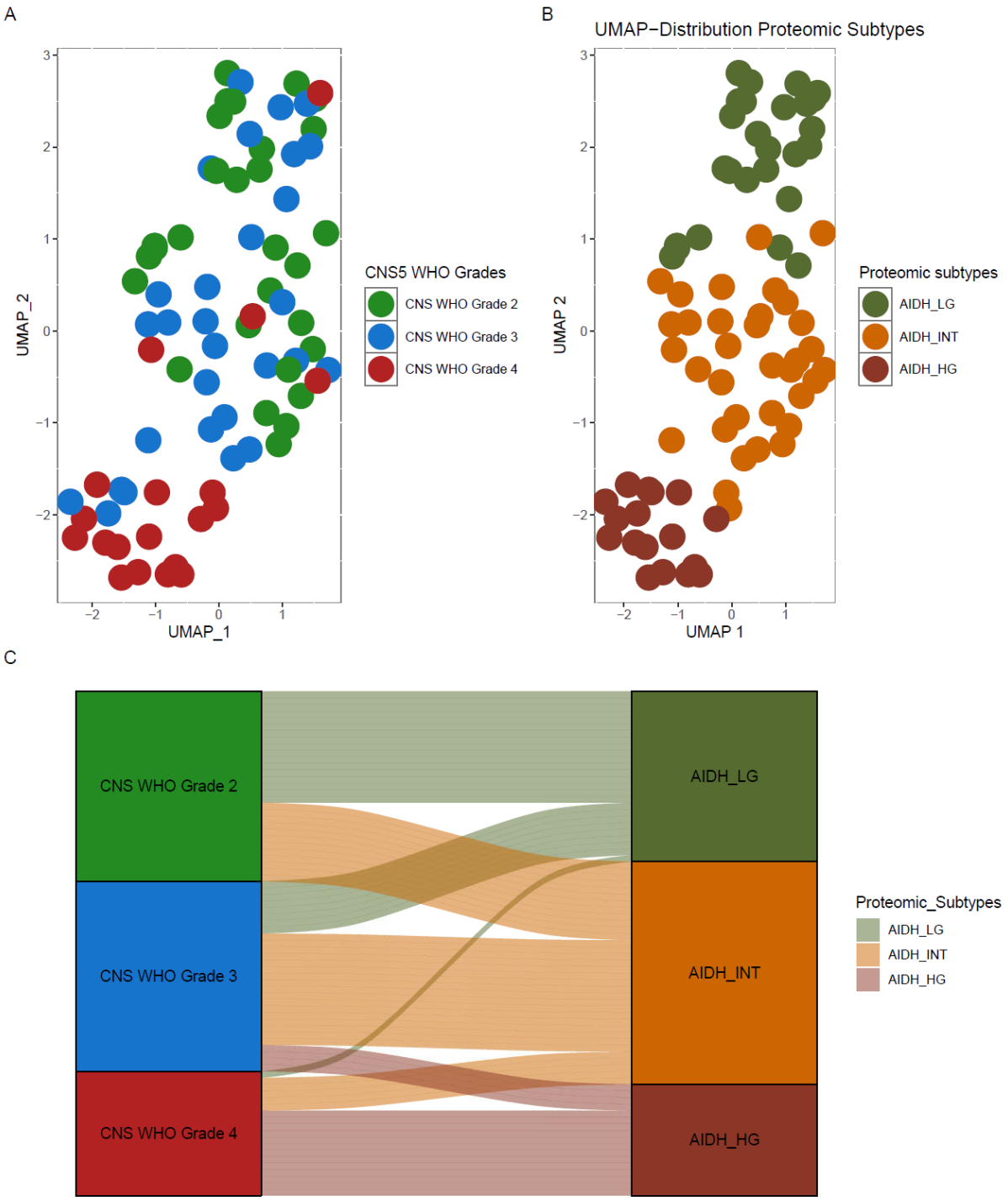


Figure 8: Unsupervised analysis reveals three novel proteomic AIDH subtypes (n=77).
Plots depicting main findings from unsupervised analysis of AIDH proteomes. UMAP plots showing AIDH represented as dots and colours assigned to A) CNS5 WHO grading B) proteomic subtypes. C) Sankey plot showing the distribution of CNS5 WHO grades in proteomic subtypes. UMAP was estimated on the initial 8 principal components preselected by the elbow approach and k= 15 for the number of neighbours. Proteomic subtypes were identified by random walk clustering using the same k used for the UMAP to preserve consistency with the groups shown. Proteomic subtypes AIDH_HG, AIDH_INT and AIDH_LG are represented by the colours red, orange and green. WHO grades are represented by light green, blue and light red. Abbreviations: AIDH_HG, Astrocytoma IDH mutant high-grade; AIDH_INT, Astrocytoma IDH mutant intermediate; AIDH_LG, Astrocytoma, IDH mutant low-grade; CNS, central nervous system; IDH, Isocitrate dehydrogenase; UMAP Uniform Manifold Approximation and Projection; WHO, World Health Organization.

To assess the link between the discovered proteomic subtypes and tumour malignancy, the association with clinical outcomes was evaluated. A survival analysis compared patient overall (OS) and progression-free survival (PFS) based on the proteomic subtypes. Kaplan-Meier analysis of the overall survival of AIDH patients showed consistency with the postulated malignant progression states of proteomic subtypes (Figure 9 A). As anticipated, AIDH_LG had the most favourable survival probabilities with a median OS of ~22 years, while patients grouped into AIDH_INT had an intermediate outcome (median OS =~16; Log-rank p =0.029). Patients stratified as AIDH_HG demonstrated the most dismal outcome (median OS= ~5 years; Log-rank p p=0.001) which is consistent with the overall worse survival predicted for WHO grade 4 AIDH in this group (Figure 9 A). Investigation of the progression-free survival (PFS) showed a similar decrease in association with observed patient OS (Figure 9 B); however, the overall effect was minor (Log-rank p =0.054) compared to patient OS (Log-rank p=0.00052).

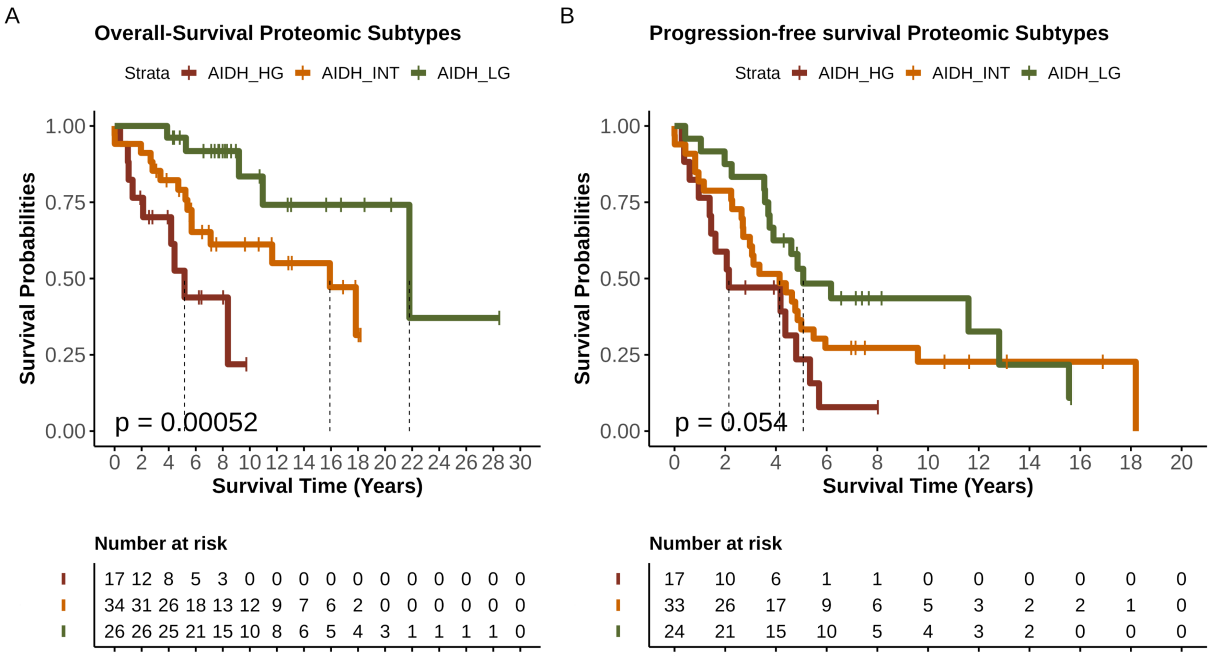


Figure 9: Proteomic subtypes of AIDH show significant association with overall survival (n=77).
Kaplan Meier Plots showing the overall survival A) and progression-free survival B) probabilities of AIDH patients over years stratified by proteomics subtypes. Patients were stratified into proteomic subtypes by random walk-based clustering and significance of survival outcomes was ascertained by performing the log-rank test. Proteomic subtypes AIDH_HG, AIDH_INT and AIDH_LG are represented by red, orange and green colours, respectively. The plot show significant differences in patient overall survival AIDH_LG and AIDH_INT even in absence of grade 4 AIDH cases in AIHD_INT. Abbreviations: AIDH_HG, Astrocytoma IDH mutant high-grade; AIDH_INT, Astrocytoma IDH mutant intermediate; AIDH_LG, Astrocytoma, IDH mutant low-grade; OS, Overall survival; PFS, progression free survival

To further validate that the survival differences observed between AIDH_LG and AIDH_INT were not confounded by clinical aggressive WHO grade 4, the survival analysis was re-evaluated, under their exclusion. Nevertheless, the Kaplan-Meier analysis showed a significant (p = 0.029) distinction between the OS of low-grade AIDH stratified into AIDH_LG or

AIDH_INT (Figure 10). Thus, this analysis identified three novel proteomic AIDH subtypes, which showed an association with the progressive continuum of AIDH. Notably, it also highlighted high heterogeneity in the distribution of WHO grade 2 and 3 in low-grade AIDH, introducing novel insights into these tumour types. Furthermore, evaluation of patient overall survival demonstrated a clinical association between proteomic subtypes and favourable, intermediate, and dismal outcomes, suggesting a potential link between protein expression and AIDH malignancy.

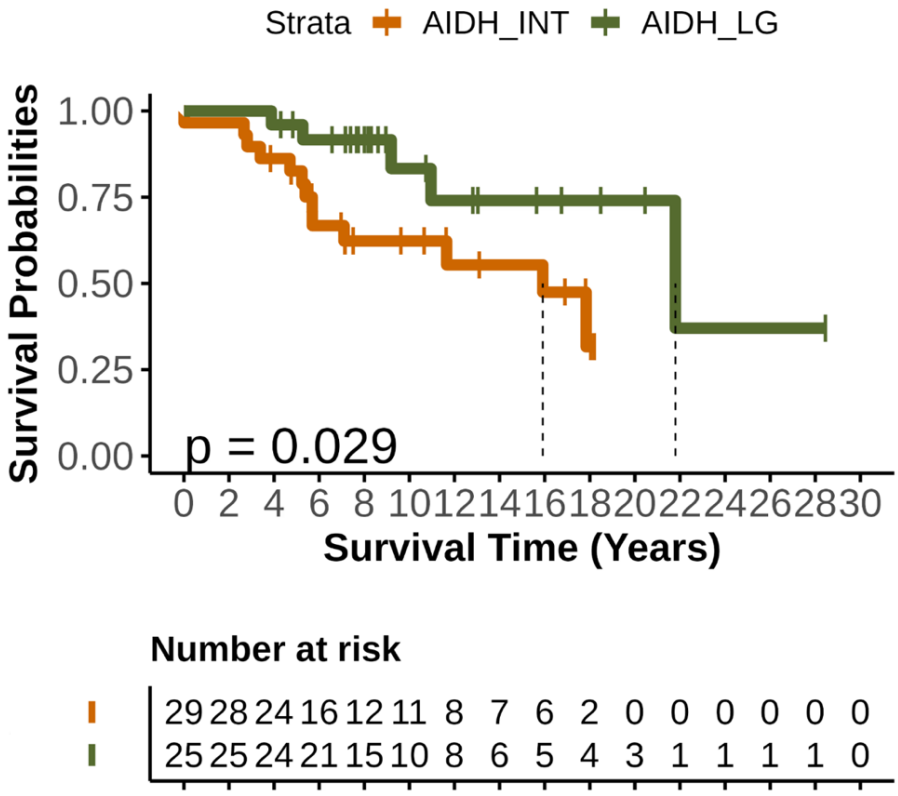


Figure 10: Comparison of the overall survival of low-grade AIDH patients shows significant differences between AIDH_INT and AIDH_LG.
Kaplan-Meier plot showing the overall survival probabilities of low-grade AIDH cases according to the proteomic subtypes AIDH_LG and AIDH_INT (n=54). Cases were stratified into proteomic subtypes by random walk-based clustering and significance of survival outcomes was ascertained by performing the log-rank test. Proteomic subtypes AIDH_INT and AIDH_LG are represented by orange and green colours, respectively. The plot shows significant differences in patient overall survival AIDH_LG and AIDH_INT even in absence of grade 4 AIDH cases in AIHD_INT. Abbreviations: AIDH_INT, Astrocytoma IDH mutant intermediate; AIDH_LG, Astrocytoma IDH mutant low-grade; OS, Overall survival.

3.1.3. Clinical evaluation of proteomic subtypes

Given the observed associations between proteomic subtypes and patient overall survival, I conducted a clinical evaluation to determine their prognostic significance and potential to enhance current low-grade AIDH patient risk stratification. In this context, proteomic subtyping was compared to CNS5 WHO grading to determine whether it provides better stratification for low-grade AIDH patients than the established standard. Furthermore, the distribution of genetic and epigenetic were investigated in the context of proteomic subtypes to evaluate their association with distinct clinical outcomes.

3.1.3.1. Proteomic subtyping is superior to WHO grading

Survival outcomes of patients were evaluated by Kaplan Meier and cox proportional hazard models adjusted by patient age, gender and therapy modality. Analysis of the different survival outcomes showed a superior clinical separation for AIDH stratified by proteomic subtypes (Figure 11). In this context, the Cox regression models adjusted by age, gender and therapy modality showed a better survival discrimination (log-rank $p=4.6 \bullet 10^{-5}$) and better overall model fit (AIC = 164.11) (Supplemental Figure 1 B) for proteomic stratification compared to WHO grading (log-rank $p = 9.2 \bullet 10^{-4}$; AIC=170.88) (Supplemental Figure 1 A). Cases grouped into AIDH_INT exhibited an 8.5-fold higher hazard ratio (HR) of death (HR=8.53; 95% CI 1.91-38.0) compared to patients grouped as AIDH_LG. Importantly, the clinical association between AIDH_LG and AIDH_INT remained significant even after exclusion of WHO grade 4 cases (Figure 11, Supplemental Figure 1 C). In contrast, grade 2 and grade 3 AIDH showed only minor difference (HR 2.24; 95% CI 0.67-6.4; p=0.086) suggesting less discriminatory power. Unexpectedly, comparison of the high-grade groups AIDH_HG and WHO 4 revealed that proteomic subtyping resulted in a better hazard prediction (HR 16.46; 95% CI 3.98-69.6; p <0.001) than the WHO grade 4 classification (HR 6.46; 95% CI 2.19-6.64; p<0.001).

These results show that proteomic subtypes provide a superior prognostic resolution than WHO grading. Furthermore, it underscores the prognostic value of the proteomic subtypes and suggests it may refine clinical stratification making not only for low but also for high grade AIDH.

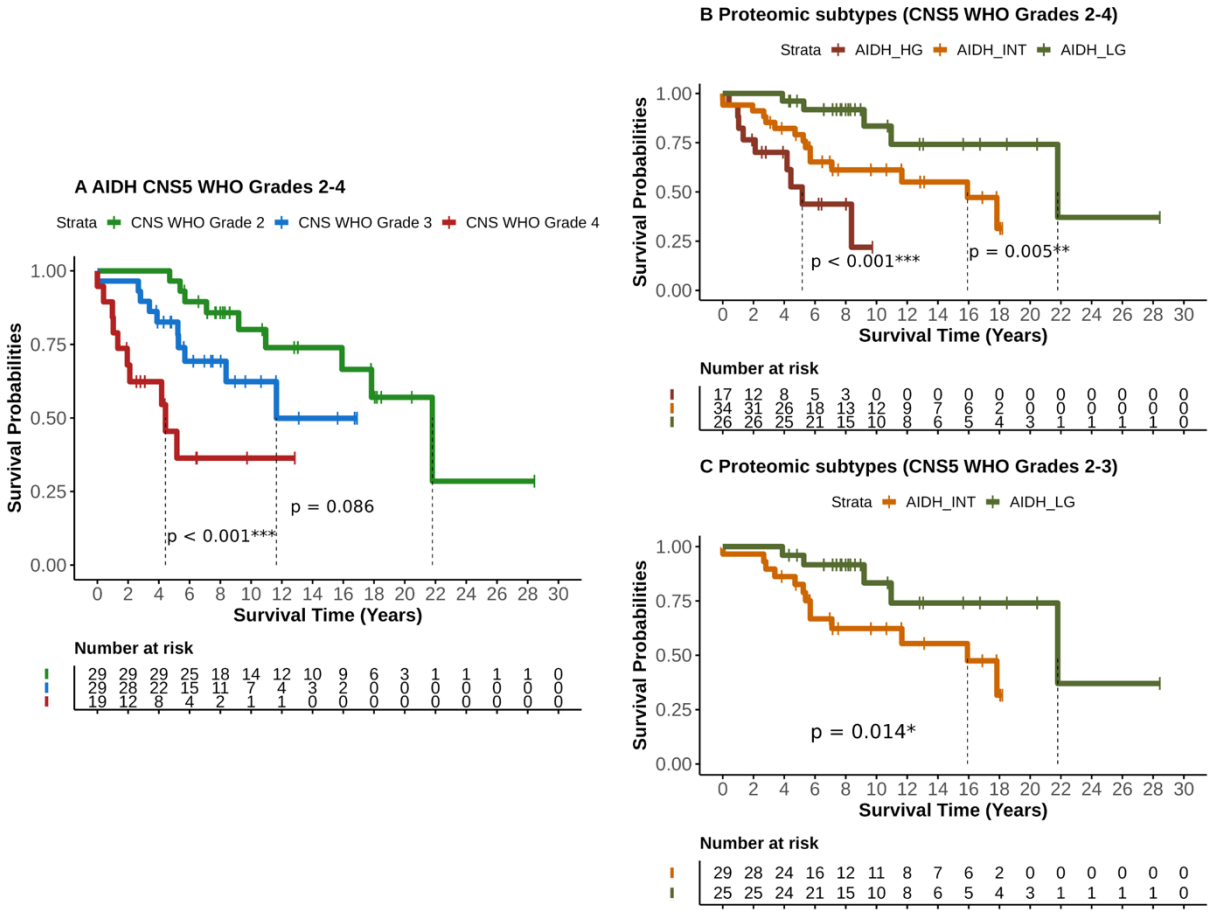


Figure 11: Proteomic subtypes show superior stratification of low-grade AIDH patients.
Kaplan-Meier plots showing the comparison of overall survival probabilities of AIDH patients stratified by A) CNS5 WHO grading criteria and by proteomic subtypes in presence B) and absence C) of grade 4 AIDH. P-values were estimated by Cox proportional hazard models adjusted for patient age, gender and therapy modality. Proteomic subtypes AIDH_HG, AIDH_INT and AIDH_LG are represented by the colours red, orange and green. WHO grades are represented by light green, blue and light red. The plots show that patients stratified by proteomic subtypes show a superior clinical separation compared to CNS WHO grading. Abbreviations: AIDH_HG, Astrocytoma IDH mutant high-grade; AIDH_INT, Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma IDH mutant low-grade; CNS, central nervous system; IDH, Isocitrate dehydrogenase; UMAP Uniform Manifold Approximation and Projection; WHO, World Health Organization.

3.1.3.2. Prognostic differences between AIDH_LG and AIDH_INT are independent from global CNV alterations

Increased burden in copy number variations (CNV) have recently been associated with a dismal prognosis in high grade AIDH [13, 17, 28, 29, 35, 186]. Based on this, I evaluated the influence of CNV burden in the context of proteomic subtypes. Subsequently, copy number alteration profiles of the AIDH were determined to analyse the association between chromosomal copy number alterations (CNVs) in context of proteomic subtypes. The CNV burden was calculated for each sample with available epigenetic data (n=71) by determining the total number of altered chromosomal regions.

As shown in Figure 12 A, AIDH_HG had a much higher CNV burden (>400 Mbp), which is consistent with recent studies [13, 17, 28, 29, 35, 186] describing genomic instability linked with high grade AIDH. In contrast, AIDH_LG and AIDH_INT showed comparable levels of CNV burden suggesting that increase of global CNV is primary associated with high grade AIDH. This hypothesis was further supported by the prognostic association of CNV burden in context of proteomic subtypes. Evaluation with Cox proportional models showed that CNV burden was only relevant in the univariate model and in presence of high-grade AIDH (Figure 12 B) while it lost significance in the absence of grade 4 AIDH (Figure 12 C). Thus, these findings show that the prognostic differences between AIDH_LG and AIDH_INT are independent from the global CNV-burden.

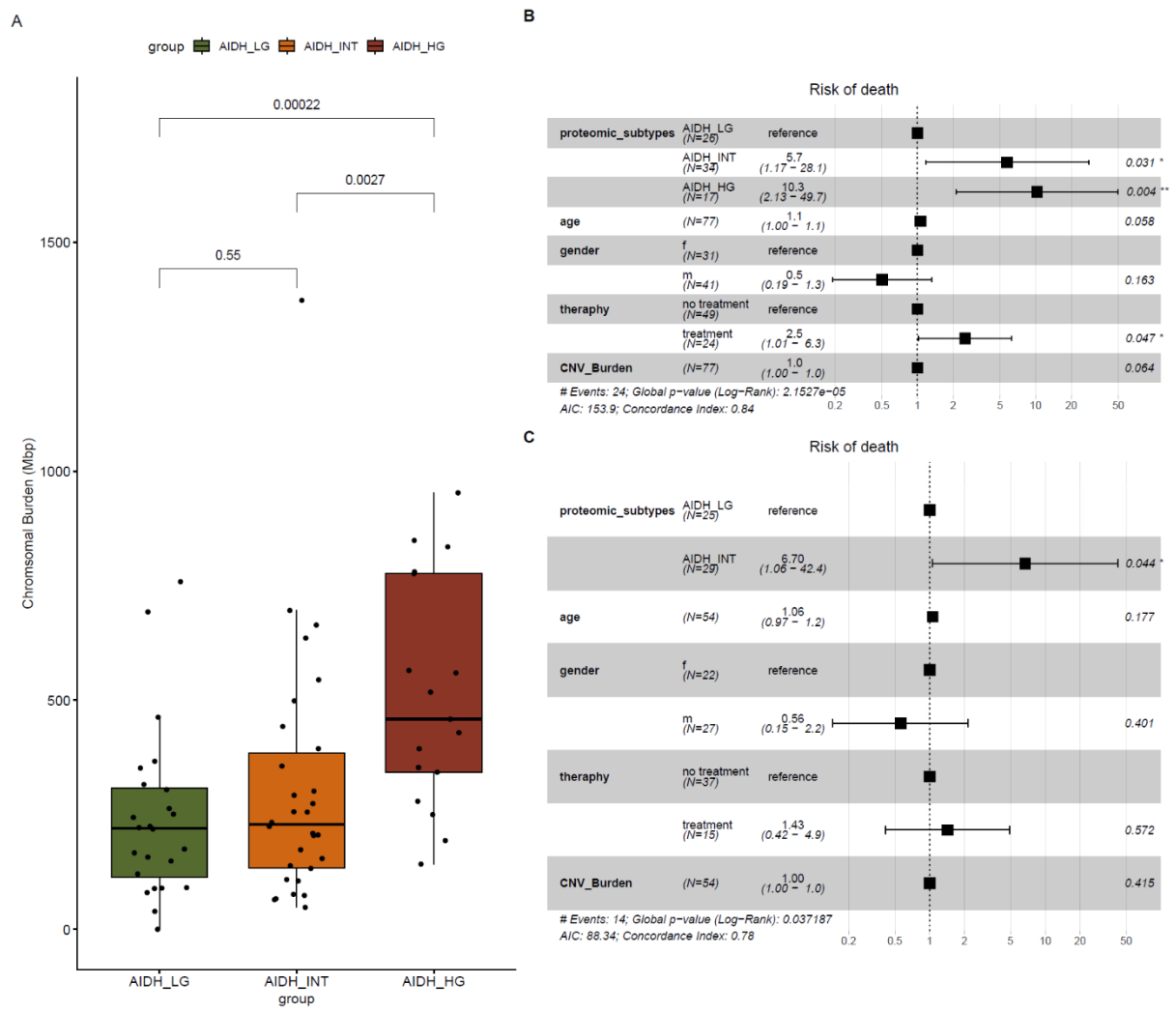


Figure 12: Prognostic differences between AIDH_LG and AIDH_INT are independent from copy number variation (CNV) burden.
Plots showing the A) distribution of CNV burden in context of proteomic subtypes and prognostic significance in forest plots in the B) presence and C) absence of WHO grade 4 AIDH cases. The CNV burden (Mbp) was estimated by calculating the total length of altered segments per sample, and a Wilcoxon-Rank Sum test was performed to compare groups. Proteomic subtypes AIDH_HG, AIDH_INT and AIDH_LG are represented by red, orange and green colours, respectively. The prognostic relevance was determined using a cox proportional hazard model adjusted for proteomic subtype, patient age, gender and therapy modality. The plots indicate that CNV burden is primarily relevant in high-grade AIDH in AIDH_HG. Abbreviations: AIDH_HG, Astrocytoma IDH mutant high-grade; AIDH_INT, Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma IDH mutant low-grade;

3.1.3.3. AIDH_INT and AIHD_HG with partial loss of chromosome 14q are associated with worse clinical outcome

Next, I investigated whether specific segmental CNVs are associated with proteomic subtypes. Therefore, segmental deletions and amplifications across chromosomal regions 1-22 were evaluated and quantified to identify recurrent CNVs in the context of proteomic subtypes (Figure 13). Consistent with earlier findings from the CNV burden analysis, AIDH_HG showed an increased genomic instability. More than 60% of the AIDH in AIHD_HG had losses on focal end of chromosomal arms 11p, 10q and 19q. In contrast, AIDH_LG and AIDH_INT showed minor genomic instability and shared most of the acquired segmental CNVs. Interestingly, comparison of AIDH_INT with AIDH_HG identified a subset of patients showing segmental losses within a locus of the chromosomal arm 14q, while AIDH_LG showed largely no alterations in this region. Approximately 25% of the cases from AIDH_INT and 35 % from AIHD_HG cases showed recurrent deletions within the genomic locus chr14:62025000-86950000.

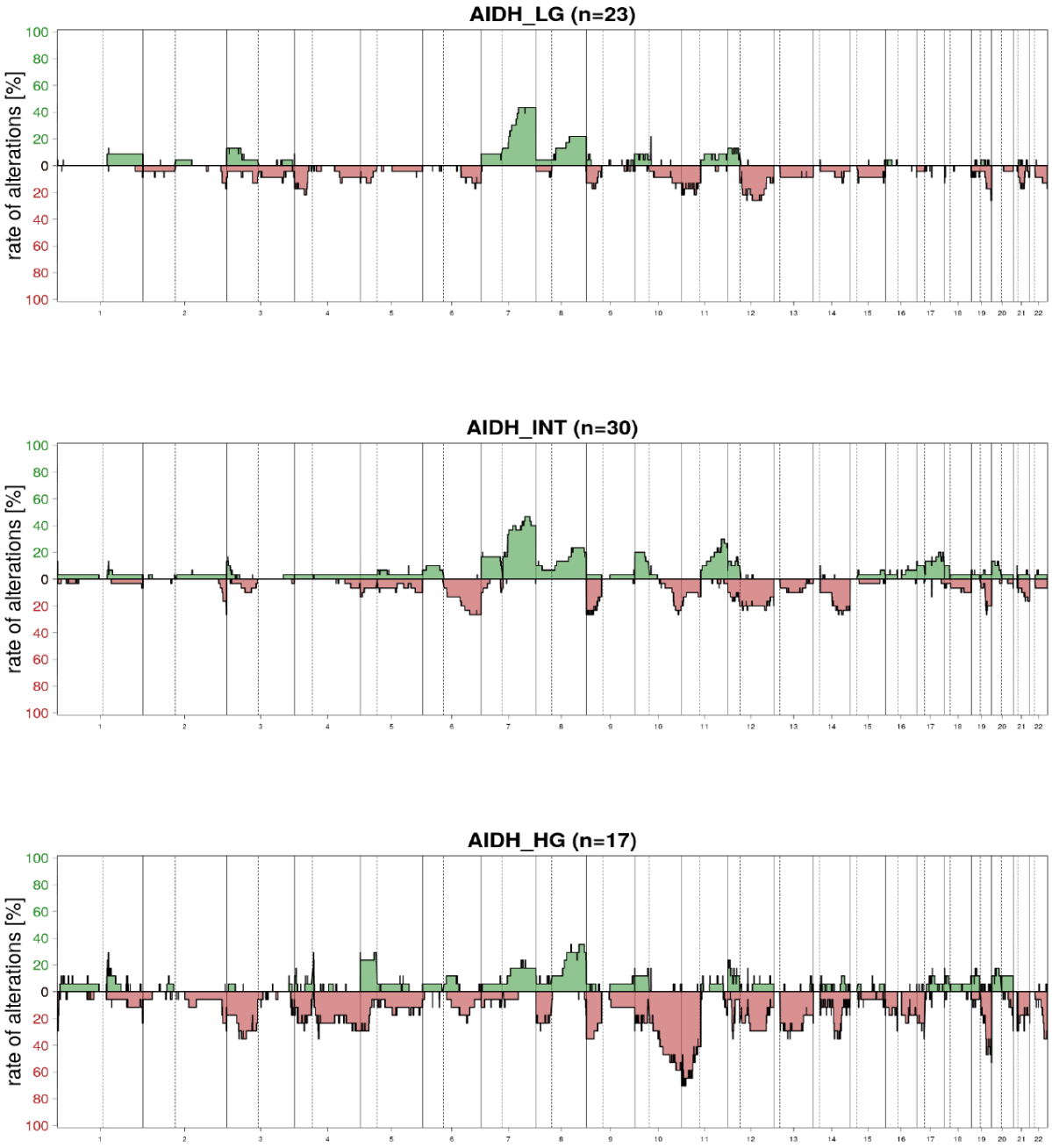


Figure 13: Analysis of frequent segmental copy number variation in context of proteomic subtypes indicates a partial loss in chromosome 14q in AIDH_INT and AIHD_HG.
Copy number variation (CNV) summary plots showing the rate of altered segments across the autosomes (1-22) in context of proteomic subtypes. Summary CNV plots were generated by segmenting the chromosomal arms into bins of 1% and calculating the relative frequency of gains and losses within each proteomic subtype followed by plotting the rate of alteration alongside their position in the genomic axis. The plots shows that AIDH_HG have high genomic instability compared to the other groups. Further it shows a common deletion within a region of 14q in AIDH_HG and AIDH_INT. Abbreviations: AIDH, Astrocytoma IDH mutant; AIDH_HG, Astrocytoma, IDH mutant high-grade; AIDH_INT, Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma, IDH mutant low-grade.

Subsequently I evaluated the CNV status of the altered chr14q loci (chr14:62025000-86950000) across all samples to determine the association with patient outcome. Cases with a CNV region score below -0.1 were considered to have a deletion. Comparison of the overall survival of AIDH patients demonstrated a significant worse outcome when the locus was deleted in the context of high- and low-grade AIDH (Figure 14). Thus, these results suggest, that a segmental loss of the 14q loci might be an additional potential marker for malignant AIDH. Furthermore, it is likely that genes affected by this loss may have a potential contribution for disease progression and clinical outcome.

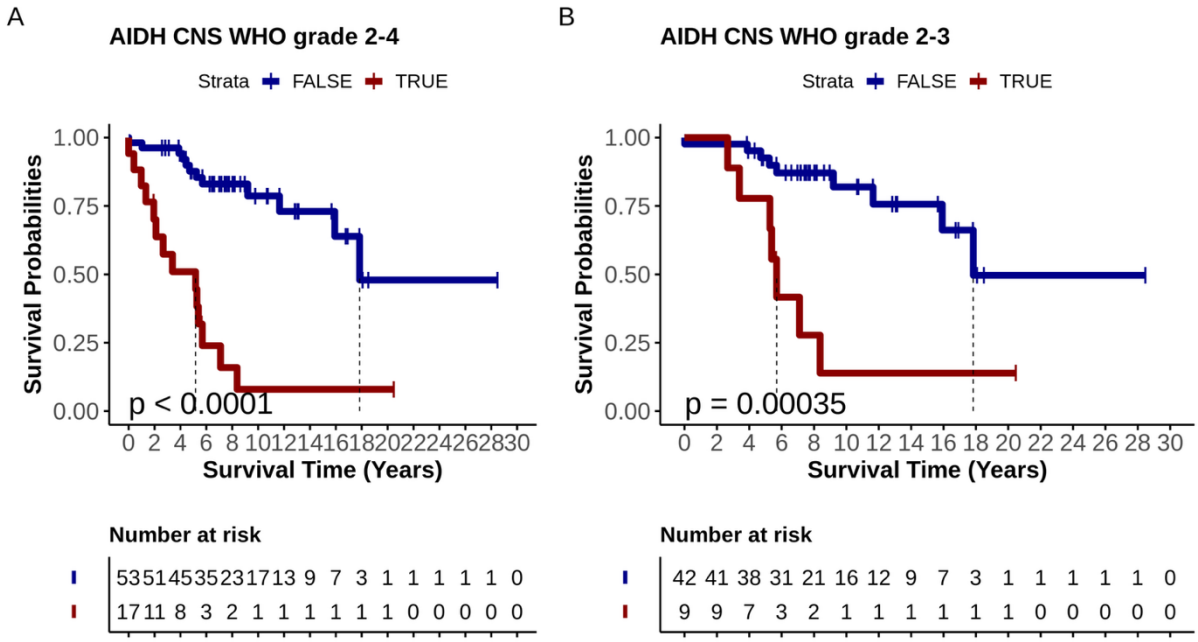


Figure 14: Stratification of patients by deletion in chr14:62025000-86950000 confers association with clinical outcome.
Kaplan-Meier plots showing the survival probabilities of AIDH patients stratified by occurrence of a deletion in chr14:62025000-86950000 in A) presence and B) absence of CNS5 WHO grade 4 AIDH tumours. The CNV score for the region was determined and a threshold of -0.1 was set to stratify cases into balanced and deleted groups. The plots show significant worse clinical outcome in cases with an deletion in the analysed chromosomal region. Abbreviations: AIDH_INT, Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma, IDH mutant low-grade; OS, Overall survival.

3.1.3.4. Prognostic differences of proteomic subtypes are independent of malignancy associated gene-level SCNA

To evaluate whether gene-level somatic copy number alterations (SCNA) were associated with the proteomic subtypes, I examined alterations frequently described to be associated with malignant progression of AIHD. In this regard, alterations of the cell-cycle regulators CDKN2A/B, RB1, CCND2, CDK4 and CDK6 as well as the oncogenes MYC, MYCN, MET and PDGFRA were assessed. The distribution of SCNA in context of the proteomic subtypes is summarized in Figure 15. Unexpectedly, the analysis showed that almost none of the examined SCNAs displayed a subtype-specific enrichment. SCNA were observed across all three proteomic subtypes and frequencies were mostly comparable between AIDH_LG, AIHD_INT and AIDH_HG.

CDKN2A/B alterations were found in up to 31% of AIDH cases (Figure 15). Homozygous CDKN2A/B deletions were associated with both subtypes, AIDH_HG and AIDH_INT, suggesting that this marker might potentially be predictive for both subtypes. PDGFRA gains and amplifications were found in the context of AIHD_HG, suggesting a strong association with a dismal outcome. Notably, 40% (29/71) of AIDH cases showed an alteration of MYCN, with 32% exhibiting hemizygous deletions. MYCN gains were most frequently seen in 50% of AIHD_LG (12/24) and 30% of AIDH_INT (9/30). Furthermore, 21% of AIDH cases showed a MYC gain, which was mostly enriched (35%) in the context of AIHD_HG. Nevertheless, these findings indicate that the investigated SCNA are potentially independent from the identified proteomic subtypes.

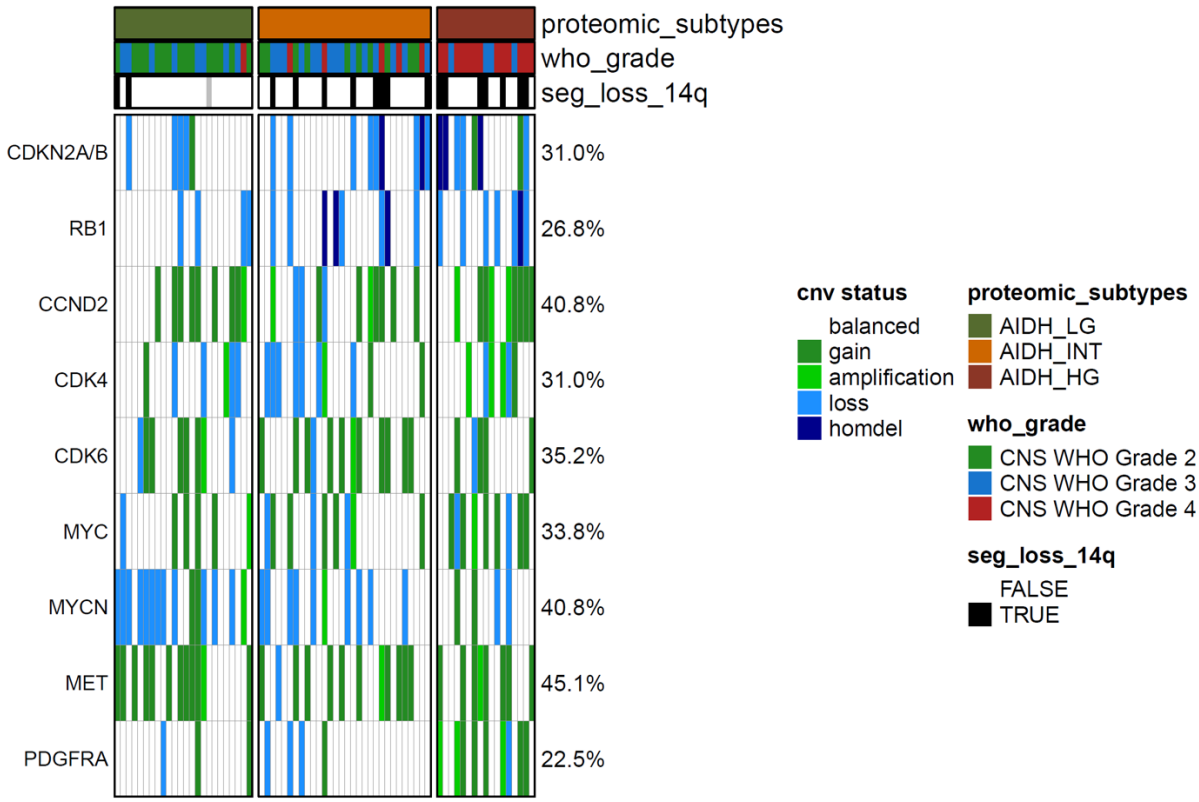


Figure 15: Enrichment of gene level somatic copy number alteration across proteomic subtypes (n=71). Oncoprint showing the gene level somatic copy number alteration (SCNA) for the genes CDKN2A/B, RB1, CCND2, CDK4 and CDK6 as well as the oncogenes MYC, MYCN, MET and PDGFRA in context of proteomic subtypes, CNS5 WHO grading. Colours in the oncoprint indicate copy number statuses balanced, gain, amplification, loss and homozygous deletion. Colours in the oncoprint header indicate proteomic subtypes and CNS5 WHO grade. SCNA status was assigned according to artificial threshold defined by *Shirahata et al* [13]. In this context, a homozygous deletion was considered when the CNV score was below -.415 and an amplification over 0.35. In addition, values below -0.1 were considered as loss and above 0.1 as gain. Abbreviations: AIDH, Astrocytoma IDH mutant; AIDH_HG, Astrocytoma IDH mutant high-grade; AIDH_INT Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma IDH mutant low-grade.

3.1.3.5. Prognostic differences between AIDH_LG and AIDH_INT are independent from epigenetic profiles

Subsequently, I examined whether proteomic subtypes exhibit distinct global methylation levels and if they can be distinguished by their CpG methylation profiles. Therefore, the matching methylation data of 71 samples was investigated in an unsupervised manner using UMAP to assess the relation of proteomic subtypes in the context of their epigenetic profiles. Furthermore, the global methylation status and methylation class according to the Heidelberg Methylation Classifier were determined to characterize the CpG global methylation status.

The analysis showed that epigenetic differences primarily arise in the context of high-grade cases in AIDH_HG while two other subtypes being less distinguishable. The UMAP in Figure 16 A depicted two groups of AIDH with each either showing enrichment of the methylation classes A_IDH_LG or A_IDH_HG. Evaluation of the distribution of proteomic subtypes in the two groups showed that AIDH_LG and AIDH_INT were strongly associated with the class A_IDH_LG (Figure 16 B). Furthermore, AIDH_HG were substantially enriched in the smaller group in the UMAP and strongly association with the methylation class A_IDH_HG. AIDH_LG and AIDH_INT had comparable high global methylation levels, which is consistent with the hypermethylation phenotype of A_IDH_LG (Figure 16 C). As expected, AIDH_HG were linked with a significant decrease in global CpG levels which is coherent with the hypomethylation phenotype of the class A_IDH_HG. Moreover, the findings show that AIDH_INT and AIDH_LG are highly similar in their epigenetic profiles as indicated by heterogeneous distribution in the UMAP as well as by absence of any significantly different methylated CpGs (Figure 16 D). Thus, these results show that the prognostic differences between AIDH_LG and AIDH_INT are independent from changes from the epigenetic landscape and primarily only indicative in the context of high-grade AIDH.

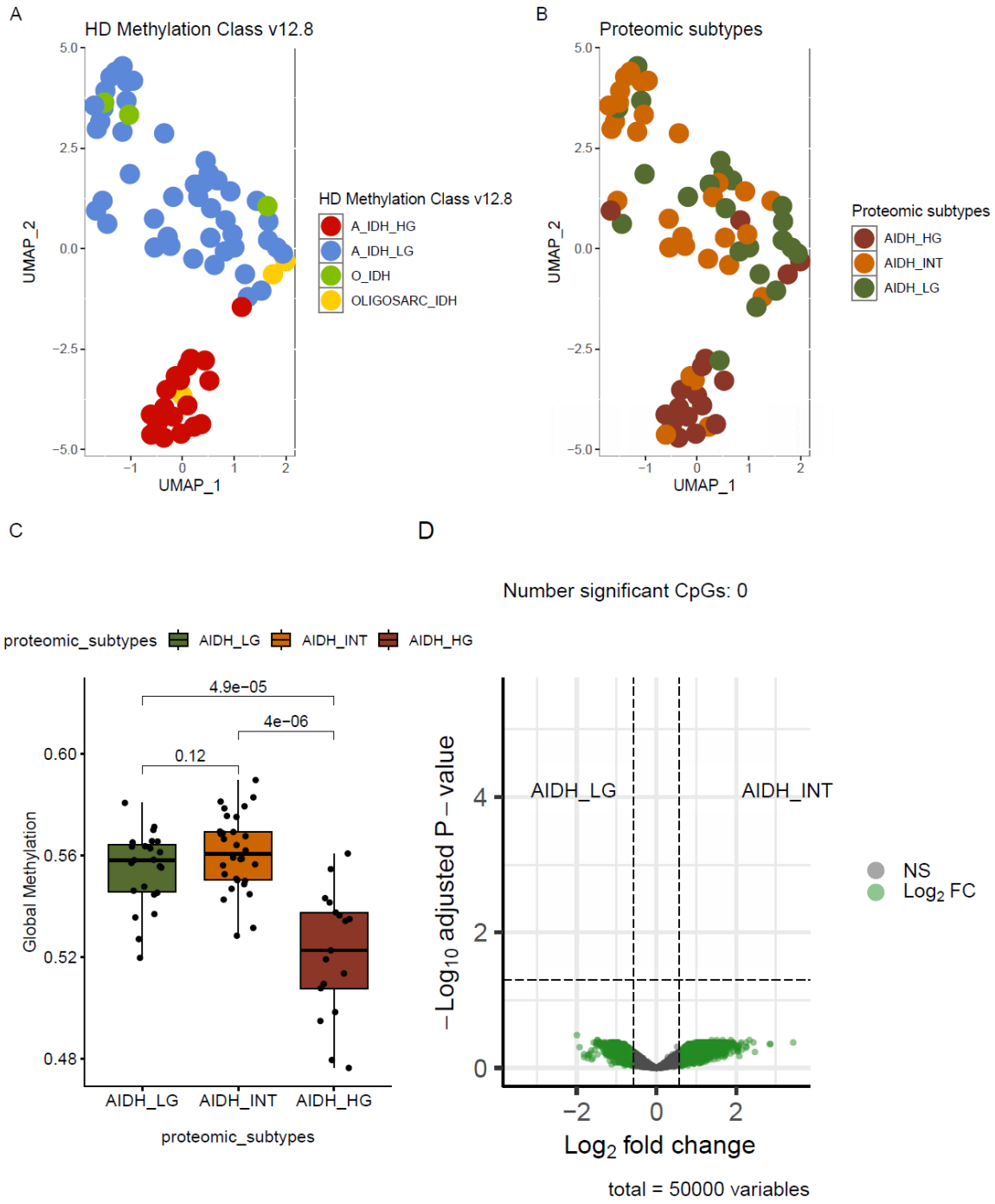


Figure 16: Prognostic differences between AIDH_LG and AIDH_INT are independent from epigenetic landscape.
Plots showing the main findings from interrogation of the epigenetic landscape in the context of proteomic subtypes. UMAP plots showing AIDH cases as dots with colours assigned to A) Heidelberg Methylation version 12.8, B) matched proteomic subtype. UMAPs were calculated using the initial 9 PCs preselected by the elbow approach and k=15 neighbours. The boxplots in C) show a comparison between global methylation levels and proteomic subtypes. The Wilcoxon rank test was used to assess the significance of hypermethylation differences, with normal distribution evaluated by the Shapiro Wilks approach. D) Volcano plot showing the Log_2 fold changes and $-log_{10}$ adjusted P-values from the differential methylated position analysis between AIDH_INT and AIDH_LG as dots. CpGs which are above the defined absolute Log_2 fold threshold of 0.58 are coloured in green, respectively. The plot shows that no significant CpGs could be found between AIDH_INT and AIDH_LG. The plots show that only AIDH_HG are distinguishable by their epigenetic pattern, while AIDH_LG and AIDH_INT are not distinguishable. Abbreviations: AIDH, Astrocytoma IDH mutant; AIDH_HG, Astrocytoma, IDH mutant high-grade; AIDH_INT, Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma, IDH mutant low-grade, PreditBrain_v12.8, Heidelberg Methylation Class version 12.8, PC Principal Component.

3.1.3.6. Proteomic subtypes are linked with unique non-immune and immune tumour microenvironments

The distribution of cell types in each sample was investigated to determine if changes in the tumour microenvironment (TME) are linked to distinct clinical outcomes of proteomic subtypes. Cell types proportions were determined by deconvolution of the methylation data using MethylCIBERSORT (Figure 17).

The analysis revealed several significant shifts in immune and non-immune cell types in relation to the proteomic subtypes. In terms of non-immune cells, AIDH_INT showed a substantial increase in the proportion of cancer cells compared to the other subtypes, suggesting a higher density of tumour cells in this tissue. In contrast, in terms of immune cells, a higher fraction of CD4 effector cells was identified in AIDH_LG while there is a progressive decrease in AIDH_INT and AIDH_HG suggests higher immune surveillance in AIDH_LG. In contrast, AIDH_HG cases showed a higher fraction of CD8 positive immune cells often described in the context of inactivation of cancer cells [187], indicating an TME characterised by cytotoxic immune cells as possible response to tumour burden. Interestingly, cases in AIDH_HG and AIDH_LG showed a significantly higher distribution of monocytic immune cells compared to AIHD_INT which suggests a potential immune remodelling in these subtypes. Thus, these shifts collectively show that proteomic subtypes associate with unique non-immune and immune TME.

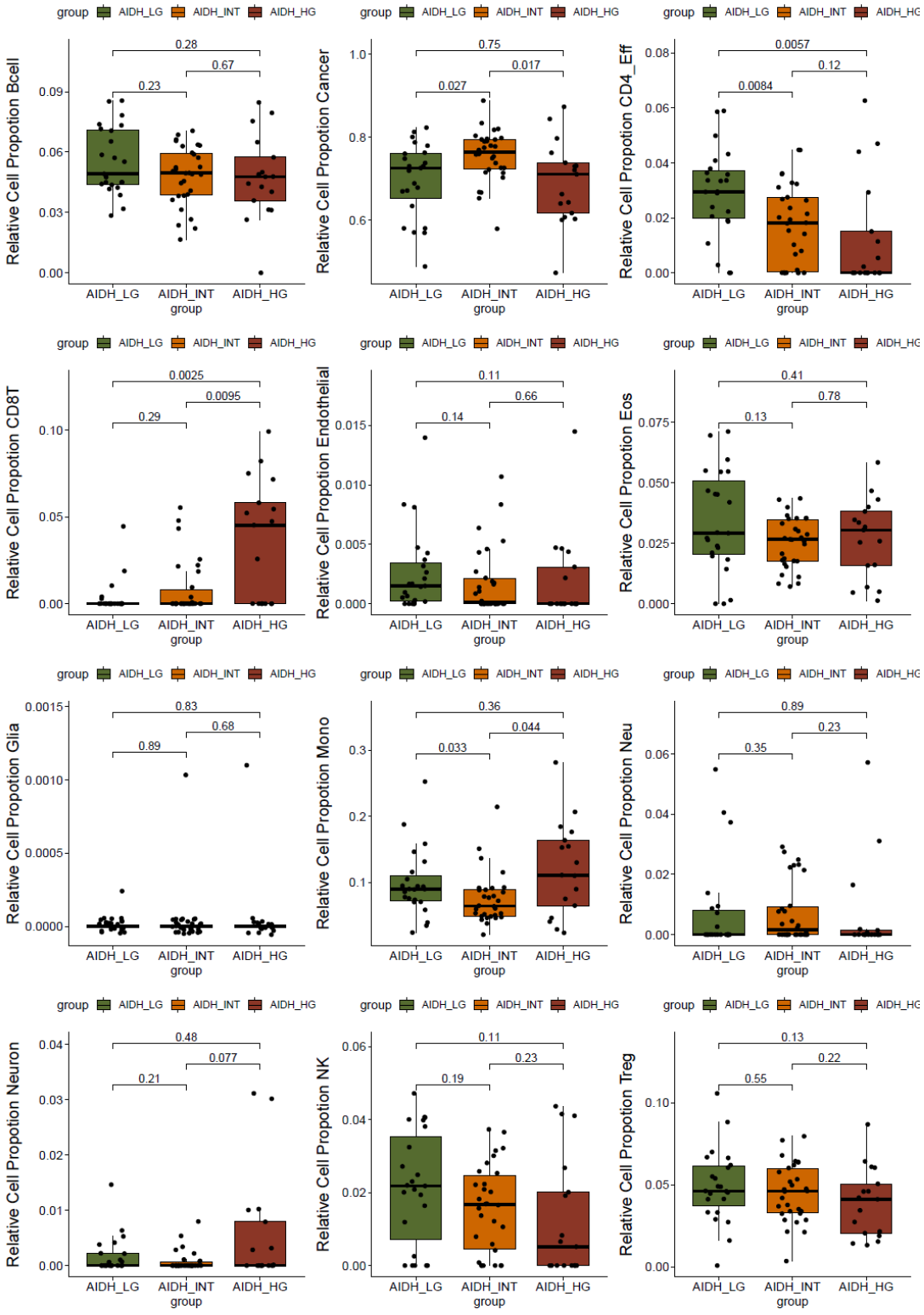


Figure 17: Cell deconvolution reveals significant shifts of immune and non-immune cell types in the context of proteomic subtypes.

Box plots showing cell type proportions in relation to proteomic subtypes estimated from methylation data. Cell type proportions were determined per sample by submitting beta values to MethylCIBERSORT. Statistical significance was determined using a Wilcoxon rank sum test. The plots show a higher proportion of cancer cells is linked to AIDH_INT and a gradual change in immune cell types, CD4 Effector and CD8-T cell in, in the context of proteomic subtypes. Abbreviations: AIDH, Astrocytoma IDH mutant; AIDH_HG, Astrocytoma IDH mutant high-grade; AIDH_INT, Astrocytoma IDH mutant intermediate; AIDH_LG, Astrocytoma IDH mutant low-grade; NK Natural-Killer cell; Treg T-regulatory cell; CD4_eff, CD4 Effector cell; CD8T, CD 8 Cytotoxic lymphocyte; Eos, Eosinophile cell; Neu Neutrophile cell.

3.2. Characterization of proteins linked with AIDH malignancy

Previous findings showed a link between proteomic subtypes and clinical outcomes, suggesting potential biological differences with prognostic significance. Therefore, the aim was to characterise protein expression in the context of proteomic subtypes to understand their biological significance. Proteins related to the AIDH subtypes and their associated biological processes were investigated in detail. To enhance the biological insight in the molecular characterization of the proteomic subtypes, the previous preprocessing of the proteomic data was repeated by using a filtering strategy accounting for missingness arising from differences between biological groups. Proteins with missing values were retained only if they had less than 80% missingness in at least one subtype. For those proteins, a mixed imputation strategy was used to substitute the missing values yielding a total of 6853 proteins. Afterwards the protein intensities were log2 median normalised and batch corrected for further downstream analysis.

3.2.1 Proteomic subtypes are characterised by divergent biological expression programmes

To investigate whether proteomic subtypes are biologically distinct, I characterized substantial differences in protein expression patterns and their potential biological implications. A differential expression analysis (DEA) was performed to identify proteins specifically expressed in each proteomic subtype by comparing the average expression of each subtype against the averaged expression of the remaining subtypes. Proteins were considered differentially expressed if they had an adjusted P-value below 0.05 and an absolute $\log_2$-fold change (LogFC) bigger than 0.58 (Supplemental Tables 1-3, Supplemental Figure 2). To infer biological processes upregulated in each subtype, a Gene set enrichment analysis was conducted using gene sets form the Gene Ontology (GO) biological process database. Subsequently a semantic similarity of enriched pathways with an adjusted p-value below 0.05 was conducted to characterize the major underlying biological phenotype for each proteomic subtype.

In total, 2788 significantly differentially expressed proteins (DEP) were identified. Although DEA was designed to identify subtype specific protein expression, a subset of DEP from AIDH_LG as well as AIDH_HG common with AIDH_INT suggesting partial overlap in expression programs (Supplemental Figure 3). By interrogating subtype specific DEP, I discovered that neuronal protein expression was enriched in AIDH_LG, metabolic expression in AIDH_INT and mesenchymal expression in AIHD_HG (Figure 18 A)

The neuronal expression signature of AIDH_LG was characterized by proteins enriched in processes such as synaptic signalling as well as the transport and uptake of neurotransmitters (Figure 18 B). Consistent with this pattern, AIDH_LG exhibited strong expression of proteins contributing to canonical neuronal function including CAMK2A, SYN1, SVB2, TUBB4B, STX1A, SNCB and SLC30A3 but also blood barrier proteins PRRT2, CLDN11 and the brain metabolism protein CRYM. These findings suggest that AIDH_LG cases either retain a more neuronal differentiated phenotype or includes more neurons relative to the other subtypes.

AIDH_INT showed substantial enrichment of proteins contributing to mitochondrial (MT) metabolism including amino acid metabolism but also showed enrichment for MT biogenesis processes including mitochondrial gene expression and translation (Figure 18 C). Furthermore, top expressed proteins in AIDH_INT included metabolic proteins such as PCCA, PCCB, OXNAD1, ALDH5A1, HIBADH. Notably, AIDH_INT also showed high expression of mitochondrial stress regulator proteins SIRT3 and TRAP1, which have been implicated to contribute to metabolic reprogramming under metabolic stress promoting neoplastic progression [188]. Thus, these findings indicate that increased malignancy of AIDH_INT might derive from an alteration of the MT metabolism.

The mesenchymal signature of AIDH_HG was characterised by a strong enrichment of proteins involved in extracellular matrix organisation, wound healing, and inflammatory response (Figure 18 D). These processes are typical hallmarks of epithelial-to-mesenchymal transition (EMT), which has been described in other cancer types [189]. In association with this results, AIDH_HG exhibited strong expression of proteins contributing or associated to EMT including transcription factors IGF2BP2 [190] , IGF2BP3 [191], PBK [192], the wound healing protein F13A1, cell migration regulators (NRP-1, LAMB1) as well as NNMT a known regulator of cellular methylation levels [193]. these patterns suggest that that the high-grade phenotype of AIHD_HG is linked to a mesenchymal phenotype.

Thus, these results demonstrate that proteomic subtypes are not only clinically but also biologically functional divergent providing strong support for their relevance and validity.

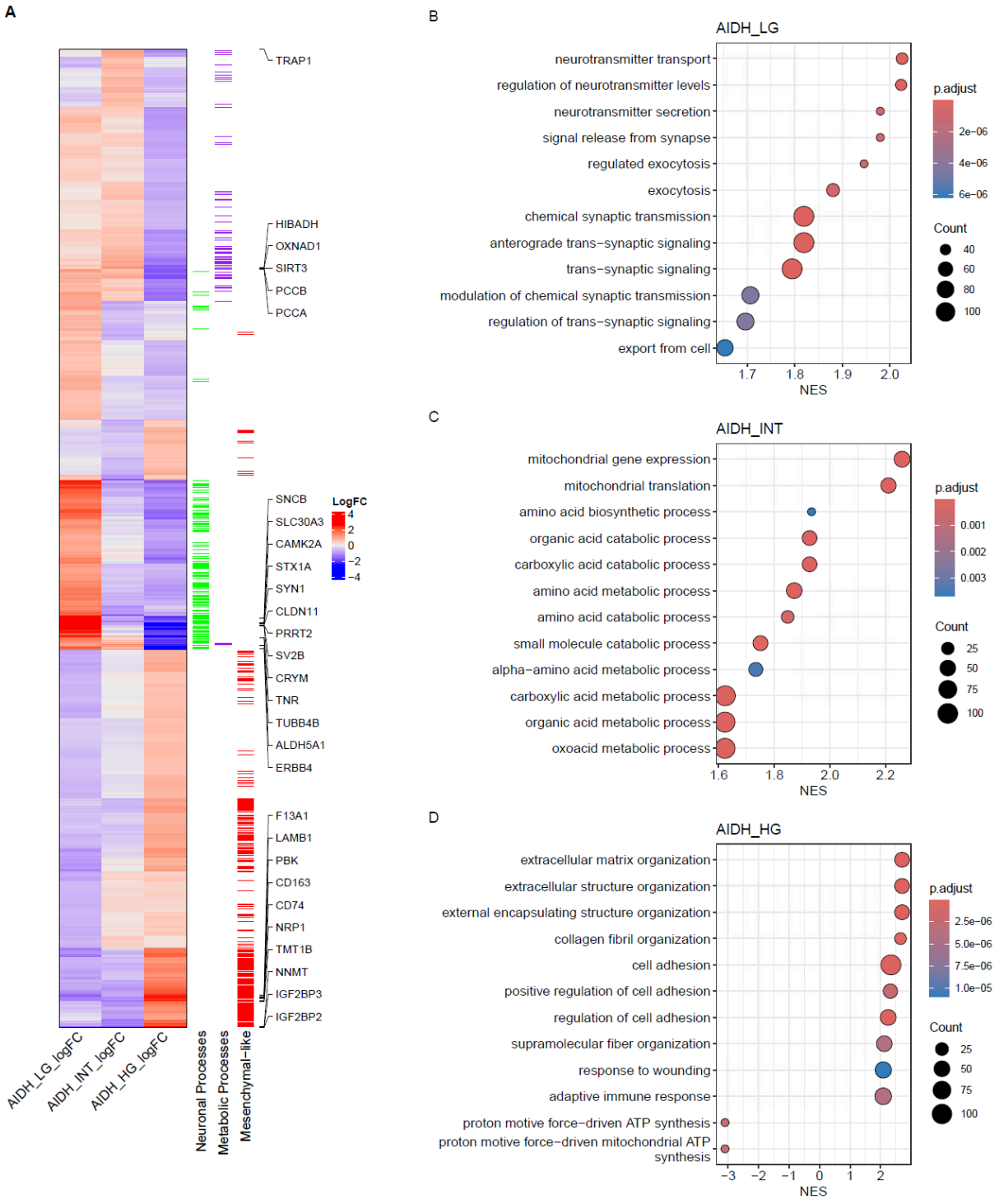


Figure 18: Proteomic subtypes show biological distinct phenotypes linked with metabolic, neuronal and mesenchymal expression programs.
Plots showing the key findings from the comparative analysis of proteomic subtypes. A) Heatmap representation showing the logFC of 2788 DEP. The top 10 proteins with the highest positive logFC of each group are highlighted. Gene set enrichment analysis combined with semantic similarity was performed to identify underlying major biological phenotypes. DEP contributing to the phenotypes are colourized in the row annotation of the heatmap with green, purple and red, respectively. Dot plots depicting the NES of the 12 strongest enriched biological processes in AIDH_LG, AIDH_INT, and AIDH_HG. AIDH_LG shows high enrichment of neuronal processes, AIDH_INT shows metabolic processes, and AIDH_HG shows processes linked with a mesenchymal phenotype. Abbreviations: AIDH, Astrocytoma IDH mutant; AIDH_HG, Astrocytoma, IDH mutant high-grade; AIDH_INT, Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma, IDH mutant low-grade; DEP Differential expressed protein; logFC $\log_2$-fold change; NES normalized enrichment score.

3.2.2. In-depth analysis of AIHD_INT reveals upregulation of mitochondrial biogenesis in concordance with complex I dysregulation

Given the recent findings from proteomic subtype analysis, it was questioned how the altered metabolic protein expression in AIDH_INT might contribute to malignant transformation. Therefore, I performed an analysis of the mitochondrial and metabolic proteins and processes altered in AIDH_INT to identify the candidates and potential drivers for malignancy. The enrichment of subtype specific DEP in MT processes was analysed by using GSEA on the gene sets provided in the MitoDB3.0 database and filtering for gene sets with an adjusted p-value below 0.05.

The enrichment results showed a striking upregulation of MT proteins in AIDH_INT compared to the other subtypes (Figure 19 A). In this regard, AIDH_INT exhibited enrichment in alternative metabolic pathways (amino acid metabolism, lipid metabolism) but also MT apoptosis, autophagy and mitophagy. Interestingly, the strongest enrichment of proteins was found in the context Mitochondrial central dogma which includes the machinery involved for transcription and translation of MT encoded genes (Figure 19 B) suggesting an increased mitochondrial biogenesis in AIDH_INT [194]. Nevertheless, evaluation of complexes of the electron transport chain (ETC) showed a downregulation of proteins from Complex I while the expression of proteins from the other complexes III-V was still retained.

Interestingly, investigation of the ETC Complex I proteins revealed several significant shifts in the expression of complex I subunits (Figure 19 C). Striking downregulation were discovered in the context of core and accessory subunits including nuclear (NDUF) and MT-DNA (MT-ND 1-6) encoded subunits. Conversely, regulatory and assembly proteins of Complex I including AIFM1, NUBPL and NDUFAF1 were strongly upregulated suggesting a compensatory response as answer to dysregulation core and accessory subunit. Taken together, these findings characterize an unique inverse association between dysregulation of complex I units and the upregulation of mitochondrial protein biogenesis with malignancy of AIDH_INT.

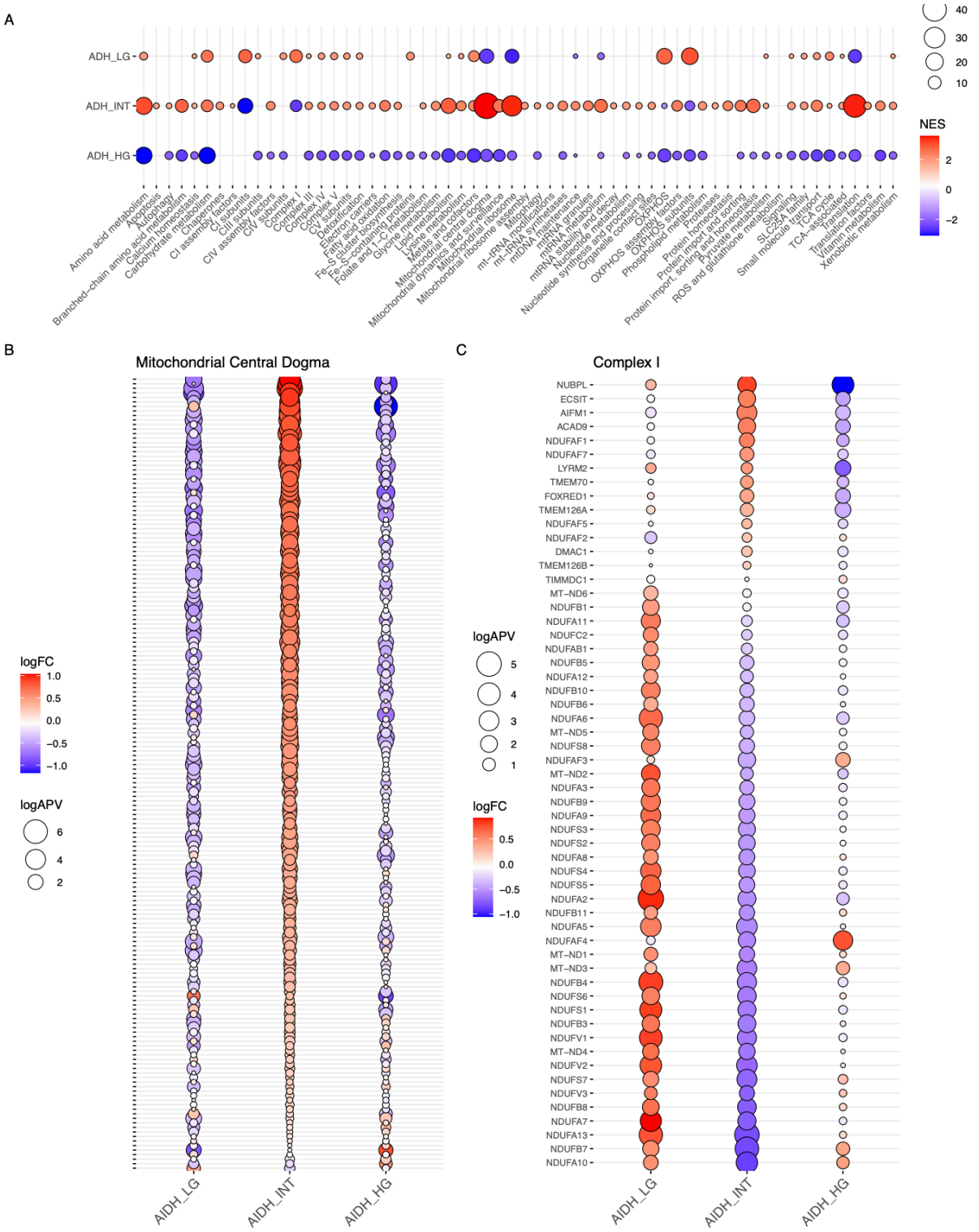


Figure 19: AIDH_INT shows strong upregulation of mitochondrial processes in concordance with complex I dysregulation.
Key findings from the interrogation of proteins contributing to metabolic and MT processes in context of the proteomic subtypes. Ballon plots showing the A) normalised enrichment scores (NES) and log adjusted P-value (logAPV) of MT and metabolic processes in proteomic subtypes and log_2-fold changes (logFC) and logAPV of proteins of the central MT dogma and ETC complex I. Enriched pathways were obtained by analysing gene sets from the MitoDB3.0 database and results from subtype-specific differential expression analysis (DEA) in a gene set enrichment analysis (GSEA). Abbreviations: AIDH, Astrocytoma IDH mutant; AIDH_HG, Astrocytoma, IDH mutant high-grade; AIDH_INT, Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma, IDH mutant low-grade; NES normalized enrichment score; MT Mitochondria; logFC, log_2-fold changes.

3.2.3. AIDH_INT and AIDH_HG show enriched expression of MYC target proteins

The recent results showed that an increase of malignancy in AIDH_INT associates with alteration of their mitochondrial metabolism. Importantly, dysregulation of complex I has been described to promote tumour progression through AKT1 activation [195] an important player for a series of oncogenic pathways including MYC [196]. I therefore examined whether AIDH_INT with complex I dysregulation also show an activation of MYC by evaluating the enrichment of MYC target proteins in the context of proteomic subtypes. MYC hallmark gene sets (MYC_Targets_V1, MYC_Targets_V2) were curated from MSigDB and their enrichment was determined using GSEA. Notably, MYC target proteins were found significantly enriched in both subtypes AIDH_INT and AIDH_HG compared to AIDH_LG (Figure 20 A). In this regard, evaluation of the protein expression pattern of MYC targets from the MYC genesets V1 showed overall consistent upregulation of MYC target proteins in AIDH_INT. To summarize, the results show and enrichment of MYC target proteins in AIDH_INT and AIDH_HG suggesting a potential activation of MYC.

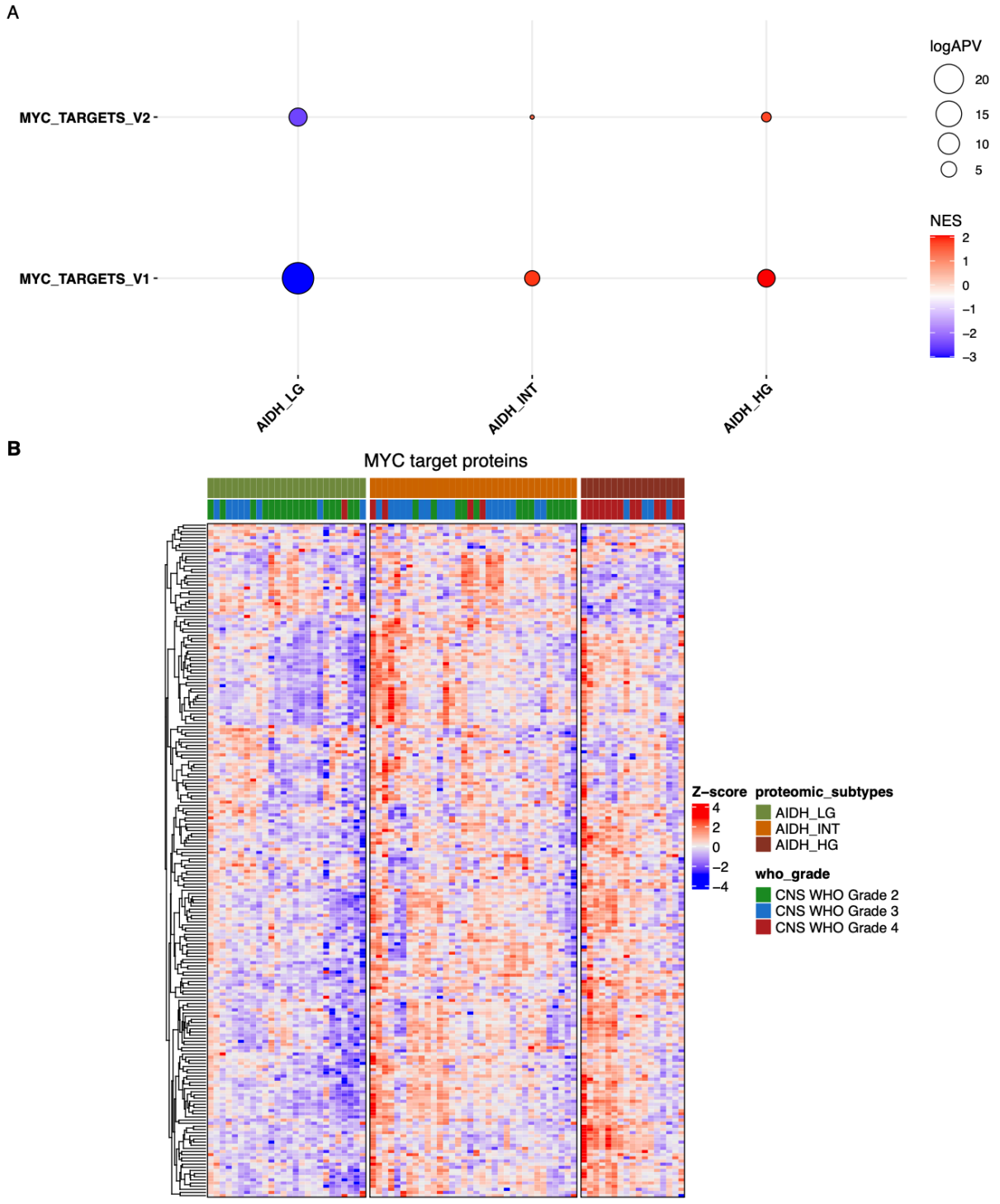


Figure 20: AIDH_INT and AIDH_HG show enriched expression of MYC target proteins.
Key findings from the investigation of MYC activity in proteomic subtypes. A) Ballon plots showing the enrichment of proteins from MYC target signatures V1 and V2 in context of the proteomic subtypes. MYC genesets were acquired from the MSigDB database. B) Heatmap showing the Z-scored protein intensities from MYC target proteins from the MYC V1 genesets. The plots show that both AIDH_INT and AIDH_HG show an overall higher expression of MYC target proteins compared to AIHD_LG. Abbreviations: AIDH, Astrocytoma IDH mutant; AIDH_HG, Astrocytoma IDH mutant high-grade; AIDH_INT Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma IDH mutant low-grade

3.3. Evaluation of a proteomic prediction framework for biological validation and clinical stratification of AIDH

The recent findings presented a stratification of AIDH into three clinically and biologically distinct proteomic subtypes with favourable outcome being characterized by neuronal protein expression, intermediate with metabolic and dismal outcome by mesenchymal protein expression. Therefore, a predictive framework was developed using the previous information and the utilization for prognostic prediction of AIDH patient outcomes was tested by applying it for validation of recent biological and clinical findings.

3.3.1. A protein signature score-based decision tree model shows effective prediction of proteomic subtypes

To test if the identified proteomic subtypes can be reliably identified based on their protein expression hallmarks, I trained and tested a framework for the prediction of proteomic subtypes. Therefore, proteomic subtype-specific proteins (see 3.2.1) were summarized into signatures scores as it has been shown to ensure better accuracy for prognostic stratification in a single sample manner [134, 135]. The protein signature scores were tested for suitability in subtype prediction by examining their ability to separate proteomic subtypes by hierarchical clustering. This approach tested whether a simple clustering method could correctly group samples into their corresponding subtypes.

The results demonstrate that protein signature scores effectively separate the proteomic subtypes. Samples showed strong separated into three major branches, with depicting almost homogenous contributions for each subtype. Notably, only five cases were clustered into the wrong subtype cluster (Figure 21).

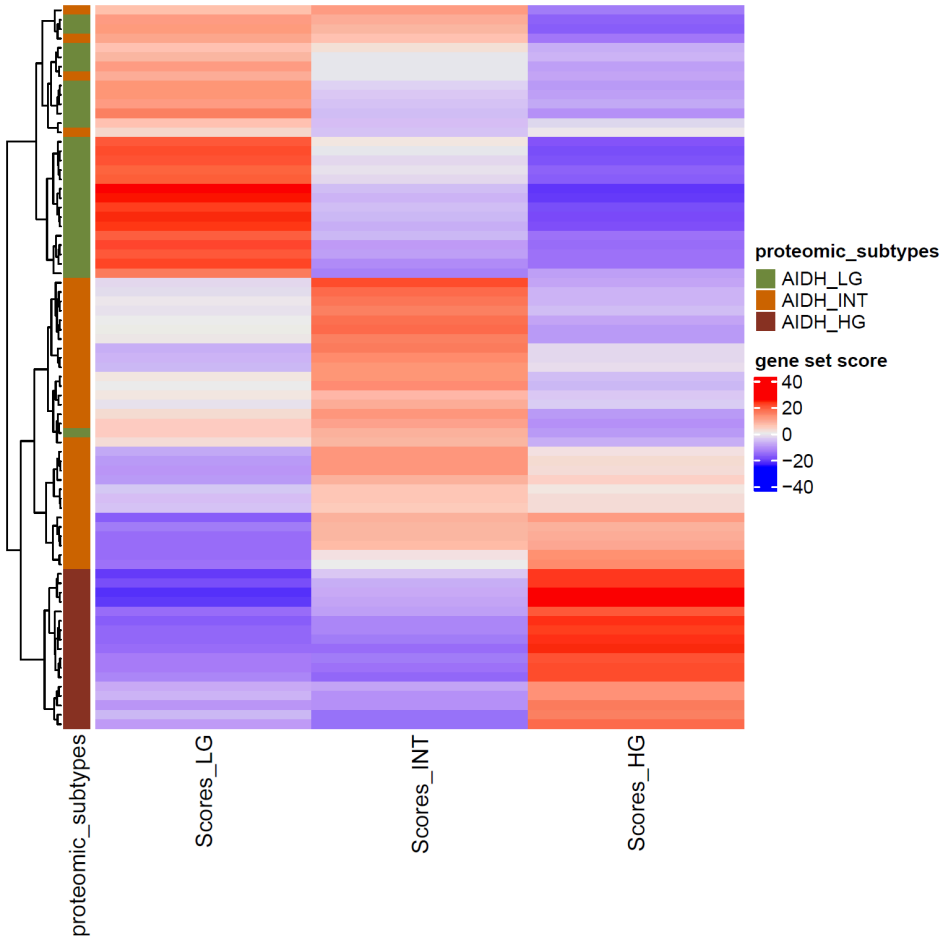


Figure 21: Protein signature scores effectively distinguish between proteomic subtypes.
Heatmap showing the hierarchical clustering of AIDH samples based on the subtype specific protein signature scores. Significantly upregulated proteins from each subtype (3.2.1) were and used summarized into single sample scores using the genes set scoring algorithm from the R-package viper. The heatmap depicts that the three major groups have almost homogenous distribution of the three proteomic subtypes. Abbreviations: AIDH, Astrocytoma IDH mutant; AIDH_HG, Astrocytoma IDH mutant high-grade; AIDH_INT, Astrocytoma IDH mutant intermediate; AIDH_LG, Astrocytoma IDH mutant low-grade

To further evaluate whether AIDH patients can be stratified by their proteomic signature scores, a decision tree model was trained on a subset of 70% and prediction accuracy was tested on 30% of the remaining cohort. The decision tree model was optimized for model balanced accuracy (BACC) to account for imbalanced groups (Figure 22). Evaluated hyperparameters included the complexity parameters (cp=0.01- 0.05), maximum tree depth (maxdepth=1-5) and the minimum tree split (minsplit =2-10). Evaluation of the confusion matrices of the test data before tuning (Figure 22 A) already showed a good performance (BACC=0.886) and depicted that only minor wrong classifications were found. In comparison the confusion matrix of the tuned model (cp=0.01768; maxdepth = 5; minsplit = 3) indicates no misclassifications between the groups of the test data (BACC =1) (Figure 22 B). Thus, these results suggest that the tuned decision tree model could provide effective stratification by capturing relevant differences between the proteomic subtypes by their protein signature scores.

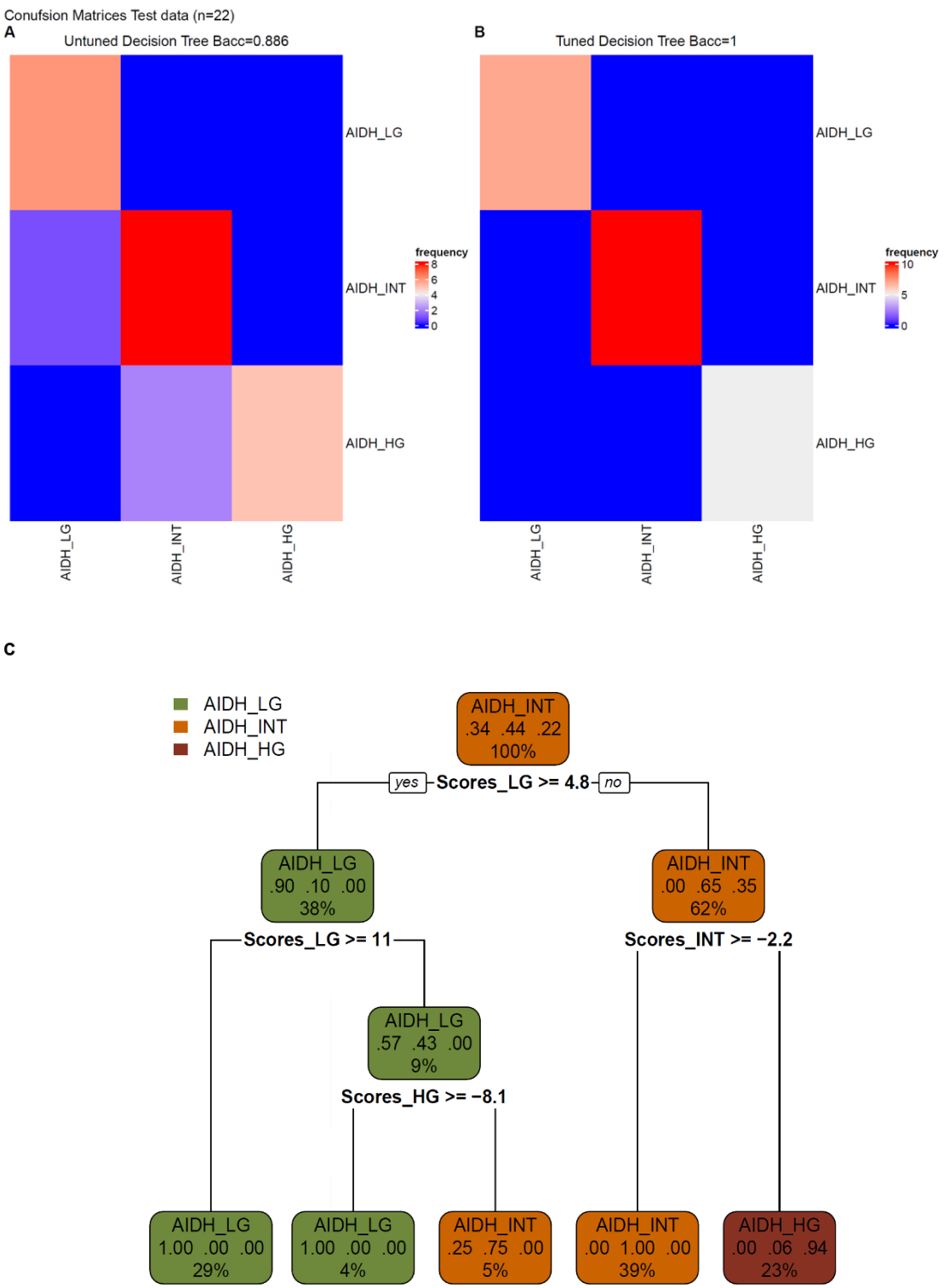


Figure 22: A Decision-Tree models enables prediction of proteomic subtypes from protein signature scores. Key findings form the training and testing of the decision-tree model for proteomic subtype prediction. Confusion Heatmaps showing the predicted proteomic subtypes of the test data (n=22) before A) and after B) optimizing the hyperparameters of the model for BACC. C) Decision tree model showing insight into the decision about proteomic subtypes based on protein signature scores. Abbreviations: AIDH, Astrocytoma IDH mutant; AIDH_HG, Astrocytoma, IDH mutant high-grade; AIDH_INT, Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma, IDH mutant low-grade; BACC Balanced Accuracy.

3.3.2. Validation of biological and clinical findings in in-house and publicly available data

Given the possibilities provided by the decision tree model, I evaluated whether proteomic subtypes predicted from AIDH cases from other proteomic studies align with recent key biological and clinical findings. Therefore, the decision tree model was tested on four datasets comprehending low- and high-grade AIDH. Two in-house DDA datasets were ascertained, the first dataset included the protein expression profiles 66 low-grade AIDH patients (26 WHO grade 2, 40 WHO grade 3) acquired form the GGN [139] and NOA trail (Inhouse GGN_NOA) [140] while the second included high- and low-grade AIDH tumours of 24 patients (9 WHO grade 2,11 WHO grade 3, 4 WHO grade 4) from our recent study investigating proteomic markers for AIDH and OIDH (Inhouse HIP1R_VIM)[130]. Furthermore, publicly available external data, from two recent proteomic studies were included, compromising DDA protein expression profiles which from 11 AIDH [131] (9 WHO grade 3,2 WHO grade 4) and 12 AIDH (8 WHO grade 3,4 WHO grade 4) [132].

3.3.2.1. Stratification of AIDH from in-house and external proteomic studies show consistency in association between proteomic subtypes and WHO grading

To get a first impression whether the decision tree model agrees with the biological findings, it was evaluated whether the composition of WHO grades in the subtypes compare with that of the DIA study cohort. The comparison of predicted proteomic subtypes with their WHO grades showed consistency with recent findings form the DIA study cohort (Figure 23). Here, it was shown that majority of CNS WHO grade 3 grouped into AIDH_INT but also distribution into AIDH_LG and AIDH_HG. In addition, most AIDH_LG show high abundance of CNS WHO grade 2 AIDH while most CNS WHO grade 4 were classified either as AIDH_INT or AIDH_HG.

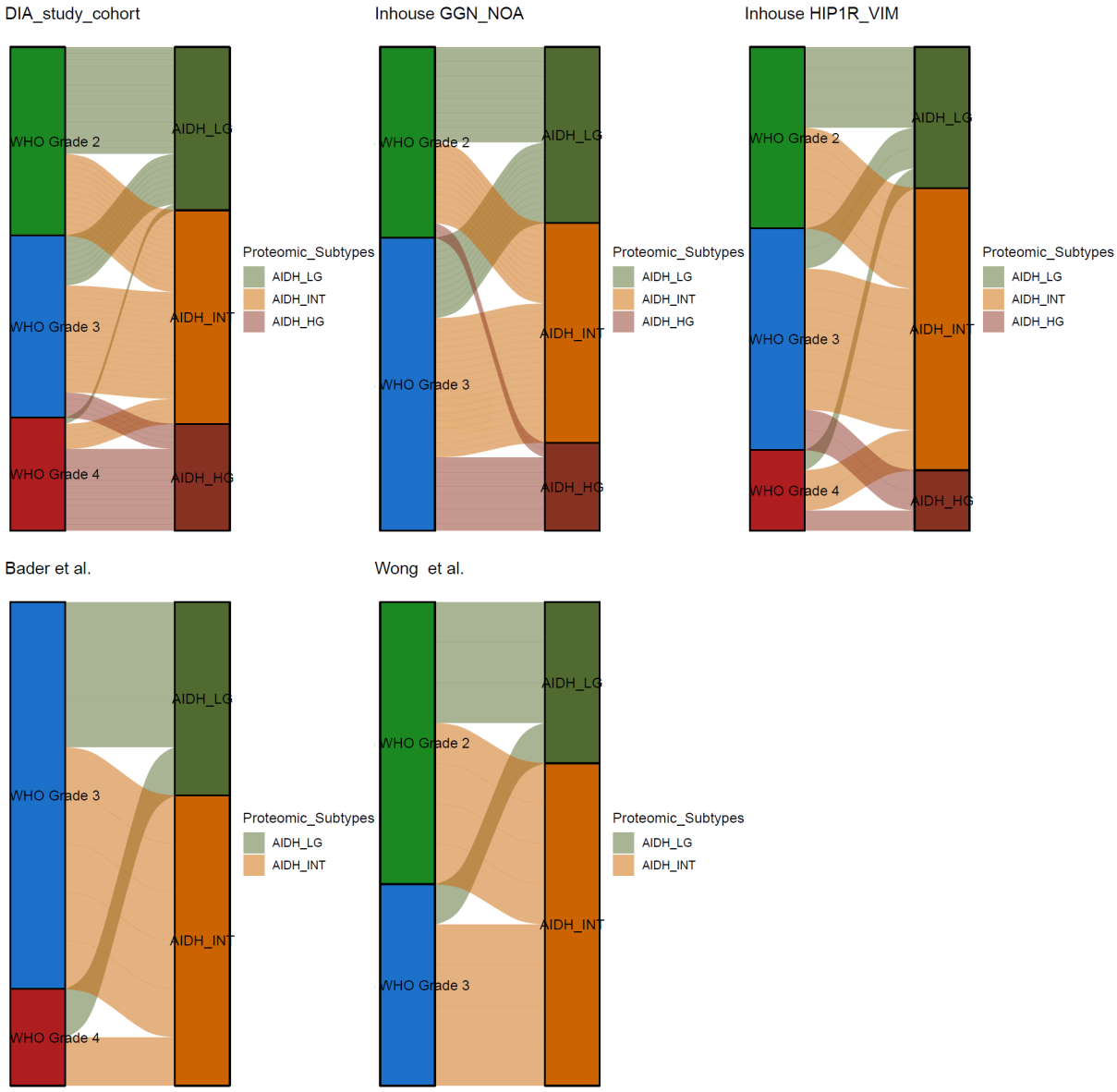


Figure 23: Relation of predicted proteomic subtypes with WHO grades from different datasets emphasizes with recent findings from the DIA study cohort.
Sankey plots showing the association between CNS WHO grades and the decision-tree predicted proteomic subtypes in two in-house DDA datasets (Inhouse GGN_NOA, Inhouse HIP1R_VIM) and two recent proteomic studies from Bader et al [131] and *Wong et al.* [132]. The results show consistency in the found association between WHO grades and Proteomic subtypes in the DIA study cohort by showing that AIDH_LG consists mostly of CNS WHO grade 2 AIDH while AIDH_INT more associates with CNS WHO 3. Abbreviations: AIDH, Astrocytoma IDH mutant; AIDH_HG, Astrocytoma, IDH mutant high-grade; AIDH_INT, Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma, IDH mutant low-grade; WHO, World Health Organization

3.3.2.2. Interrogation of in-house and external proteomic studies confirm complex I dysregulation in predicted AIDH_INT cases

To further validate AIDH stratification model and biological findings, I evaluated protein expression pattern between the predicted proteomic subtypes in regard of the mitochondrial alterations identified in AIDH_INT. A differential expression analysis was conducted as previously described (see 3.2.1) by comparing the expression of one subtype against the mean of the others. Interrogation of mitochondrial proteins from complex I and the mitochondrial central dogma validated the recent identified mitochondrial phenotype of AIDH_INT. The results showed consistency with recent results by showing strong expression for proteins linked with the mitochondrial central dogma in concordance with complex I downregulation (Figure 24). In addition, predicted AIDH_LG and AIDH_HG show consistency with recent findings by depicting lower expression of mitochondrial dogma proteins and AIDH_LG showing higher expression of complex I in all evaluated datasets. Thus, these findings provide validation for the unique mitochondrial expression pattern discovered in the context of AIHD_INT.

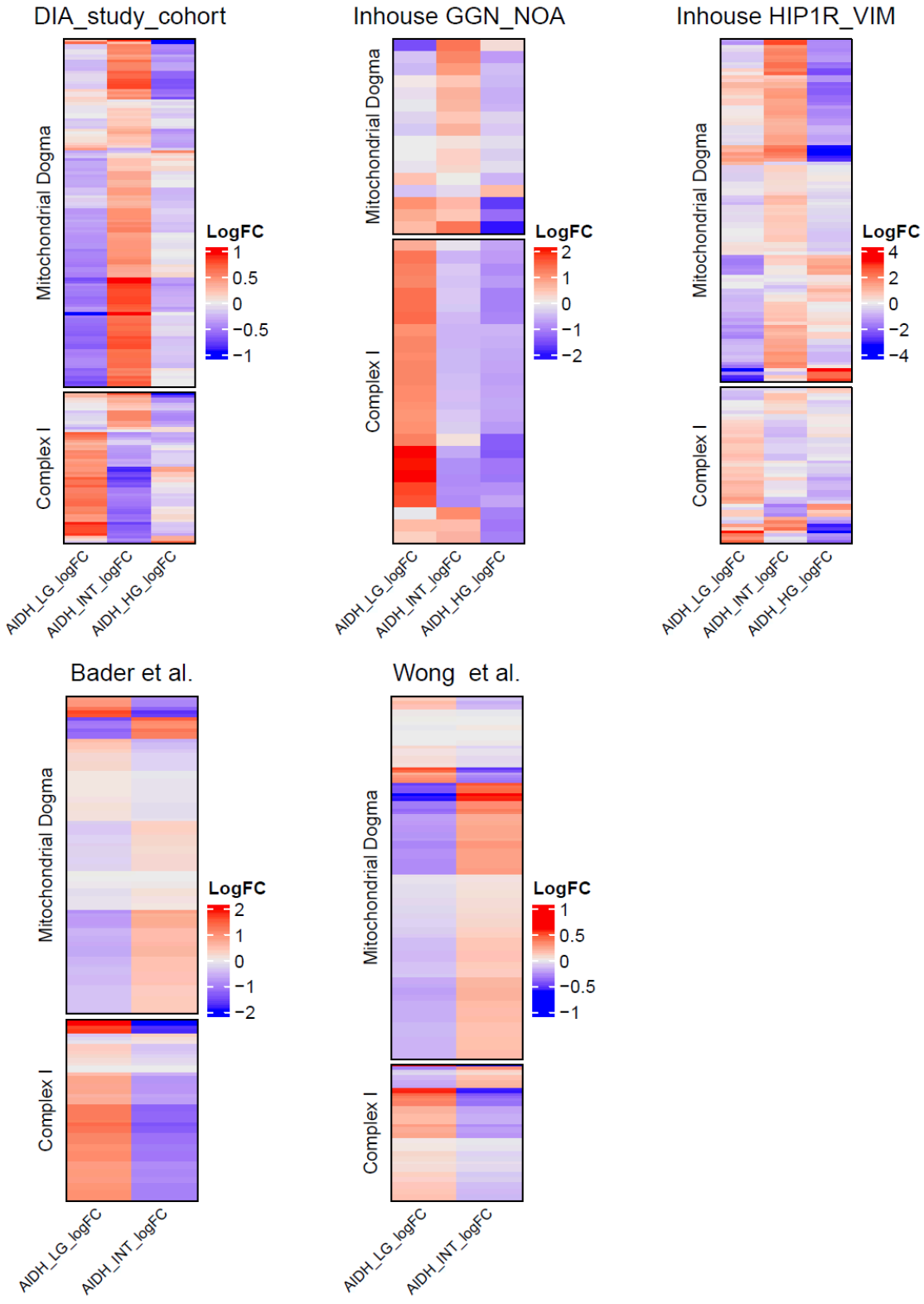


Figure 24: Comparison of protein expression across proteomic datasets validates mitochondrial phenotype observed in AIDH_INT.
Heatmap representations depicting the log_2-fold changes of mitochondrial proteins from complex I and the central mitochondrial dogma gene sets of the MitoCoP Database (version 3.0) between predicted proteomic subtypes from various in-house and publicly available proteomic datasets from *Bader et al* [131] and *Wong et al.* [132]. Abbreviations: AIDH, Astrocytoma IDH mutant; AIDH_HG, Astrocytoma, IDH mutant high-grade; AIDH_INT, Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma, IDH mutant low-grade; LogFC, Log_2-fold change

3.3.2.3. Upregulation of MYC proteins in AIDH_INT is consistently found across in-house and external studies

To further validate whether predicted AIDH_INT and AIDH_HG are consistent with biological findings, I determined the overall upregulation of MYC proteins for each evaluated dataset. The ssGSEA scores of the MYC hallmark gene set were each estimated for each sample and compared between the proteomic subtypes. The results showed consistency with recent findings from the DIA study cohort by demonstrating a significant higher score for MYC hallmark proteins in the groups AIDH_INT and AIDH_HG for most of the samples (Figure 25). Interestingly, AIDH_HG in the Inhouse GGN_NOA data showed an even higher score compared to AIDH_INT emphasizing with the suggested higher malignancy in this group.

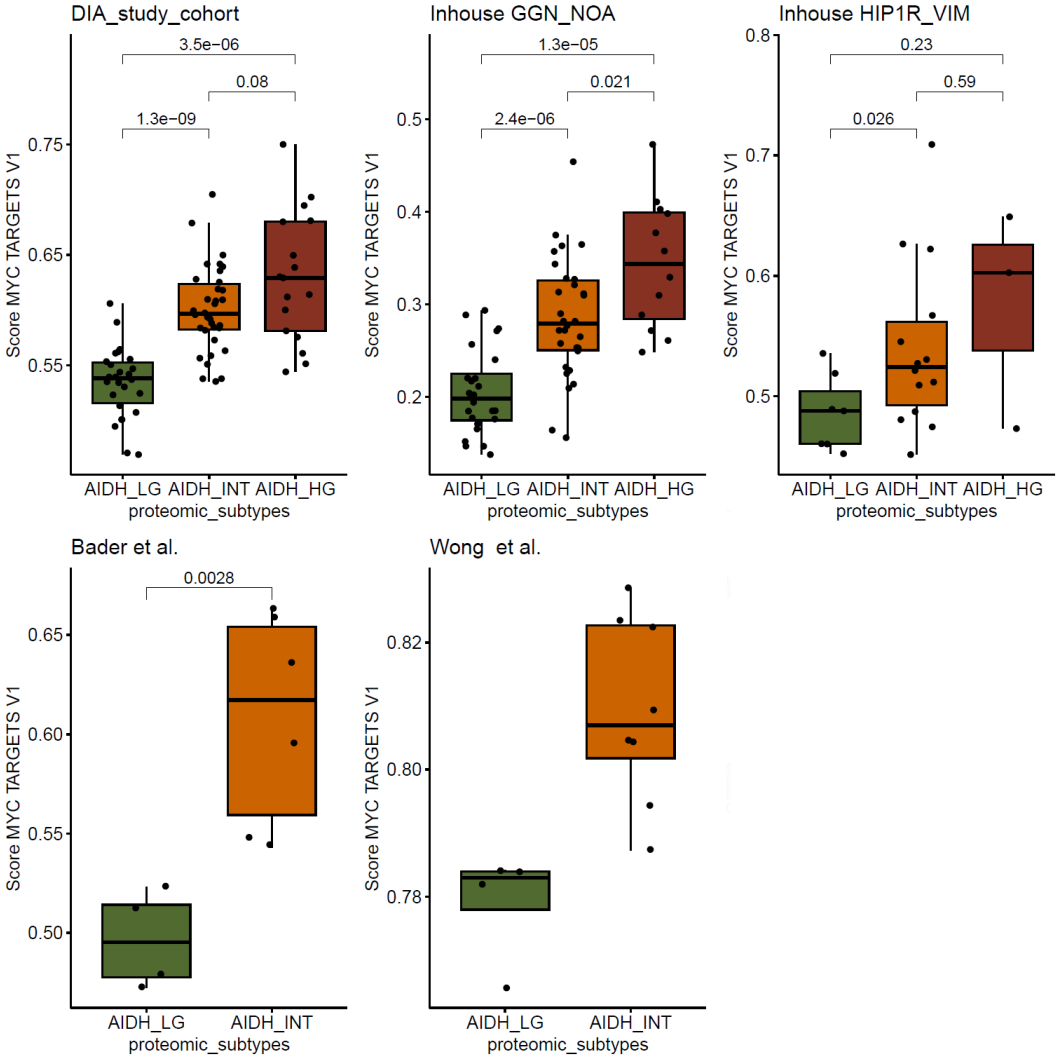


Figure 25.: Evaluation of proteomic subtypes demonstrates consistent upregulation of MYC target proteins in AIDH_INT and AIDH_HG across all tested studies.
Boxplots showing a comparison of MYC scores between the predicted proteomic subtypes showing an upregulation of MYC targets in AIDH_INT and AIDH_HG. gene set score of the MYC hallmark gene set were each estimated for each sample. Statistical significance of between groups was ascertained by conducting a student's-test. Abbreviations: AIDH, Astrocytoma IDH mutant; AIDH_HG, Astrocytoma, IDH mutant high-grade; AIDH_INT, Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma, IDH mutant low-grade.

3.3.2.4. Proteomic subtypes show distinct phosphoproteomic Kinase activities including upregulation of AKT1

To further investigate whether proteomic subtypes can also be distinguished on the phosphoproteomic scale, I examined the adjacent phospho-proteomic data obtained from the study of *Bader et al.* [131] by inferring kinase activities. Next, kinase activity was estimated by performing a differential phosphorylated sites analysis between the predicted cases (4 AIDH_LG, 6 AIDH_INT). Furthermore, phospho-site and kinase relationships were curated from the Omnipath database and contrasting activities between kinases were calculated using the weighted mean function from the R-package decoupler [197].

The results that proteomic subtypes AIDH_LG and AIDH_INT exhibit distinct kinase activities (Figure 26). Importantly, AIDH_INT showed high kinase scores for AKT1 which is in line with the hypothesized AKT1 activation due to dysregulation of complex I [195] identified in this group. Moreover, AIDH_INT also showed a downregulation of the AKT1 target GSK3β [196] providing further consitency into the biological function of AKT1. Interestingly, AIDH_INT also showed high kinase scores of several cell-cycle checkpoint kinases including WEE1, AURKB and ROCK1 in combination with downregulation of several CDKs (CDK12, CDK13) suggesting adaption to replication and transcriptional stress. Thus, the findings demonstrate that proteomic subtypes exhibit distinct phosphoproteomic profiles. Furthermore, increased AKT1 activity consistent with complex I dysregulation provides further biological confidence into the metabolic driven malignant progression assumed for AIDH_INT. Moreover, these findings provide further confidence into the decision tree model by showing strong consistency in the identified key biological processes in the corresponding subtypes identified.

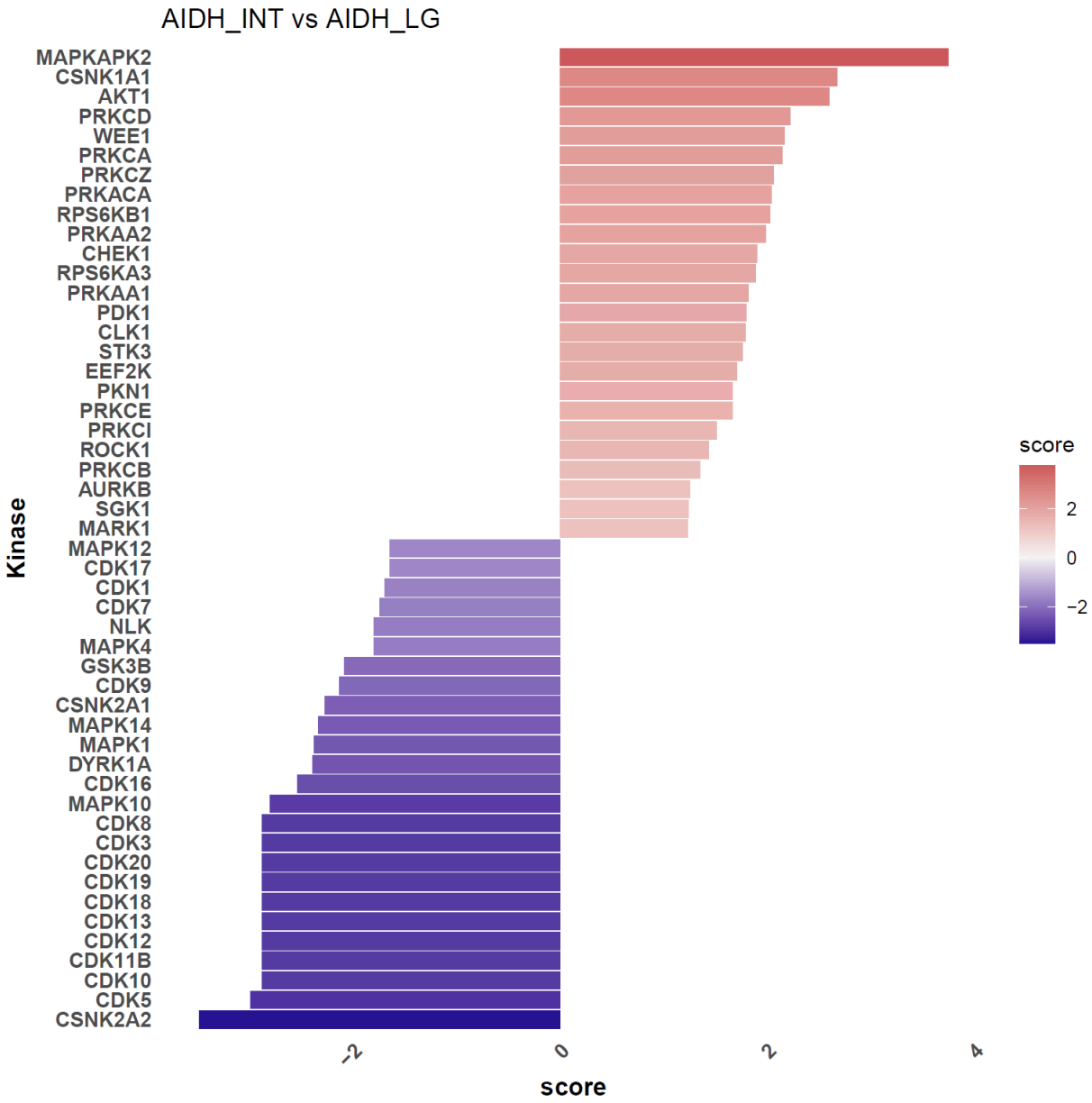


Figure 26: Proteomic subtypes show distinct phosphoproteomic kinase activities.
Bar plot depicting positive and negative kinase activity score in AIDH_INT compared to AIDH_LG. Differential abundant phopsho-sites were determined between the predicted proteomic subtypes from Bader et al [131] using the functions from the R-package limma. Kinase activity scores were calculated by using the run_wmean function from the R-package decoupler [197]. Abbreviations: AIDH, Astrocytoma IDH mutant; AIDH_INT, Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma, IDH mutant low-grade;

3.3.2.5. Evaluation of predicted subgroups in in-house data shows consistency with distinct clinical outcomes between AIHD_LG and AIHD_INT

To further validate recent clinical findings using the AIDH stratification model, the prognostic differences predicted proteomic subtypes were evaluated. Clinical information of patient OS from the two in-house datasets (n=86) were curated and analysed by Kaplan-Meier to assess the differences in clinical outcome between the predicted subtypes. The stratification of patients according to proteomic subtypes in the DDA datasets shows that low grade AIDH predicted into AIDH_INT (p=0.038) and AIDH_HG (p=0.009) had a significant worse survival compared to AIDH_LG (Figure 27 A). Notably, AIDH_HG had previously been described to be most aggressive group it only shows a minor tendency to have a worse outcome in comparison to AIDH_INT. In comparison, risk stratification by WHO grading (Figure 27 B) shows only minor differences in the survival of patients from both groups (p=0.081).

Thus, these results provide consistency with the clinical associations made between proteomic subtypes in the DIA study cohort as they show that stratification identified as AIDH_INT or AIDH_HG perform worse compared AIDH_LG. In addition, the result also confirms prognostic potential by showing a better clinical stratification compared to WHO grading. In addition, these results provide confidence that the developed decision tree model can at least effectively discern between AIDH patients with favourable and unfavourable clinical outcome.

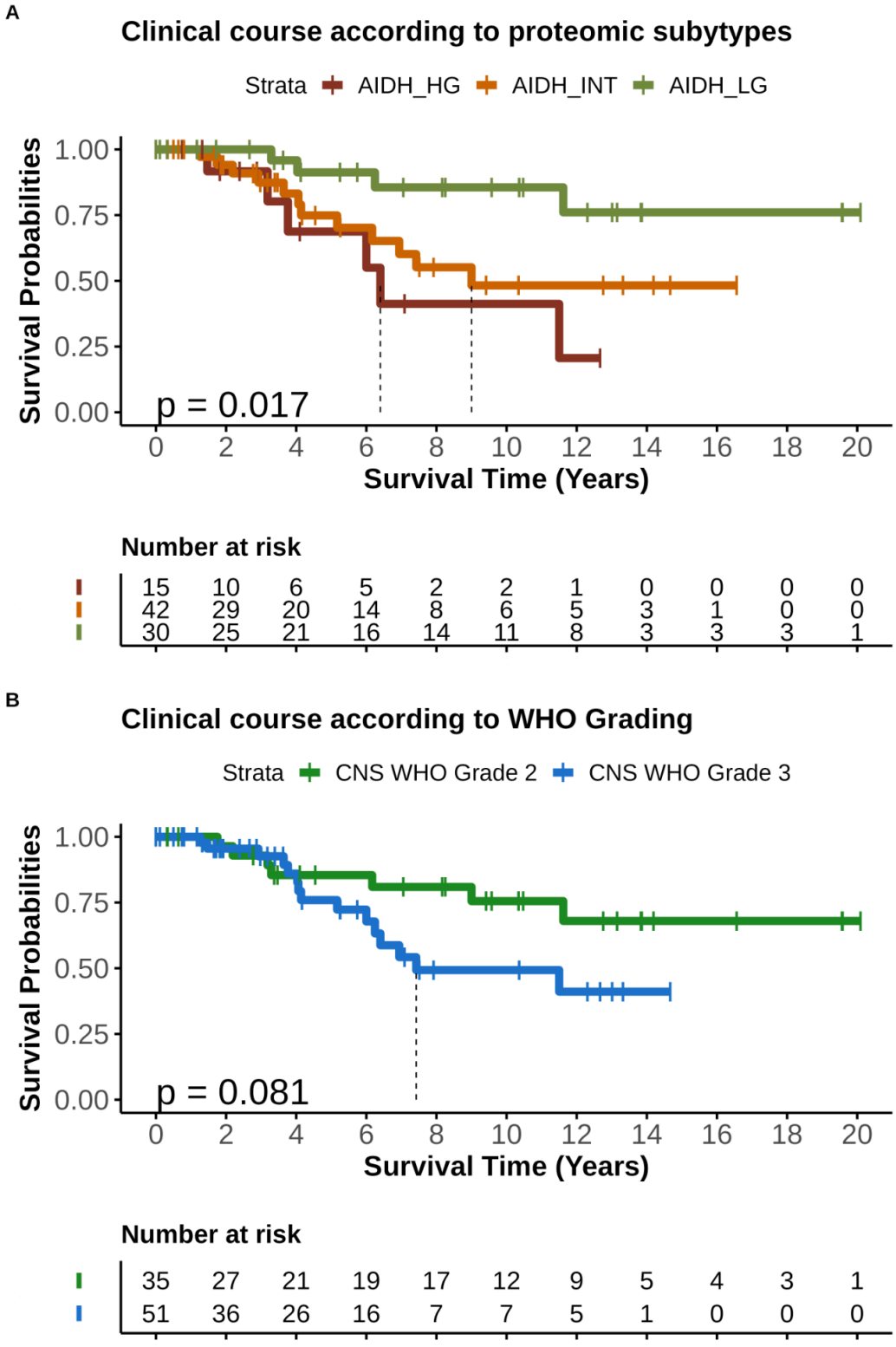


Figure 27: Prediction of proteomic subtypes of AIDH from DDA in-house data (n=86) validates worse clinical outcome of AIDH_INT from the DIA study cohort.
Kaplan Meier Plots depicting the overall survival probabilities of low grade primary AIHD cases from the combined in-house datasets GGN_NOA and HIP1R_VIM stratified by A) predictions from the proteomic subtype decision tree model and B) WHO grading Proteomic subtypes AIDH_HG, AIDH_INT and AIDH_LG are represented by red, orange and green colours and WHO grade 2 and 3 by light green and blue, respectively. The plot shows superior clinical separation in the proteomic groups compared to WHO grading. Abbreviations: AIDH_HG, Astrocytoma IDH mutant high-grade; AIDH_INT, Astrocytoma IDH mutant intermediate; AIDH_LG, Astrocytoma IDH mutant low-grade; CNS, Central Nervous system; WHO World Health Organization.

4. Discussion

Despite the wealth of genetic, transcriptomic, and epigenetic data gathered from numerous studies, identifying a molecular marker for clinical stratification of AIDH remains a challenge [15, 198, 199]. Moreover, the prognostic potential of the AIDH proteome is currently poorly understood. To address this current knowledge gap, the following questions must be considered: 1) How does the proteome contribute to distinct clinical outcomes of primary AIDH patients? 2) Which proteins associate with AIDH malignancy? 3) How can the proteome be used for prognostic stratification AIDH patients? To answer these questions, I investigated the association of the proteome with clinical outcome in a cohort of primary AIDH. Here, I discovered three novel biological and clinical distinct proteomic subtypes of AIDH which showed a better risk stratification of AIDH compared to the current WHO grading system. A neuronal protein expression profile was associated low grade AIDH with favourable outcome while low grade AIDH with a mitochondrial showed intermediate outcome. Mesenchymal protein expression was strongly linked with high-grade AIDH and dismal outcomes. In this context, dysregulation of the mitochondrial complex I and increased mitochondrial biogenesis emerged as a potential hallmark for malignant progression of low-grade AIDH. Exploiting the prognostic potential of the AIDH proteome in a decision-tree model as potential clinical application enabled validation of the study findings in in-house and publicly available datasets and might therefore represent a potential framework for improving clinical stratification of low-grade AIDH.

4.1. Separation of AIDH patients with distinct clinical outcomes based on their proteome.

4.1.1. Curation of a molecular and clinically well documented cohort of primary AIDH cases

A major limiting factor for clinical studies of low-grade AIDH is that these are slow-growing tumours with high heterogeneity in clinical outcomes. While some patients survive for decades, others decease within less than ten years [5]. This variability raises the need of long-term follow-up data for making reliable clinical conclusions, which are often difficult to obtain in large number e.g. due to high cost and logistical constraints. To address this limitation, this study cohort compromised 80 patients including cases with follow-up data up to 20 years after the first diagnosis, providing a robust foundation to draw clinical conclusions (section 3.1.1). Moreover, the cohort represents clinically and molecularly well-documented cases from AIDH from different WHO grades from patients of different genders, ages, and treatments thereby offering a new pivotal resource for further biological and clinical interrogations of this type of tumour.

4.1.2. The proteomic landscape has the potential for prognostic stratification of AIDH

Unsupervised proteomic analysis of the AIDH revealed three novel proteomic subtypes associated with favourable (AIDH_LG), intermediate (AIDH_INT) and dismal (AIDH_HG) patient overall survival (OS) (section 3.1.2). The AIDH_HG subtype was almost exclusively enriched with grade 4 AIDH consistent with previous reports describing grade 4 AIDH as molecularly more distinct from lower grades 2 and 3 [1]. Notably, a small subset of WHO grade 3 AIDH enriched in AIDH_HG, suggesting a comparable aggressive phenotype. Although, no conclusion could be drawn as the number of samples was too small. Larger cohorts will be required to confirm and validate this observation. Comparatively, AIDH_LG and AIDH_INT showed heterogeneous enrichment of grade 2 and grade 3 AIDH, suggesting incoherence between the histological WHO grading and their grouping according to the low-grade AIDH proteome. Importantly, even in absence of WHO grade 4 AIDH cases, there was a significant survival difference between both groups, suggesting that proteomic stratification captures clinical meaningful differences among low-grade AIDH. Findings from the UMAP further illustrated the three discovered subtypes as connected groups suggesting a gradual transition in clinical outcomes from favourable to poor. Importantly, this finding aligns with previous molecular studies indicating that these tumours represent a biological continuum [198] linked with malignant transformation from less aggressive towards more aggressive states [3, 4]. In

summary, these findings show that the AIDH proteome alters with increasing tumour malignancy which can be resolved into three groups corresponding to low, intermediate and highly aggressive AIDH types. Together these results indicated that the proteomic landscape has the potential for prognostic stratification of AIDH. In line with this, the prognostic performance of proteomic subtypes was further compared with that of WHO grading which is the current gold standard for their risk stratification.

The comparison of the WHO grading system with proteomic subtypes in predicting patient OS showed superior performance of the novel stratification framework in both Cox regression models (section 3.1.3.1.). Importantly, when grade 4 tumours were excluded, proteomic subtyping continued to show stronger separation between AIDH_LG and AIDH_INT than WHO grades 2 and 3 in both unilinear and multilinear regression models. This finding is consistent with recent studies reporting that traditional histopathological grading of low-grade AIDH is often inadequate for predicting the variable clinical behaviour of low-grade AIDH [16-18, 20, 200]. These results suggest that the proteome more accurately reflects tumour malignancy and outcome-related alterations than histopathological grading used in the WHO grading system. Interestingly, the prediction model also showed an improved prognostic stratification of grade 4 AIDH suggesting that it could even enhance risk stratification of high-grade AIDH. Unfortunately, only a small subset of grade 4 AIDH was grouped into the lower grade groups AIDH_INT (n=5) and AIDH_LG (n=1) limiting further insight. Consequently, future studies with larger cohorts of primary grade 4 AIDH would aid to further draw a definitive conclusion. To summarize, these findings demonstrate that clinical stratification of AIDH by their proteome provides stronger prognostic information than WHO grading and may have the potential to improve the current grading system of AIDH.

4.1.3. Differences between low grade subtypes substantially arise from variation on the proteomic landscape

The observation that proteomic subtypes are linked with distinct outcomes also further raises the question if these differences might be contributed by genomic alterations including global cop number variations (CNV), gene somatic copy number alterations (SCNA) or epigenetic alterations including CpG hyper/hypomethylation. In line with this it was also questioned if there is an altered distribution of immune cell types in the tumour microenvironment (TME) indicating possible responses of the immune system to tumour malignancy.

Evaluation of CNV burden in proteomic subtypes only demonstrated an increase in the context of AIDH_HG and low burden no differences between AIDH_LG and AIDH_INT (section 3.1.3.2.). This result is coherent with recent studies which showed that an increasing CNV

burden and aneuploidy is often only relevant in the context of high-grade AIDH [13, 28, 29, 35, 186, 201]. To summarize, these findings show that the prognostic differences between AIDH_LG and AIDH_INT are independent from global CNV burden.

Interrogating specific segmental CNVs in the context of proteomic subtypes revealed that some cases in AIHD_INT and AIDH_HG have a partial loss of chromosome 14q, which strongly associated with poor clinical outcomes of cases in both groups (section 3.1.3.3.). This finding is highly interesting, as it suggests loss of the 14q locus as another potential marker for dismal overall survival in AIDH patients. Furthermore, it is likely that genes affected by this loss may have a potential contribution for disease progression and clinical outcome. Moreover, deletion status of 14q has already been discussed to be relevant in the context of shorter progression-free survival in AIDH patients [17]. Perhaps, the differences in prognostic significance might derive from the fact that the discovered alteration only includes a locus of 14q and not the entire arm which narrows down the possible candidates contributing to malignancy. Nevertheless, the study size was not large enough to draw further prognostic conclusions. Consequently, future studies including larger cohorts would aid to further draw a definitive conclusion from this promising novel target.

Gene SCNA including inactivation of tumour suppressors CDKN2A and RB1 trough homozygous deletions as well as activation of oncogenes CDK4, CDK6, PDGFRA, CCND2, MYC, MYCN and PDFRA trough amplifications have recently brought into association with malignant progression of grade AIDH [13, 15, 22-27]. However, analysis of gene level SCNA did not show any subtype-specific enrichments (section 3.1.3.4.) suggesting that prognostic differences between the proteomic subtypes are mostly independent from SCNA.

Findings from the global methylation analysis confirmed that AIDH_LG and AIDH_INT belong to the G-CIMP-high methylation class while AIDH_HG belongs to G-CIMP-low (section 3.1.3.5). These results together with recent literature demonstrated an association between dismal outcome and G-CIMP-low phenotype which are mostly grade 4 [26, 28-30]. Importantly, further evaluation between AIDH_LG and AIHD_INT demonstrated similarity in the methylation pattern between both groups suggesting only minor differences on the scale of CpG methylation. To conclude, these results demonstrate that epigenetic alterations have only minor contribution to the prognostic effects observed between AIDH_LG and AIDH_INT.

Investigation of the TME showed significant shifts in the distribution of tumour and immune cells in the context of the proteomic subtypes (section 3.1.3.6). This suggests that there is an adaptation of the TME in response to the malignancy of the proteomic subtypes. Interestingly,

a higher proportion of tumour cells was linked with a worse outcome in low-grade AIDH in AIDH_INT which could be due to acquisition of stronger proliferative potential leading to higher density of tumour cells. This finding emphasises grading strategies suggesting presence of ≥ 2 mitoses per 10 high-power fields (HPF) as grade 3 criterion [18, 202]. On the other hand, a resolution in efficacy of the immune system is indicated as AIDH_INT showed a decrease of CD4-Effector cells which are responsible for recruitment and activation of immune cells [187]. Based on these findings, further evaluation of proteomic subtypes on the scale of their single cell protein or mRNA expression could provide further insight into the immune cell types and their contribution to tumour malignancy.

In conclusion, the findings from the evaluation of known risk-associated genetic and epigenetic alterations provide confidence that the prognostic differences between proteomic subtypes AIDH_LG and AIDH_INT primarily arise from variations in the proteomic landscape. These differences are independent of CNV burden, epigenetic alterations, as well as gene-level SCNA linked with malignant progression. Moreover, partial loss of 14q and shifts in the TME composition in the context of proteomic subtypes indicate novel biomarkers which could be used for additional AIDH stratification and further could lead to a better understanding of their malignant progression.

4.2. Investigation of proteins and processes linked with malignancy of AIDH subtypes

4.2.1. Proteomic stratification reveals distinct biological states associated with AIDH malignancy

By obtaining the differential expressed proteins (DEP) across proteomic subtypes, distinct functional biological programs were discovered (section 3.2.1.) and characterized. Neuronal protein expression was linked with AIDH_LG, metabolic and mitochondrial proteins with AIDH_INT, and stark expression of extracellular and mesenchymal proteins was found in AIDH_HG. These patterns provide biological significance to the proteomic subtypes and their distinct clinical outcomes, demonstrating that their clustering mirrors molecular meaningful differences rather than technical artifacts. Furthermore, association with the corresponding clinical outcomes suggests a potential contribution of the identified proteins and biological processes to AIDH malignancy by suppressing or promoting malignant progression.

In this context, low tumour malignancy in AIDH_LG was associated with high expression of neuronal proteins involved in synaptic functions including neurotransmitter transmission, transport and uptake. As recent results from the deconvolution analysis did not show any increased proportions of glial or neuronal cells which could mean on the one hand that tumour cells reflect a more differentiated phenotype compared with the other subtypes. On the other hand, it could also indicate that AIDH_LG have long-range synaptic connections with neuronal cells without their presence in tissue [203]. Nevertheless, this result indicates that neuronal proteins expression may serve as a potential marker for favourable prognosis of AIDH. This observation is coherent with proteomic investigations of IDH WT GBM and IDH mutant glioma reporting favourable prognosis in patient groups with elevated neuronal protein expression [204-206]. In particular, high expression of potential tumour suppressor proteins including Claudin-11 (CLDN11), proline-rich transmembrane protein 2 (PRRT2) as well as Syntaxin-Binding Protein 1 (STXBP1) [207-210] could contribute to reduced malignancy of this group. Notably, STXBP1 knockdown has been linked with increased expression of markers of epithelial mesenchymal transition (EMT) as well as enhanced proliferation, migration and invasion in glioma [210] suggesting it a potential suppressor of malignant progression of AIDH. This implication is further supported as recent findings show an association with low expression of STXBP1 with a shift into mesenchymal protein expression in the context of aggressive AIDH_HG. To summarize, these findings suggest that neuronal protein expression as a key

biological feature of low malignant AIDH, contrasting to the expression programs in more malignant subtypes.

A frequently discussed question in malignant AIDH progression is the mechanism which drives the metabolic independence form the IDH mutations. Low grade AIDH have been shown to depend in their metabolism on mitochondrial OXPHOS [40] while late stage AIDH have been shown to shift their metabolism and increase glycolytic uptake and lactate production which is a hallmark for the Warburg effect [39]. In this context, the finding that differential proteins expressed in AIDH_INT contribute to mitochondrial biogenesis and metabolic processes was highly interesting. This suggest that mitochondrial remodelling may contribute to malignant transformation of low-grade AIDH. This is supported by findings showing high expression of mitochondrial regulators TRAP1 and SIRT3. Notably, SIRT3 is a NAD+-depended deacetylase and regulator for mitochondrial integrity and metabolic reprogramming. SIRT3 has a two-sided role in cancer as it has been described as oncogene as well as tumour suppressor in different cancer types [211]. In contrast, TRAP1 is a mitochondrial chaperon frequently highly expressed in hypoxic conditions and controlled by the transcription factor HIF-1α [212]. High TRAP1 levels have been shown to contribute to neoplastic growth of tumours by inhibiting cytochrome c oxidase [213] and Succinate dehydrogenase (SDH) which leads to further stabilization of HIF-1α [214]. Importantly, recent studies described a crosstalk between SIRT3 and TRAP1 leading to adaption of metabolic changes to cope with metabolic stress to drive neoplastic progression [188]. In this regard, these findings suggest metabolic remodelling mediated by SIRT3 and TRAP1 crosstalk as one possible mechanism of how low-grade AIDH could become more independent from IDH1 mutation and acquire more malignant potential. To conclude, the expression pattern of metabolic and mitochondrial proteins emerges as a key biological feature of AIDH with intermediate prognosis.

In contrast, AIDH_HG showed a mesenchymal shift in protein expression, towards an EMT-like phenotype. This shift was indicated by an enriched expression of EMT markers and decrease of Cadherins as well as proteins involved in extracellular matrix organisation, wound healing, and inflammation. These processes have been described as hallmark biological processes linked to epithelial-to-mesenchymal transition (EMT) in other cancer types [189]. Notably, this finding is concordant with the clinical course of ADIH_HG, as current literature demonstrating that EMT often associates with poor outcomes in GBM [215] by contributing to increased tumour motility, aggressive invasive growth and potentially chemotherapy resistance [216]. Moreover, these findings align also with results from two other proteomic studies of IDH mutant glioma which described an WHO grade 4 enriched immune-hot proteomic subtypes

with dismal clinical outcome and mesenchymal expression profile [205]. Interestingly, nicotinamide-N-methyltransferase (NNMT) a protein which has already been reported in several aggressive tumour types was the highest expressed protein in this group suggesting a potential contribution to the observed dismal prognosis. NNMT plays an important role in tumours metabolism as it transfers methyl group from S-adenosyl-L-methionine (SAM) to nicotinamide (NA) leading to the generation of S-adenosylhomocysteine (SAH) and 1-methylnicotinamide (1MNA) which has been shown to limit the pool of the universal methyl donor SAM thereby reducing the global methylation capacity. Although this mechanism proved mainly to be mainly relevant for the methylation of histones and several other proteins, an influence on DNA methylation remains an interesting possibility, especially in highly malignant IDH-mutant tumours where global demethylation is an important molecular marker [217]. To conclude, these findings show a mesenchymal expression shift in AIHD_HG as the endpoint of malignant progression.

Taken together, the identified proteomic patterns provide insight that AIDH malignancy is linked with distinct biological expression programs. Furthermore, this has important implications on that malignant progression of AIDH might be reflected by a loss of neuronal differentiation towards a mesenchymal state, with metabolic remodelling in AIDH_INT serving as a bridging step to gain independence for the IDH1 mutation.

4.2.2. Dysregulation of complex I subunits is a hallmark of metabolic reprogramming in AIDH_INT

Characterisation of mitochondrial (MT) proteins in AIDH_INT revealed striking enrichment of proteins contributing to MT biogenesis and metabolic processes (section 3.2.2). This finding is in contrast to current established reports of low-grade AIHD metabolism, which is highly dependent on the electron transport chain (ETC) complexes I for oxidative phosphorylation (OXPHOS) and suggests a potential selective reprogramming of MTs in AIDH_INT [40]

Furthermore, findings showing enrichment of lipid and amino acid metabolism as alternative energy pathways and MT biogenesis through increased expression of Mito-ribosomes could indicate possible responses to MT stress or a compensatory program upon complex I dysregulation. Nevertheless, given the scarcity of NADH in AIDH due to the IDH mutation, dysregulation of complex I could also be a compensatory step to impair NADH oxidation and reduce the oxidative [57, 67] and lipid de novo synthesis stress [58] caused by the lack of NADH. Intriguingly, dysregulation and mutations of complex I have been described to promote tumour progression by activation of AKT1 [195] an important player for a series of oncogenic pathways contributing to further malignant progression.

In this regard, investigation of oncogenic pathways linked with AKT1 revealed strong enrichment of proteins of the oncogenic Myelocytomatosis (MYC) pathway in AIHD_INT (section 3.2.3). AKT1 has been described to promote MYC activity by inhibiting GSK3β and preventing its ubiquitination through phosphorylation at Thr58 [196]. This is biologically coherent with the observations as MYC has been reported as enhancer of mitochondrial biogenesis and metabolic processes including fatty acid metabolism and OXPHOS [218]. Furthermore, MYC amplifications have been described as genetic event which associated with AIDH patients WHO grades 2-4 with shorter overall survival [22, 27] supporting that a mechanism to activate MYC in might be relevant for tumour progression. Thus, it implicates complex I deficiency of AIHD_INT could indirectly lead to a metabolic rewiring and acquisition of a more malignant phenotypes through the AKT1 promoted activation of MYC.

To conclude, these findings provide an unexpected heterogeneity of metabolic and mitochondrial proteins among low grade AIDH and identify increased mitochondrial biogenesis and complex I dysregulation as potential hallmark for metabolic adaption to the IDH1 mutation and malignant progression. The cause for complex I deficiency in AIDH is unknown. Interestingly, complex I deficiency in other tumor types is often due to mutations in the mitochondrial genome which has not been comprehensively analyzed in AIDH. Following studies could focus on identifying the molecular cause for complex I deficiency as this might represent a critical step in initiating tumor progression in AIDH [219].

4.3. Molecular stratification of AIDH by their proteome

4.3.1. Decision-tree based classification of AIDH by protein expression

Current molecular studies report low-grade AIDH as a biological continuum linked with the malignant transformation rather than distinct biological tumour groups. Consequently, due to this lack of a clear biological breaking point, empirical evidence was used to define an artificial cut-off on an identified biomarkers to divide low grade AIDH into favourable and poor outcomes [198]. However, this procedure comes with several setbacks and pitfalls possibly leading to wrong conclusion about clinical outcome [220, 221]. To overcome this limitation, a decision-tree-based model was developed utilizing the recent biological specific programs identified to predict the proteomic subtype of AIDH from their proteomic expression profile. A decision-tree classifier was used as stratification strategy due to the following reasons: Firstly it performs comparable to advanced machine learning models like XGBoost and Random Forest for smaller datasets [222] while still providing biological interpretability [223]. Secondly, number of features to train was low as well as subtype-specific biological protein expression programmes were summarized into signature scores supporting the use of a simple prediction model like decision-tree. Signature scoring was used as recent studies investigating the proteomes of glioblastomas showed better efficacy in prognostic stratification as it broadens the molecular stratification and ensure accuracy [134, 135]. In this context the decision-tree algorithm demonstrated high discriminatory power by utilizing the signature scores of the three subtype-specific biological expression programmes (section 3.3.1), highlighting the choice of the selected model and the robustness of the biological distinction observed between the subtypes.

Classification and characterisation AIDH cases form in-house and publicly available datasets provided validation of the metabolic phenotype as well as enriched expression of MYC target proteins in AIDH_INT (section 3.3.2). This provides further confidence on the significance of the discovered biological signature by showing that it can be reproducible found across other datasets as well. Intriguingly, proteomic subtypes predicted from AIDH from the dataset of *Bader et al.* [131] showed concordance between the groups HGG-IDHmut-A and AIHD_LG and HGG-IDHmut-B and AIDH_INT. Nevertheless, the findings from *Bader et al.* [131] suggested no clinical differences in the identified groups and a decline of mitochondrial proteins including the ETC complexes as response to the Warburg effect. Perhaps, these differences might come from to the low number of AIDH cases in the cohort of and the limited coverage of the proteome due to DDA based whole proteome analysis.

Moreover, evaluation of kinase activities inference of matched phospho-proteomic data of AIDH from *Bader et al.* [131] showed an increased activity of kinases contributing to tumour malignancy in AIDH_INT (section 3.3.2.4.). In this context, AIDH_INT suggested increased activity of several cell-cycle proteins including WEE1 [224], ROCK1 [225] and AURKB [226] suggesting an increased mitotic activity in AIDH_INT cases compared to AIDH_LG. Moreover, kinase activities inference also revealed AKT1 phosphorylation and decreased GSK3β dephosphorylation which is in line with previous assumptions of a MYC based activation by the stabilization through AKT1 driven inactivation of the GSK3β [196]. Taken together, these findings show that AIDH_INT also exhibit malignant progression associated profiles on the level of their phospho-proteome.

By classifying in-house AIDH cases, I was able to show distinct clinical outcomes between AIDH_LG and AIDH_INT (section 3.3.2.5.). This provides further confidence that accurate subtype prediction could support risk stratification of low-grade AIDH patients. Nevertheless, while some low-grade AIDH cases classified into AIDH_HG showed a worse outcome compared to AIDH_LG and AIDH_INT, the differences to AIDH_INT were less compared to the DIA study cohort. In this context, this might suggest, that the current classification model might be less efficient to stratify between intermediate and high-risk AIDH cases. Nevertheless, it was also observed that a small subset of low-grade AIDH clustered into AIDH_HG suggesting that these cases might represent an intermediate stage between AIDH_INT and the grade 4 AIDH in AIDH_HG.

To conclude the introduced decision tree model provides a framework to improve molecular prognostic stratification of low-grade AIDH patients using LC-MS based proteomics. Subsequently, further clinical refinement from larger multicentre cohorts such as the GLASS [227] or the GGN network [139] would provide further granularity and confidence and could aid to provide further confidence on this hypothesis.

.

5. Conclusion

The findings of the introduced thesis enhance the molecular and clinical understanding of AIDH by conducting bioinformatic analysis of the proteomes of a clinically, epigenetically, and histologically well characterized cohort of primary AIDH. It characterizes three novel AIDH subtypes at the proteomic level each with distinct clinical outcome and biological phenotypes linked with neuronal, metabolic and mesenchymal expression programs. Furthermore, it provides insight into proteomic alterations in the context of distinct types of AIDH and has potentially important implications regarding our understanding of malignant progression in AIDH. Moreover, the developed decision-tree model may serve as a potential framework for clinical stratification of low-grade AIDH patients.

6. References

1. Louis, D.N., et al., *The 2021 WHO Classification of Tumors of the Central Nervous System: a summary.* Neuro Oncol, 2021. **23**(8): p. 1231-1251.
2. Weller, M., et al., *EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood.* Nat Rev Clin Oncol, 2021. **18**(3): p. 170-186.
3. Board, W.C.o.T.E., *Central Nervous System Tumours: Who Classification of Tumours*. 2022: World Health Organization.
4. Miller, J.J., et al., *Isocitrate dehydrogenase (IDH) mutant gliomas: A Society for Neuro-Oncology (SNO) consensus review on diagnosis, management, and future directions.* Neuro Oncol, 2023. **25**(1): p. 4-25.
5. Katzendobler, S., et al., *Determinants of long-term survival in patients with IDH-mutant gliomas.* J Neurooncol, 2024. **170**(3): p. 655-664.
6. Zong, H., R.G. Verhaak, and P. Canoll, *The cellular origin for malignant glioma and prospects for clinical advancements.* Expert Rev Mol Diagn, 2012. **12**(4): p. 383-94.
7. Ostrom, Q.T., et al., *CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014-2018.* Neuro Oncol, 2021. **23**(12 Suppl 2): p. iii1-iii105.
8. 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.
9. 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.
10. Appay, R., et al., *CDKN2A homozygous deletion is a strong adverse prognosis factor in diffuse malignant IDH-mutant gliomas.* Neuro Oncol, 2019. **21**(12): p. 1519-1528.
11. Lu, V.M., et al., *The prognostic significance of CDKN2A homozygous deletion in IDH-mutant lower-grade glioma and glioblastoma: a systematic review of the contemporary literature.* J Neurooncol, 2020. **148**(2): p. 221-229.
12. Reis, G.F., et al., *CDKN2A loss is associated with shortened overall survival in lower-grade (World Health Organization Grades II-III) astrocytomas.* J Neuropathol Exp Neurol, 2015. **74**(5): p. 442-52.
13. Shirahata, M., et al., *Novel, improved grading system(s) for IDH-mutant astrocytic gliomas.* Acta Neuropathol, 2018. **136**(1): p. 153-166.
14. Daumas-Duport, C., et al., *Grading of astrocytomas. A simple and reproducible method.* Cancer, 1988. **62**(10): p. 2152-65.
15. Reuss, D.E., *Updates on the WHO diagnosis of IDH-mutant glioma.* J Neurooncol, 2023. **162**(3): p. 461-469.
16. Aoki, K., et al., *Prognostic relevance of genetic alterations in diffuse lower-grade gliomas.* Neuro Oncol, 2018. **20**(1): p. 66-77.
17. 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.
18. 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.
19. Yang, R.R., et al., *IDH mutant lower grade (WHO Grades II/III) astrocytomas can be stratified for risk by CDKN2A, CDK4 and PDGFRA copy number alterations.* Brain Pathol, 2020. **30**(3): p. 541-553.
20. 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.
21. Stupp, R., et al., *Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma.* N Engl J Med, 2005. **352**(10): p. 987-96.

22. Lee, K., et al., *Genomic profiles of IDH-mutant gliomas: MYCN-amplified IDH-mutant astrocytoma had the worst prognosis.* Sci Rep, 2023. **13**(1): p. 6761.
23. Wong, Q.H., et al., *Correction to: Molecular landscape of IDH-mutant primary astrocytoma Grade IV/Glioblastomas.* Mod Pathol, 2021. **34**(6): p. 1231.
24. Zhao, Z., et al., *Chinese Glioma Genome Atlas (CGGA): A Comprehensive Resource with Functional Genomic Data from Chinese Glioma Patients.* Genomics Proteomics Bioinformatics, 2021. **19**(1): p. 1-12.
25. Jonsson, P., et al., *Genomic Correlates of Disease Progression and Treatment Response in Prospectively Characterized Gliomas.* Clin Cancer Res, 2019. **25**(18): p. 5537-5547.
26. Ceccarelli, M., et al., *Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma.* Cell, 2016. **164**(3): p. 550-63.
27. Korshunov, A., et al., *Integrated molecular characterization of IDH-mutant glioblastomas.* Neuropathol Appl Neurobiol, 2019. **45**(2): p. 108-118.
28. Tesileanu, C.M.S., et al., *Prognostic significance of genome-wide DNA methylation profiles within the randomized, phase 3, EORTC CATNON trial on non-1p/19q deleted anaplastic glioma.* Neuro Oncol, 2021. **23**(9): p. 1547-1559.
29. Mamatjan, Y., et al., *Integrated molecular analysis reveals hypermethylation and overexpression of HOX genes to be poor prognosticators in isocitrate dehydrogenase mutant glioma.* Neuro Oncol, 2023. **25**(11): p. 2028-2041.
30. Li, K.K., et al., *Identification of subsets of IDH-mutant glioblastomas with distinct epigenetic and copy number alterations and stratified clinical risks.* Neurooncol Adv, 2019. **1**(1): p. vdz015.
31. Bai, H., et al., *Integrated genomic characterization of IDH1-mutant glioma malignant progression.* Nat Genet, 2016. **48**(1): p. 59-66.
32. Mazor, T., et al., *DNA Methylation and Somatic Mutations Converge on the Cell Cycle and Define Similar Evolutionary Histories in Brain Tumors.* Cancer Cell, 2015. **28**(3): p. 307-317.
33. Capper, D., et al., *DNA methylation-based classification of central nervous system tumours.* Nature, 2018. **555**(7697): p. 469-474.
34. 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.
35. Kling, T., et al., *Refinement of prognostication for IDH-mutant astrocytomas using DNA methylation-based classification.* Brain Pathol, 2024. **34**(5): p. e13233.
36. Watanabe, T., et al., *IDH1 mutations are early events in the development of astrocytomas and oligodendrogliomas.* Am J Pathol, 2009. **174**(4): p. 1149-53.
37. Johnson, B.E., et al., *Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma.* Science, 2014. **343**(6167): p. 189-193.
38. Lai, A., et al., *Evidence for sequenced molecular evolution of IDH1 mutant glioblastoma from a distinct cell of origin.* J Clin Oncol, 2011. **29**(34): p. 4482-90.
39. Johannessen, T.A., et al., *Rapid Conversion of Mutant IDH1 from Driver to Passenger in a Model of Human Gliomagenesis.* Mol Cancer Res, 2016. **14**(10): p. 976-983.
40. Grassian, A.R., et al., *IDH1 mutations alter citric acid cycle metabolism and increase dependence on oxidative mitochondrial metabolism.* Cancer Res, 2014. **74**(12): p. 3317-31.
41. Horton, J.R., et al., *Characterization of a Linked Jumonji Domain of the KDM5/JARID1 Family of Histone H3 Lysine 4 Demethylases.* J Biol Chem, 2016. **291**(6): p. 2631-46.
42. Ciccarone, F., L. Di Leo, and M.R. Ciriolo, *Aberrations of the TCA Cycle in Cancer*, in *Reference Module in Biomedical Sciences*, P. Boffetta and P. Hainaut, Editors. 2018, Academic Press: Oxford. p. 429-436.
43. Kayabolen, A., E. Yilmaz, and T. Bagci-Onder, *IDH Mutations in Glioma: Double-Edged Sword in Clinical Applications?* Biomedicines, 2021. **9**(7).
44. Hartmann, C., et al., *Type and frequency of IDH1 and IDH2 mutations are related to astrocytic and oligodendroglial differentiation and age: a study of 1,010 diffuse gliomas.* Acta Neuropathol, 2009. **118**(4): p. 469-74.
45. Mellinghoff, I.K., et al., *Vorasidenib in IDH1- or IDH2-Mutant Low-Grade Glioma.* N Engl J Med, 2023. **389**(7): p. 589-601.

46. Ye, D., Y. Xiong, and K.L. Guan, *The mechanisms of IDH mutations in tumorigenesis.* Cell Res, 2012. **22**(7): p. 1102-4.
47. Xu, W., et al., *Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases.* Cancer Cell, 2011. **19**(1): p. 17-30.
48. Chowdhury, R., et al., *The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases.* EMBO Rep, 2011. **12**(5): p. 463-9.
49. Turcan, S., et al., *IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype.* Nature, 2012. **483**(7390): p. 479-83.
50. Lu, C., et al., *IDH mutation impairs histone demethylation and results in a block to cell differentiation.* Nature, 2012. **483**(7390): p. 474-8.
51. Fu, X., et al., *2-Hydroxyglutarate Inhibits ATP Synthase and mTOR Signaling.* Cell Metab, 2015. **22**(3): p. 508-15.
52. Sasaki, M., et al., *D-2-hydroxyglutarate produced by mutant IDH1 perturbs collagen maturation and basement membrane function.* Genes Dev, 2012. **26**(18): p. 2038-49.
53. Viswanath, P., et al., *2-Hydroxyglutarate-Mediated Autophagy of the Endoplasmic Reticulum Leads to an Unusual Downregulation of Phospholipid Biosynthesis in Mutant IDH1 Gliomas.* Cancer Res, 2018. **78**(9): p. 2290-2304.
54. Lee, S.M., et al., *Cytosolic NADP(+)-dependent isocitrate dehydrogenase status modulates oxidative damage to cells.* Free Radic Biol Med, 2002. **32**(11): p. 1185-96.
55. Jo, S.H., et al., *Control of mitochondrial redox balance and cellular defense against oxidative damage by mitochondrial NADP+-dependent isocitrate dehydrogenase.* J Biol Chem, 2001. **276**(19): p. 16168-76.
56. Koh, H.J., et al., *Cytosolic NADP+-dependent isocitrate dehydrogenase plays a key role in lipid metabolism.* J Biol Chem, 2004. **279**(38): p. 39968-74.
57. Badur, M.G., et al., *Oncogenic R132 IDH1 Mutations Limit NADPH for De Novo Lipogenesis through (D)2-Hydroxyglutarate Production in Fibrosarcoma Sells.* Cell Rep, 2018. **25**(4): p. 1018-1026 e4.
58. Lita, A., et al., *IDH1 mutations induce organelle defects via dysregulated phospholipids.* Nat Commun, 2021. **12**(1): p. 614.
59. Fack, F., et al., *Altered metabolic landscape in IDH-mutant gliomas affects phospholipid, energy, and oxidative stress pathways.* EMBO Mol Med, 2017. **9**(12): p. 1681-1695.
60. Salamanca-Cardona, L., et al., *In Vivo Imaging of Glutamine Metabolism to the Oncometabolite 2-Hydroxyglutarate in IDH1/2 Mutant Tumors.* Cell Metab, 2017. **26**(6): p. 830-841 e3.
61. Newsholme, P., et al., *Glutamine and glutamate--their central role in cell metabolism and function.* Cell Biochem Funct, 2003. **21**(1): p. 1-9.
62. Altman, B.J., Z.E. Stine, and C.V. Dang, *From Krebs to clinic: glutamine metabolism to cancer therapy.* Nat Rev Cancer, 2016. **16**(10): p. 619-34.
63. Seltzer, M.J., et al., *Inhibition of glutaminase preferentially slows growth of glioma cells with mutant IDH1.* Cancer Res, 2010. **70**(22): p. 8981-7.
64. Chen, R., et al., *Hominoid-specific enzyme GLUD2 promotes growth of IDH1R132H glioma.* Proc Natl Acad Sci U S A, 2014. **111**(39): p. 14217-22.
65. van Lith, S.A., et al., *Glutamate as chemotactic fuel for diffuse glioma cells: are they glutamate suckers?* Biochim Biophys Acta, 2014. **1846**(1): p. 66-74.
66. Chan, S.M., et al., *Isocitrate dehydrogenase 1 and 2 mutations induce BCL-2 dependence in acute myeloid leukemia.* Nat Med, 2015. **21**(2): p. 178-84.
67. Gelman, S.J., et al., *Consumption of NADPH for 2-HG Synthesis Increases Pentose Phosphate Pathway Flux and Sensitizes Cells to Oxidative Stress.* Cell Rep, 2018. **22**(2): p. 512-522.
68. Ou, X., et al., *Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV.* Nat Commun, 2020. **11**(1): p. 1620.
69. Yalaza, C., et al., *R132H Mutation in IDH1 Gene is Associated with Increased Tumor HIF1-Alpha and Serum VEGF Levels in Primary Glioblastoma Multiforme.* Ann Clin Lab Sci, 2017. **47**(3): p. 362-364.

70. Fujiwara, S., et al., *Silencing hypoxia-inducible factor-1alpha inhibits cell migration and invasion under hypoxic environment in malignant gliomas.* Int J Oncol, 2007. **30**(4): p. 793-802.
71. Schwenk, J.M., et al., *The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays.* J Proteome Res, 2017. **16**(12): p. 4299-4310.
72. Chen, D., et al., *Proteomics and microstructure profiling of goat milk protein after homogenization.* J Dairy Sci, 2019. **102**(5): p. 3839-3850.
73. Mohanta, T.K., et al., *The molecular mass and isoelectric point of plant proteomes.* BMC Genomics, 2019. **20**(1): p. 631.
74. Aslam, B., et al., *Proteomics: Technologies and Their Applications.* J Chromatogr Sci, 2017. **55**(2): p. 182-196.
75. Liu, Y., A. Beyer, and R. Aebersold, *On the Dependency of Cellular Protein Levels on mRNA Abundance.* Cell, 2016. **165**(3): p. 535-50.
76. 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.
77. Woods, A.G., et al., *Mass Spectrometry for Proteomics-Based Investigation.* Adv Exp Med Biol, 2019. **1140**: p. 1-26.
78. Aebersold, R. and M. Mann, *Mass-spectrometric exploration of proteome structure and function.* Nature, 2016. **537**(7620): p. 347-55.
79. Orton, D.J. and A.A. Doucette, *Proteomic Workflows for Biomarker Identification Using Mass Spectrometry - Technical and Statistical Considerations during Initial Discovery.* Proteomes, 2013. **1**(2): p. 109-127.
80. Schirle, M., M. Bantscheff, and B. Kuster, *Mass spectrometry-based proteomics in preclinical drug discovery.* Chem Biol, 2012. **19**(1): p. 72-84.
81. Crutchfield, C.A., et al., *Advances in mass spectrometry-based clinical biomarker discovery.* Clin Proteomics, 2016. **13**: p. 1.
82. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.
83. Di Meo, A., M.D. Pasic, and G.M. Yousef, *Proteomics and peptidomics: moving toward precision medicine in urological malignancies.* Oncotarget, 2016. **7**(32): p. 52460-52474.
84. Toby, T.K., L. Fornelli, and N.L. Kelleher, *Progress in Top-Down Proteomics and the Analysis of Proteoforms.* Annu Rev Anal Chem (Palo Alto Calif), 2016. **9**(1): p. 499-519.
85. Po, A. and C.E. Eyers, *Top-Down Proteomics and the Challenges of True Proteoform Characterization.* J Proteome Res, 2023. **22**(12): p. 3663-3675.
86. Dupree, E.J., et al., *A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of this Field.* Proteomes, 2020. **8**(3).
87. Rogers, J.C. and R.D. Bomgarden, *Sample Preparation for Mass Spectrometry-Based Proteomics; from Proteomes to Peptides.* Adv Exp Med Biol, 2016. **919**: p. 43-62.
88. Han, X., A. Aslanian, and J.R. Yates, 3rd, *Mass spectrometry for proteomics.* Curr Opin Chem Biol, 2008. **12**(5): p. 483-90.
89. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present.* Anal Bioanal Chem, 2012. **404**(4): p. 939-65.
90. Jiang, Y., et al., *Comprehensive Overview of Bottom-Up Proteomics Using Mass Spectrometry.* ACS Meas Sci Au, 2024. **4**(4): p. 338-417.
91. 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.
92. Li, J., L.S. Smith, and H.J. Zhu, *Data-independent acquisition (DIA): An emerging proteomics technology for analysis of drug-metabolizing enzymes and transporters.* Drug Discov Today Technol, 2021. **39**: p. 49-56.
93. Yates, J.R., 3rd, et al., *Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database.* Anal Chem, 1995. **67**(8): p. 1426-36.

94. Stahl, D.C., et al., *Data-controlled automation of liquid chromatography/tandem mass spectrometry analysis of peptide mixtures.* J Am Soc Mass Spectrom, 1996. **7**(6): p. 532-40.
95. Nesvizhskii, A.I., *Protein identification by tandem mass spectrometry and sequence database searching.* Methods Mol Biol, 2007. **367**: p. 87-119.
96. Verheggen, K., et al., *Database Search Engines: Paradigms, Challenges and Solutions.* Adv Exp Med Biol, 2016. **919**: p. 147-156.
97. 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.
98. Tyanova, S., T. Temu, and J. Cox, *The MaxQuant computational platform for mass spectrometry-based shotgun proteomics.* Nat Protoc, 2016. **11**(12): p. 2301-2319.
99. Carr, S.A., et al., *Targeted Peptide Measurements in Biology and Medicine: Best Practices for Mass Spectrometry-based Assay Development Using a Fit-for-Purpose Approach *.* Molecular & Cellular Proteomics, 2014. **13**(3): p. 907-917.
100. Bourmaud, A., S. Gallien, and B. Domon, *Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: Principle and applications.* Proteomics, 2016. **16**(15-16): p. 2146-59.
101. Shi, J., et al., *Comparison of protein expression between human livers and the hepatic cell lines HepG2, Hep3B, and Huh7 using SWATH and MRM-HR proteomics: Focusing on drug-metabolizing enzymes.* Drug Metab Pharmacokinet, 2018. **33**(2): p. 133-140.
102. Nakamura, K., et al., *Large-scale multiplex absolute protein quantification of drug-metabolizing enzymes and transporters in human intestine, liver, and kidney microsomes by SWATH-MS: Comparison with MRM/SRM and HR-MRM/PRM.* Proteomics, 2016. **16**(15-16): p. 2106-17.
103. Vildhede, A., et al., *Comparison of Proteomic Quantification Approaches for Hepatic Drug Transporters: Multiplexed Global Quantitation Correlates with Targeted Proteomic Quantitation.* Drug Metab Dispos, 2018. **46**(5): p. 692-696.
104. Li, J. and H.J. Zhu, *Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)-Based Proteomics of Drug-Metabolizing Enzymes and Transporters.* Molecules, 2020. **25**(11).
105. Ludwig, C., et al., *Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial.* Mol Syst Biol, 2018. **14**(8): p. e8126.
106. Shi, T., et al., *Advances in targeted proteomics and applications to biomedical research.* Proteomics, 2016. **16**(15-16): p. 2160-82.
107. Huang, Q., et al., *SWATH enables precise label-free quantification on proteome scale.* Proteomics, 2015. **15**(7): p. 1215-23.
108. Suomi, T. and L.L. Elo, *Enhanced differential expression statistics for data-independent acquisition proteomics.* Sci Rep, 2017. **7**(1): p. 5869.
109. Tsou, C.C., et al., *DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics.* Nat Methods, 2015. **12**(3): p. 258-64, 7 p following 264.
110. Demichev, V., et al., *DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput.* Nat Methods, 2020. **17**(1): p. 41-44.
111. Apweiler, R., et al., *UniProt: the Universal Protein knowledgebase.* Nucleic Acids Res, 2004. **32**(Database issue): p. D115-9.
112. Nesvizhskii, A.I. and R. Aebersold, *Interpretation of shotgun proteomic data: the protein inference problem.* Mol Cell Proteomics, 2005. **4**(10): p. 1419-40.
113. Audain, E., et al., *In-depth analysis of protein inference algorithms using multiple search engines and well-defined metrics.* J Proteomics, 2017. **150**: p. 170-182.
114. Sinitcyn, P., et al., *MaxQuant goes Linux.* Nat Methods, 2018. **15**(6): p. 401.
115. Bielow, C., G. Mastrobuoni, and S. Kempa, *Proteomics Quality Control: Quality Control Software for MaxQuant Results.* J Proteome Res, 2016. **15**(3): p. 777-87.
116. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment.* J Proteome Res, 2011. **10**(4): p. 1794-805.

117. 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.
118. Kong, A.T., et al., *MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics.* Nat Methods, 2017. **14**(5): p. 513-520.
119. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-3567.
120. Kim, S. and P.A. Pevzner, *MS-GF+ makes progress towards a universal database search tool for proteomics.* Nat Commun, 2014. **5**: p. 5277.
121. Craig, R. and R.C. Beavis, *A method for reducing the time required to match protein sequences with tandem mass spectra.* Rapid Commun Mass Spectrom, 2003. **17**(20): p. 2310-6.
122. Craig, R. and R.C. Beavis, *TANDEM: matching proteins with tandem mass spectra.* Bioinformatics, 2004. **20**(9): p. 1466-7.
123. Sinitcyn, P., et al., *MaxDIA enables library-based and library-free data-independent acquisition proteomics.* Nat Biotechnol, 2021. **39**(12): p. 1563-1573.
124. MacLean, B., et al., *Skyline: an open source document editor for creating and analyzing targeted proteomics experiments.* Bioinformatics, 2010. **26**(7): p. 966-8.
125. Zhang, F., et al., *A Comparative Analysis of Data Analysis Tools for Data-Independent Acquisition Mass Spectrometry.* Mol Cell Proteomics, 2023. **22**(9): p. 100623.
126. Doncheva, N.T., V. Schwammle, and M. Locard-Paulet, *Understanding Data Analysis Steps in Mass-Spectrometry-Based Proteomics Is Key to Transparent Reporting.* J Proteome Res, 2025. **24**(10): p. 4965-4976.
127. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
128. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
129. Suwala, A.K., et al., *Oligosarcomas, IDH-mutant are distinct and aggressive.* Acta Neuropathol, 2022. **143**(2): p. 263-281.
130. Felix, M., et al., *HIP1R and vimentin immunohistochemistry predict 1p/19q status in IDH-mutant glioma.* Neuro Oncol, 2022. **24**(12): p. 2121-2132.
131. Bader, J.M., et al., *Proteomics separates adult-type diffuse high-grade gliomas in metabolic subgroups independent of 1p/19q codeletion and across IDH mutational status.* Cell Rep Med, 2023. **4**(1): p. 100877.
132. Wong, D., et al., *Integrated proteomic analysis of low-grade gliomas reveals contributions of 1p-19q co-deletion to oligodendroglioma.* Acta Neuropathol Commun, 2022. **10**(1): p. 70.
133. Wadekar, A., et al., *Evaluating the efficacy of Hip1R, Vimentin, and H3K27me3 as surrogate markers for 1p/19q co-deletion in oligodendrogliomas.* Neurooncol Adv, 2025. **7**(1): p. vdaf060.
134. Migliozzi, S., et al., *Integrative multi-omics networks identify PKCdelta and DNA-PK as master kinases of glioblastoma subtypes and guide targeted cancer therapy.* Nat Cancer, 2023. **4**(2): p. 181-202.
135. Garofano, L., et al., *Pathway-based classification of glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities.* Nat Cancer, 2021. **2**(2): p. 141-156.
136. Liu, J., et al., *Multi-scale signaling and tumor evolution in high-grade gliomas.* Cancer Cell, 2024. **42**(7): p. 1217-1238 e19.
137. Zhang, J., et al., *Proteomic profiling of gliomas unveils immune and metabolism-driven subtypes with implications for anti-nucleotide metabolism therapy.* Nat Commun, 2024. **15**(1): p. 10005.
138. Sill, M., et al., *Advancing CNS tumor diagnostics with expanded DNA methylation-based classification.* medRxiv, 2025.
139. Weller, M., et al., *Improved prognostic stratification of patients with isocitrate dehydrogenase-mutant astrocytoma.* Acta Neuropathol, 2024. **147**(1): p. 11.

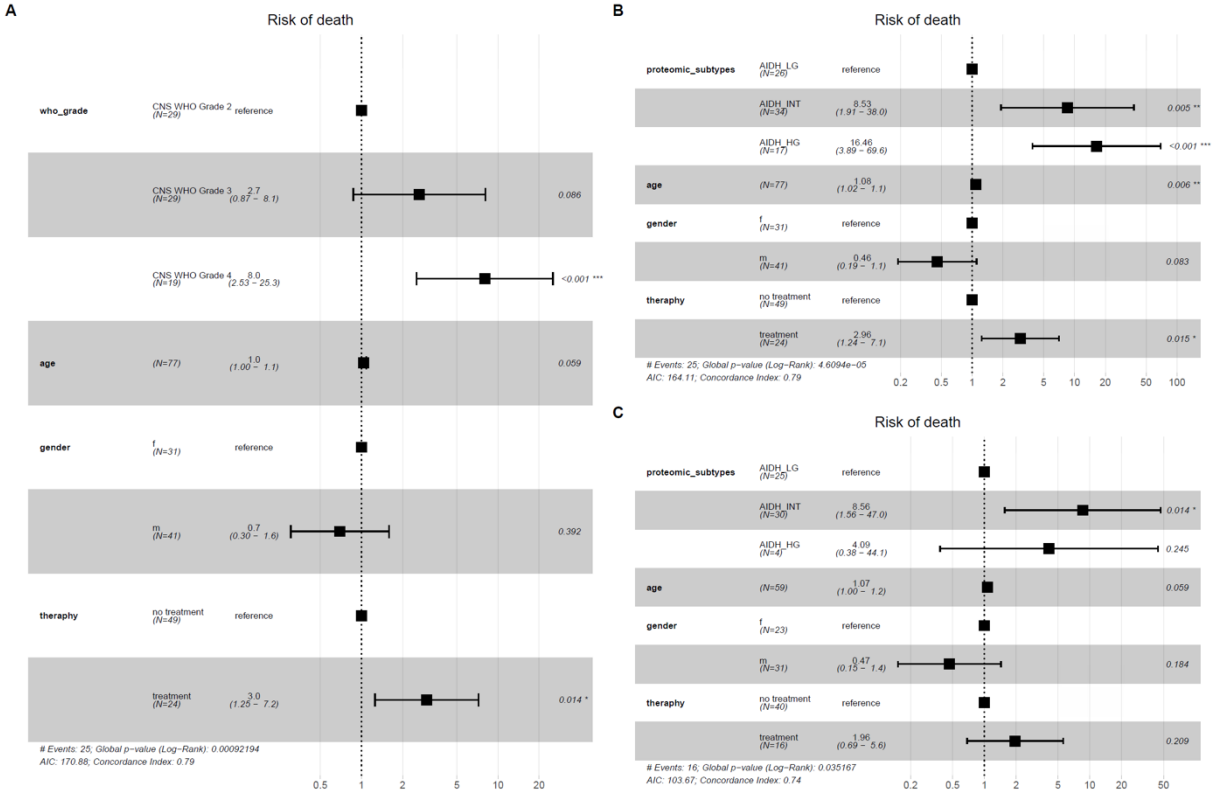
140. Wick, W., et al., *NOA-04 randomized phase III trial of sequential radiochemotherapy of anaplastic glioma with procarbazine, lomustine, and vincristine or temozolomide.* J Clin Oncol, 2009. **27**(35): p. 5874-80.
141. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection.* Cell Syst, 2015. **1**(6): p. 417-425.
142. Dolgalev, I., *MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format*. 2022.
143. Gene Ontology, C., et al., *The Gene Ontology knowledgebase in 2023.* Genetics, 2023. **224**(1).
144. Morgenstern, M., et al., *Quantitative high-confidence human mitochondrial proteome and its dynamics in cellular context.* Cell Metab, 2021. **33**(12): p. 2464-2483 e18.
145. Turei, D., et al., *Integrated intra- and intercellular signaling knowledge for multicellular omics analysis.* Mol Syst Biol, 2021. **17**(3): p. e9923.
146. Aras, F.K., et al., *Expansion of the spectrum of tumors diagnosed as myxopapillary ependymomas.* Acta Neuropathol, 2025. **150**(1): p. 37.
147. 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.
148. Zhu, Y. and T. Guo, *High-Throughput Proteomic Analysis of Fresh-Frozen Biopsy Tissue Samples Using Pressure Cycling Technology Coupled with SWATH Mass Spectrometry.* Methods Mol Biol, 2018. **1788**: p. 279-287.
149. UniProt, C., *UniProt: the universal protein knowledgebase in 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D480-D489.
150. R Developer Core Team, *R: A language and enviroment for statisitcal computing*. 2022, R Foundation for Statisitcal Computing: Vienna, Austria.
151. Martin Morgan, V.O., Jim Hester, Hervé Pagès, *SummarizedExperiment: A container (S4 class) for matrix-like assays*. 2024.
152. Cuklina, J., et al., *Diagnostics and correction of batch effects in large-scale proteomic studies: a tutorial.* Mol Syst Biol, 2021. **17**(8): p. e10240.
153. Hastie T, T.R., Narasimhan B, Chu G . *impute: Imputation for microarray data. R package version 1.80.0.* 2024.
154. Zhang, S., *Nearest neighbor selection for iteratively kNN imputation.* Journal of Systems and Software, 2012. **85**(11): p. 2541-2552.
155. Sclove, S.L., *Using Model Selection Criteria to Choose the Number of Principal Components.* Journal of Statistical Theory and Applications, 2021. **20**(3): p. 450-461.
156. Melville, L.M.a.J.H.a.J., *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.* 2020.
157. Marchette, D.J., *Class Cover Catch Digraphs.* 2022.
158. Michael Antonov, et al., *igraph enables fast and robust network analysis across programming languages.* arXiv preprint arXiv:2311.10260, 2023.
159. Korotkevich G, S.V., Sergushichev A, *Fast gene set enrichment analysis*. 2019.
160. Gu, Z. and D. Hubschmann, *simplifyEnrichment: A Bioconductor Package for Clustering and Visualizing Functional Enrichment Results.* Genomics Proteomics Bioinformatics, 2023. **21**(1): p. 190-202.
161. Hanzelmann, S., R. Castelo, and J. Guinney, *GSVA: gene set variation analysis for microarray and RNA-seq data.* BMC Bioinformatics, 2013. **14**: p. 7.
162. Xu, S., et al., *Using clusterProfiler to characterize multiomics data.* Nat Protoc, 2024. **19**(11): p. 3292-3320.
163. Lang, M., et al., *mlr3: A modern object-oriented machine learning framework in R.* Journal of Open Source Software, 2019. **4**(44).
164. Alvarez, M.J., et al., *Functional characterization of somatic mutations in cancer using network-based inference of protein activity.* Nat Genet, 2016. **48**(8): p. 838-47.
165. Therneau, T. and B. Atkinson, *rpart: Recursive Partitioning and Regression Trees.* 1999.
166. Team, R.C., *R: A language and environment for statistical computing*. 2025, R Foundation for Statistical Computing: Vienna, Austria.

167. Aryee, M.J., et al., *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.* Bioinformatics, 2014. **30**(10): p. 1363-9.
168. Chakravarthy, A., et al., *Pan-cancer deconvolution of tumour composition using DNA methylation.* Nat Commun, 2018. **9**(1): p. 3220.
169. Grabovska, Y., et al., *Pediatric pan-central nervous system tumor analysis of immune-cell infiltration identifies correlates of antitumor immunity.* Nat Commun, 2020. **11**(1): p. 4324.
170. Hovestadt V, Z.M., *conumee: Enhanced copy-number variation analysis using Illumina DNA methylation arrays*.
171. Therneau, T.M., *A Package for Survival Analysis in R*. 2024.
172. Alboukadel Kassambara, M.K., Przemyslaw Biecek, Scheipl Fabian, *survminer: Drawing Survival Curves using 'ggplot2'*. 2024.
173. Harrison, E., T. Drake, and R. Pius, *finalfit: Quickly Create Elegant Regression Results Tables and Plots when Modelling.* 2018.
174. Huber, W., et al., *Orchestrating high-throughput genomic analysis with Bioconductor.* Nat Methods, 2015. **12**(2): p. 115-21.
175. Shapiro, S.S. and M.B. Wilk, *An analysis of variance test for normality (complete samples).* Biometrika, 1965. **52**(3-4): p. 591-611.
176. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society Series B: Statistical Methodology, 1995. **57**(1): p. 289-300.
177. Gu, Z., *circlize implements and enhances circular visualization in R.* Bioinformatics, 2013. **30**(19): p. 2811-2812.
178. Gu, Z., R. Eils, and M. Schlesner, *Complex heatmaps reveal patterns and correlations in multidimensional genomic data.* Bioinformatics, 2016. **32**(18): p. 2847-9.
179. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis*. 2016: Springer-Verlag New York.
180. Kassambara, A., *ggpubr: 'ggplot2' Based Publication Ready Plots.* 2016.
181. Pedersen, T.L., *patchwork: The Composer of Plots*. 2024.
182. Neuwirth, E., *RColorBrewer: ColorBrewer Palettes.* 2002.
183. OpenAI, *ChatGPT Large language model.* . 2025.
184. Apple, *Introducing Apple Intelligence for iPhone, iPad and Mac*. 2024.
185. Halim, Z., et al., *Clustering of graphs using pseudo-guided random walk.* Journal of Computational Science, 2021. **51**.
186. Barthel, F.P., et al., *Longitudinal molecular trajectories of diffuse glioma in adults.* Nature, 2019. **576**(7785): p. 112-120.
187. Hadrup, S., M. Donia, and P. Thor Straten, *Effector CD4 and CD8 T cells and their role in the tumor microenvironment.* Cancer Microenviron, 2013. **6**(2): p. 123-33.
188. Faienza, F., A. Rasola, and G. Filomeni, *Nitric oxide-based regulation of metabolism: Hints from TRAP1 and SIRT3 crosstalk.* Front Mol Biosci, 2022. **9**: p. 942729.
189. Yang, J., et al., *Guidelines and definitions for research on epithelial-mesenchymal transition.* Nat Rev Mol Cell Biol, 2020. **21**(6): p. 341-352.
190. Latifkar, A., et al., *IGF2BP2 promotes cancer progression by degrading the RNA transcript encoding a v-ATPase subunit.* Proc Natl Acad Sci U S A, 2022. **119**(45): p. e2200477119.
191. Mancarella, C. and K. Scotlandi, *IGF2BP3 From Physiology to Cancer: Novel Discoveries, Unsolved Issues, and Future Perspectives.* Front Cell Dev Biol, 2019. **7**: p. 363.
192. Xue, G., et al., *Akt/PKB-mediated phosphorylation of Twist1 promotes tumor metastasis via mediating cross-talk between PI3K/Akt and TGF-beta signaling axes.* Cancer Discov, 2012. **2**(3): p. 248-59.
193. Huang, Q., et al., *Multi-omics analysis reveals NNMT as a master metabolic regulator of metastasis in esophageal squamous cell carcinoma.* NPJ Precis Oncol, 2024. **8**(1): p. 24.
194. McShane, E. and L.S. Churchman, *Central dogma rates in human mitochondria.* Hum Mol Genet, 2024. **33**(R1): p. R34-R41.
195. Leone, G., et al., *The Oncojanus Paradigm of Respiratory Complex I.* Genes (Basel), 2018. **9**(5).

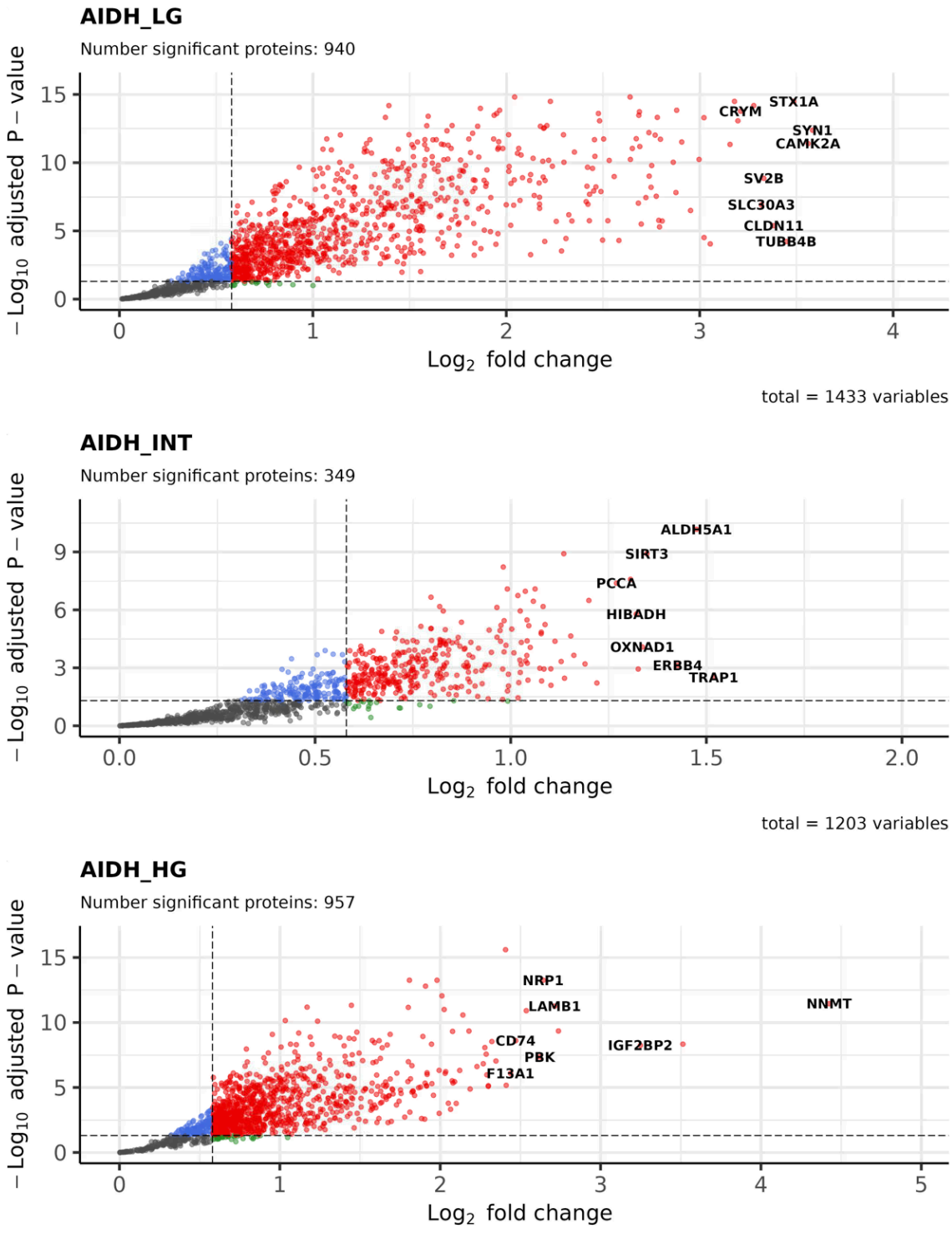
196. Lacroix, E., et al., *PI3K/AKT signaling mediates stress-inducible amyloid formation through c-Myc.* Cell Rep, 2025. **44**(5): p. 115617.
197. Badia, I.M.P., et al., *decoupleR: ensemble of computational methods to infer biological activities from omics data.* Bioinform Adv, 2022. **2**(1): p. vbac016.
198. van den Bent, M.J., et al., *The biological significance of tumor grade, age, enhancement, and extent of resection in IDH-mutant gliomas: How should they inform treatment decisions in the era of IDH inhibitors?* Neuro Oncol, 2024. **26**(10): p. 1805-1822.
199. Blobner, J., et al., *Clinical and neuropathological criteria for distinguishing between IDH-mutant astrocytomas of WHO grade 2 and 3.* J Neurooncol, 2025. **175**(2): p. 763-774.
200. Yoda, R.A., et al., *Mitotic Index Thresholds Do Not Predict Clinical Outcome for IDH-Mutant Astrocytoma.* J Neuropathol Exp Neurol, 2019. **78**(11): p. 1002-1010.
201. Richardson, T.E., et al., *Chromosomal instability in adult-type diffuse gliomas.* Acta Neuropathol Commun, 2022. **10**(1): p. 115.
202. Coons, S.W. and D.K. Pearl, *Mitosis identification in diffuse gliomas: implications for tumor grading.* Cancer, 1998. **82**(8): p. 1550-5.
203. Venkataramani, V., et al., *Synaptic input to brain tumors: clinical implications.* Neuro Oncol, 2021. **23**(1): p. 23-33.
204. Duhamel, M., et al., *Spatial analysis of the glioblastoma proteome reveals specific molecular signatures and markers of survival.* Nat Commun, 2022. **13**(1): p. 6665.
205. Tang, J., et al., *Protein-based classification reveals an immune-hot subtype in IDH mutant astrocytoma with worse prognosis.* Cancer Cell, 2025.
206. Wang, Y., et al., *Proteogenomics of diffuse gliomas reveal molecular subtypes associated with specific therapeutic targets and immune-evasion mechanisms.* Nat Commun, 2023. **14**(1): p. 505.
207. Karnati, H.K., et al., *Down regulated expression of Claudin-1 and Claudin-5 and up regulation of beta-catenin: association with human glioma progression.* CNS Neurol Disord Drug Targets, 2014. **13**(8): p. 1413-26.
208. Agarwal, R., et al., *Silencing of claudin-11 is associated with increased invasiveness of gastric cancer cells.* PLoS One, 2009. **4**(11): p. e8002.
209. Bi, G., et al., *PRRT2 inhibits the proliferation of glioma cells by modulating unfolded protein response pathway.* Biochem Biophys Res Commun, 2017. **485**(2): p. 454-460.
210. Li, X. and C. Zhang, *STXBP1 inhibits glioma progression by modulating ferroptosis and epithelial-mesenchymal transition.* Archives of Medical Science, 2025.
211. Chen, Y., et al., *Sirtuin-3 (SIRT3), a therapeutic target with oncogenic and tumor-suppressive function in cancer.* Cell Death Dis, 2014. **5**(2): p. e1047.
212. Laquatra, C., et al., *HIF1alpha-dependent induction of the mitochondrial chaperone TRAP1 regulates bioenergetic adaptations to hypoxia.* Cell Death Dis, 2021. **12**(5): p. 434.
213. Yoshida, S., et al., *Molecular chaperone TRAP1 regulates a metabolic switch between mitochondrial respiration and aerobic glycolysis.* Proc Natl Acad Sci U S A, 2013. **110**(17): p. E1604-12.
214. Sciacovelli, M., et al., *The mitochondrial chaperone TRAP1 promotes neoplastic growth by inhibiting succinate dehydrogenase.* Cell Metab, 2013. **17**(6): p. 988-999.
215. Verhaak, R.G., et al., *Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1.* Cancer Cell, 2010. **17**(1): p. 98-110.
216. Fedele, M., et al., *Proneural-Mesenchymal Transition: Phenotypic Plasticity to Acquire Multitherapy Resistance in Glioblastoma.* Int J Mol Sci, 2019. **20**(11).
217. Ulanovskaya, O.A., A.M. Zuhl, and B.F. Cravatt, *NNMT promotes epigenetic remodeling in cancer by creating a metabolic methylation sink.* Nat Chem Biol, 2013. **9**(5): p. 300-6.
218. Morrish, F. and D. Hockenbery, *MYC and mitochondrial biogenesis.* Cold Spring Harb Perspect Med, 2014. **4**(5).
219. Kim, M., et al., *Mitochondrial DNA is a major source of driver mutations in cancer.* Trends Cancer, 2022. **8**(12): p. 1046-1059.

220. Sauerbrei, W., et al., *Doug Altman: Driving critical appraisal and improvements in the quality of methodological and medical research.* Biometrical Journal, 2020. **63**(2): p. 226-246.
221. Altman, D.G. and P. Royston, *The cost of dichotomising continuous variables.* Bmj, 2006. **332**(7549).
222. Ghose, A. and B. Ravindran, *Interpretability With Accurate Small Models.* Front Artif Intell, 2020. **3**: p. 3.
223. Kokash, N. and L. Makhnist, *Using Decision Trees for Interpretable Supervised Clustering.* SN Computer Science, 2024. **5**(2).
224. Esposito, F., et al., *Wee1 Kinase: A Potential Target to Overcome Tumor Resistance to Therapy.* Int J Mol Sci, 2021. **22**(19).
225. Hu, C., et al., *ROCK1 promotes migration and invasion of non-small-cell lung cancer cells through the PTEN/PI3K/FAK pathway.* Int J Oncol, 2019. **55**(4): p. 833-844.
226. Tian, J., et al., *Aurora kinase A inhibition plus Tumor Treating Fields suppress glioma cell proliferation in a cilium-independent manner.* Transl Oncol, 2024. **45**: p. 101956.
227. Consortium, G., *Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the Glioma Longitudinal Analysis Consortium.* Neuro Oncol, 2018. **20**(7): p. 873-884.

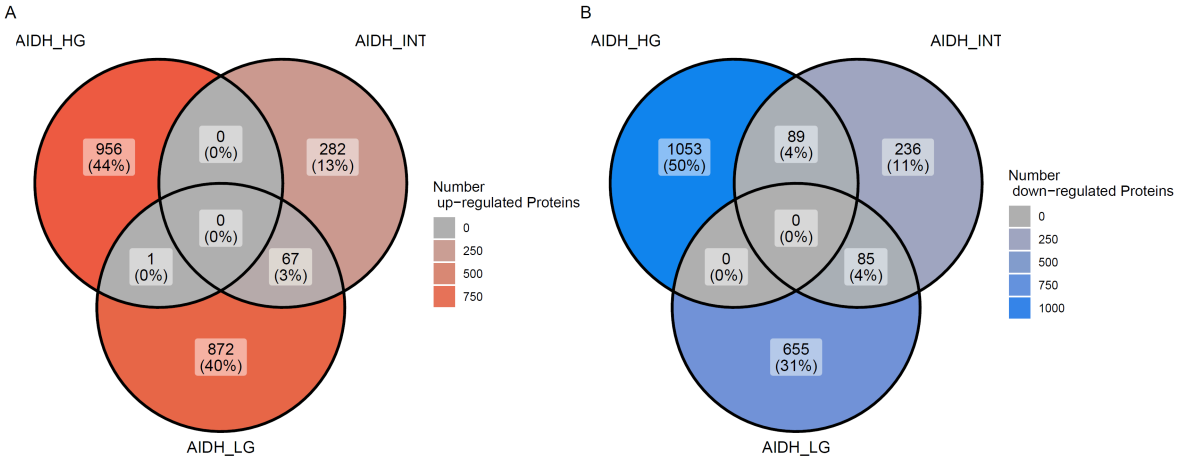
Appendix I: Supplementary figures



Supplemental Figure 1: Cox proportional regression shows superior performance of proteomic subtyping to CNS5 WHO Grading. Cox proportional hazard model of A) WHO grades and B) proteomic subtype in presence of WHO grade 4 and C) their absence. Cox proportional hazard model were adjusted by patient age, gender and therapy modality. The plots indicate that proteomic subtypes perform better in patient stratification than WHO grading. Abbreviations: AIDH_HG, Astrocytoma IDH mutant high-grade; AIDH_INT, Astrocytoma, IDH mutant intermediate; AIDH_LG, Astrocytoma IDH mutant low-grade;



Supplemental Figure 2: Proteomics subtypes show distinct expression programs with strong upregulation of proteins linked to brain function, mitochondrial metabolism and proliferation.
Volcano plots depicting the number and distribution of significant upregulated proteins determined in the context of proteomic subtypes A) AIHD_LG, B) AIDH_ INT and C) AIDH_HG. The proteins with the highest positive log_2-fold (logFC) are highlighted in text boxes. The average expression of a subtype was tested against the average expression of the remaining subtypes and p-values were obatined using a linear regression model to estimate Empirical Bayesian adjusted moderated t-statistic. Thresholds for proteins of interests were determined by using an adjusted p-value threshold of 0.05 and a logFC change threshold of 0.58. Proteins are colorized according to threshold criteria; not significant (NS) = grey, above the log-fold change threshold (Log2- FC) = green, below the p-value threshold (p-value) = blue; proteins of interest (p-value and Log2-FC) = red.



Supplemental Figure 3: Number of identified DEP between proteomic subtypes.
Venn diagrams depicting the number of A) upregulated and B) downregulated DEP determined in each proteomic subtype. Thresholds for proteins of interests were determined by using an adjusted p-value threshold of 0.05 and a logFC change threshold of 0.58. A colour code is given for the number of determined proteins with red being indicative for upregulated and blue being indicative for downregulated proteins. Abbreviations

Appendix II: Supplementary tables

Supplemental Table 1.: Top upregulated proteins AIDH_INT.
Table representation showing the log2Fold change and adjusted P-value of the top 9 upregulated proteins found in AIDH_INT. The log2Fold change and adjusted P-value of the other two groups are shown as comparison to show similarities. Differential expressed proteins characteristic for each subtype were ascertained by employing a global contrast. Here, each proteomic subtype was compared against the averaged expression of the remaining two subtypes. The top proteins were identified by using the candidates with the highest log2Fold and smallest adjusted P-value.

Gene	Log_2FC AIDH_HG	Log_2FC AIDH_INT	Log_2FC AIDH_LG	P.adj AIDH_HG	P.adj AIDH_INT	P.adj AIDH_LG
ALDH5A1	-2.426	1.475	0.952	>0.001***	>0.001***	>0.001***
ERBB4	-3.182	1.427	1.755	>0.001***	0.001	>0.001***
HIBADH	-1.498	1.321	0.178	>0.001***	>0.001***	0.516
OXNAD1	-1.476	1.336	0.140	>0.001***	>0.001***	0.692
PCCA	-1.694	1.270	0.424	>0.001***	>0.001***	0.044
PCCB	-1.668	1.306	0.362	>0.001***	>0.001***	0.090
SIRT3	-1.624	1.348	0.276	>0.001***	>0.001***	0.169
TNR	-3.884	1.326	2.558	>0.001***	>0.001***	>0.001***
TRAP1	-0.943	1.519	-0.576	0.089	0.003	0.258

Supplemental Table 2.: Top upregulated proteins AIDH_LG.
Table representation showing the log2Fold change and adjusted P-value of the top 10 upregulated proteins found in AIDH_LG. The log2Fold change and adjusted P-value of the other two groups are shown as comparison to show similarities. Differential expressed proteins characteristic for each subtype were ascertained by employing a global contrast. Here, each proteomic subtype was compared against the averaged expression of the remaining two subtypes. The top proteins were identified by using the candidates with the highest log2Fold and smallest adjusted P-value.

Gene	Log_2FC AIDH_HG	Log_2FC AIDH_INT	Log_2FC AIDH_LG	P.adj AIDH_HG	P.adj AIDH_INT	P.adj AIDH_LG
CAMK2A	-2.916	-0.644	3.560	>0.001***	0.198	>0.001***
CLDN11	-2.542	-0.841	3.383	0.001	0.291	>0.001***
CRYM	-2.898	-0.313	3.212	>0.001***	0.446	>0.001***
PRRT2	-2.466	-0.813	3.279	>0.001***	0.026	>0.001***
SLC30A3	-1.807	-1.512	3.319	0.007	0.015	>0.001***
SNCB	-2.296	-0.901	3.197	>0.001***	0.019	>0.001***
STX1A	-2.758	-0.728	3.486	>0.001***	0.058	>0.001***
SV2B	-2.649	-0.682	3.331	>0.001***	0.227	>0.001***
SYN1	-2.664	-0.919	3.584	>0.001***	0.042	>0.001***
TUBB4B	-3.777	0.329	3.448	>0.001***	0.764	>0.001***

Supplemental Table 3.: Top upregulated proteins AIDH_HG.
Table representation showing the log2Fold change and adjusted P-value of the top 10 upregulated proteins found in AIDH_HG. The log2Fold change and adjusted P-value of the other two groups are shown as comparison to show similarities. Differential expressed proteins characteristic for each subtype were ascertained by employing a global contrast. Here, each proteomic subtype was compared against the averaged expression of the remaining two subtypes. The top proteins were identified by using the candidates with the highest log2Fold and smallest adjusted P-value.

Gene	Log_2FC AIDH_HG	Log_2FC AIDH_INT	Log_2FC AIDH_LG	P.adj AIDH_HG	P.adj AIDH_INT	P.adj AIDH_LG
CD74	2.471	-1.259	-1.212	>0.001***	>0.001***	>0.001***
F13A1	2.439	-0.810	-1.629	>0.001***	0.070	>0.001***
LAMB1	2.714	-1.265	-1.449	>0.001***	>0.001***	>0.001***
NRP1	2.643	-1.485	-1.157	>0.001***	>0.001***	>0.001***
TMT1B	2.737	-1.634	-1.103	>0.001***	>0.001***	0.002
CD163	2.537	-1.012	-1.524	>0.001***	0.001	>0.001***
IGF2BP2	3.248	-1.543	-1.706	>0.001***	0.001	>0.001***
IGF2BP3	3.513	-1.801	-1.713	>0.001***	0.001	0.001
NNMT	4.425	-2.450	-1.975	>0.001***	>0.001***	>0.001***
PBK	2.621	-1.120	-1.501	>0.001***	0.007	>0.001***

Appendix III: List of publications

Publications related to this work (in preparation):

1. **Friedel D**, Sigismondo G, Schrimpf D., von Deimling A, David E. Reuss. Proteomic analysis of IDH mutant Astrocytoma reveals three novel subtypes linked with malignant progression

Publications not directly related to this work:

1. Aras Fuat Kaan, **Friedel Dennis,** Keller Felix**,** Zettl Ferdinand**,** Banan Rouzbeh**,** Sievers Philipp**,** Suwala Abigail K**,** Hinz Felix**,** Friedrich Lukas**,** Abdulrazak Ivan**,** Esmaeilibenvidi Mozhgan**,** Etminan Nima**,** Herold-Mende Christel**,** Wick Wolfgang**,** Krieg Sandro**,** Pfister Stefan M, Korshunov Andrey, Bludau Isabell, Sahm Felix, Reuss David E, Sigismondo, Gianluca, von Deimling Andreas. **Acta Neuropathol**. 2025 Sep 30;150(1):37. doi: 10.1007/s00401-025-02944-w.

2. Suwala AK, Felix M, **Friedel D**, Stichel D, Schrimpf D, Hinz F, Hewer E, Schweizer L, Dohmen H, Pohl U, Staszewski O, Korshunov A, Stein M, Wongsurawat T, Cheunsuacchon P, Sathornsumetee S, Koelsche C, Turner C, Le Rhun E, Mühlebner A, Schucht P, Özduman K, Ono T, Shimizu H, Prinz M, Acker T, Herold-Mende C, Kessler T, Wick W, Capper D, Wesseling P, Sahm F, von Deimling A, Hartmann C, Reuss DE. Oligosarcomas, IDH-mutant are distinct and aggressive. **Acta Neuropathol**. 2022 Feb;143(2):263-281. doi: 10.1007/s00401-021-02395-z.
3. Panitz V*, Končarević S*, Sadik A*, **Friedel D**, Bausbacher T, Trump S, Farztdinov V, Schulz S, Sievers P, Schmidt S, Jürgenson I, Jung S, Kuhn K, Pflüger I, Sharma S, Wick A, Pfänder P, Selzer S, Vollmuth P, Sahm F, von Deimling A, Heiland I, Hopf C, Schulz-Knappe P, Pike I, Platten M, Wick W, Opitz CA. Tryptophan metabolism is inversely regulated in the tumor and blood of patients with glioblastoma. **Theranostics**. 2021 doi: 10.7150/thno.60679. (* equal contributing authors)
4. Franziska M Ippen*, Thomas Hielscher, **Dennis Friedel**, Kirsten Göbel, David Reuss, Christel Herold-Mende, Sandro Krieg, Andreas V Deimling, Wolfgang Wick, Felix Sahm, Abigail K Suwala. The prognostic impact of CDKN2A/B hemizygous deletions in IDH-mutant glioma. **Neuro Oncol**. 2025 Mar 7;27(3):743-754. doi: 10.1093/neuonc/noae238.
5. Felix Hinz, **Dennis Friedel**, Andrey Korshunov, Franziska M Ippen, Henri Bogumil, Rouzbeh Banan, Sebastian Brandner, Martin Hasselblatt , Henning B Boldt, Vaidas Dirse , Hildegard Dohmen , Eleonora Aronica, Michael Brodhun, Marike L D Broekman, David Capper, Asan Cherkezov , Maximilian Y Deng, Vera van Dis, Jörg Felsberg, Stephan Frank, Pim J French, Rüdiger Gerlach , Kirsten Göbel, Eric Goold, Jürgen Hench, Sven Kantelhardt, Patricia Kohlhof-Meinecke, Sandro Krieg, Christian Mawrin, Gillian Morrison, Angelika Mühlebner, Koray Ozduman, Stefan M Pfister, Pietro Luigi Poliani, Marco Prinz, Guido Reifenberger, Markus J Riemenschneider, Roman Sankowski, Daniel Schrimpf, Martin Sill, Matija Snuderl, Robert M Verdijk, Mathew R Voisin, Pieter Wesseling, Wolfgang Wick, David E Reuss, Andreas von Deimling, Felix Sahm, Sybren L N Maas, Abigail K Suwala. IDH-mutant astrocytomas with primitive neuronal component have a distinct methylation profile and a higher risk of leptomeningeal spread. **Acta Neuropathol.** 2025 Feb 3;149(1):12. doi: 10.1007/s00401-025-02849-8.
6. Ling Hai, **Dennis Friedel**, Felix Hinz , Dirk C Hoffmann, Sofia Doubrovinskaia , Hannah Rohdjess, Katharina Weidenauer , Evgeniya Denisova , Georg T Scheffler , Tobias Kessler, Alexandros Kourtesakis, Christel Herold-Mende , Octavian Henegariu , Joachim M Baehring, Jorg Dietrich , Benedikt Brors , Wolfgang Wick, Felix Sahm, Leon D Kaulen. Distinct epigenetic and transcriptional profiles of Epstein-Barr virus-positive and negative primary CNS lymphomas. **Neuro Oncol.** 2025 May 15;27(4):979-992. doi: 10.1093/neuonc/noae251.
7. Leon D Kaulen, Evgeniya Denisova, Felix Hinz, Ling Hai, **Dennis Friedel**, Octavian Henegariu, Dirk C Hoffmann, Jakob Ito, Alexandros Kourtesakis , Pascal Lehnert, Sofia Doubrovinskaia , Philipp Karschnia, Louisa von Baumgarten, Tobias Kessler, Joachim M Baehring, Benedikt Brors , Felix Sahm, Wolfgang Wick. Integrated genetic analyses of immunodeficiency-associated Epstein-Barr virus- (EBV) positive primary CNS lymphomas. **Acta Neuropathol**. 2023 Sep;146(3):499-514. doi: 10.1007/s00401-023-02613-w. Epub 2023 Jul 26.
8. Felix M, **Friedel D**, Jayavelu AK, Filipski K, Reinhardt A, Warnken U, Stichel D, Schrimpf D, Korshunov A, Wang Y, Kessler T, Etminan N, Unterberg A, Herold-Mende C, Heikaus L, Sahm F, Wick W, Harter PN,

von Deimling A, Reuss DE. HIP1R and vimentin immunohistochemistry predict 1p/19q status in IDH-mutant glioma. **Neuro Oncol**. 2022 Dec 1;24(12):2121-2132. doi: 10.1093/neuonc/noac111.

Grants related to this work:

1. Sabrina Kaiser Award for Astrocytoma Research. Developing inhibitors for IDH-mutant astrocytoma with worse therapy response driven by Interleukin-4 Induced 1 (IL4I1)

Poster presentations:

1. Friedel D. *et al.;* DKFZ Phd Poster Presentation, Heidelberg, Germany (2024) (awarded with Poster Price)
2. Friedel D *et al.;* DIASym, MASS SPECTROMETRY MEETS SYSTEM MEDICINE Summer School, Mainz, Germany (2025)