

Expanding the proteome: Advancing proteomics methodologies to uncover new insights into cancers

Kenneth Weke



**Expanding the proteome: Advancing proteomics methodologies to
uncover new insights into cancers**

Kenneth Weke

A thesis submitted to the Intercollegiate Faculty of Biotechnology of UG&MUG in full
conformity with the requirements for the degree of Doctor of Philosophy (chemical biology),
University of Gdańsk

The research work presented in this doctoral thesis was completed at the International
Centre for Cancer Vaccine Science, University of Gdańsk

Supervised by Prof. **Theodore Hupp**

Co-supervised by Dr. **Sachin Kote**

© Copyright by Kenneth Weke 2023

Dedication

To patients, in life or memory, whose tissue donations form the proteome story in this thesis

Acknowledgment

Finally, at the tail end! :)

This journey has been a rollercoaster of challenges, triumphs, and, most importantly, incredible growth. My deepest appreciation in general to all who have played a part in my academic journey, and specifically to individuals whose unwavering support and guidance have been instrumental in the completion of this doctoral thesis. To my supervisor, Prof. Ted Hupp, thanks for your support and guidance. You challenged my ideas and helped refine them through constructive feedback, ultimately influencing my growth as a researcher. Dr Sachin Kote, you are a dedicated co-supervisor. You stepped in just when things seemed to go south, thanks a bunch. To past and current clinical peptidomics group members, you are terrific. The collaboration and camaraderie were real. I am thankful. To Dr. Jakub Faktor, you are selfless. Thank you so much. To Kasia, another selfless soul. Your resilience and optimism were unmatched. The fights with that chamber were swamping. I am truly grateful. To Dr. Javier Alfaro, what can I say, bro? You perfectly played the role of a big brother, starting at the coast of the Great Lakes and now at the Baltic coast, where you did the most. I am immensely thankful. To Dr. Irena Dapic, thank you for believing in me and agreeing to be my guiding light at the commencement of this journey. To Prof. Dave Goodlett, thank you for your support and mentorship. To the ICCVS fraternity, thank you. You are amazing. Special thanks to my co-authors for their collaborative efforts. To Prof. Angela Amayo, Prof. Christine Kigonde, Dr. Helen Dimaras, and Dr. Paul Boutros, thank you for shaping my academic foundation. To Dr. Alexandre Rouette, thanks for the mentorship and support. To Vern, thanks a lot, bro, for the encouragement and always informing me of available opportunities. To Beatko, you are truly awesome. Thanks for the immense support and encouragement. A big thanks to my family, whose unwavering support sustained me throughout.

Abstract

The dynamic interconnection between proteomics and cancer biology necessitates sensitive and reproducible methodologies to decode intricate molecular mechanisms, especially in clinically relevant scenarios. Mass spectrometry (MS)-based proteomics is a powerful analytical instrument for profiling and quantifying proteins in biological samples. This thesis focuses on developing and broadening the application of MS-based proteomics technologies to advance the understanding of cancer biology in order to improve the diagnosis and treatment of cancer. The first part of the thesis describes the application of a microscale proteomics method, microdroplet processing in one pot for trace samples (microPOTS), which can be used to identify and quantify proteins from small numbers of cells. microPOTS was applied to identify proteomic changes in Barrett's esophageal cells following physiologic and radiation stress exposure. From a small population of cells, microPOTS demonstrated relatively high proteome coverage with the identification of >1500 protein groups and achieved high quantitative reproducibility.

The second part of the thesis focuses on developing and applying a data-independent acquisition (DIA) MS proteomics workflow to identify and quantify proteins from formalin-fixed paraffin-embedded (FFPE) tissues. This workflow was applied to glioblastoma (GBM) FFPE tissue microdissections, and more than 1700 proteins were detected, and over 1400 proteins were quantified. GBM-relevant proteins (e.g., GFAP, FN1, VIM, and MBP) were quantified with high precision (median coefficient of variation <12%). In addition, immune-related proteins (e.g., ILF2, MIF, and CD38) were consistently detected and quantified. The strategy holds great potential for routinizing protein quantification in FFPE tissue samples.

The third part of the thesis focuses on investigating the impact of hypoxia on antigen presentation in GBM using an integrated approach combining MS-based proteomics and immunopeptidomics strategies. Hypoxic stress induces significant changes in the GBM proteome. Key enzymes (ERAP1 and ERAP2) in the antigen processing and presentation

machinery (APM) were downregulated in hypoxia. Moreover, the findings revealed that HLA-I-associated antigen peptide repertoire was reduced under hypoxia. Overall, these results open new horizons to further explore the interplay between hypoxia and antigen presentation.

Collectively, this thesis presents a series of studies that have advanced the field of proteomics and its application to cancer biology, especially GBM tumors. The development of new proteomics technologies and their application to cancer research has the potential to improve the understanding of cancer biology and to develop new diagnostic and immunotherapeutic strategies.

Streszczenie

Dynamiczność korelacji proteomiki i biologii nowotworów wymaga wrażliwych i powtarzalnych metod dekodowania złożonych mechanizmów molekularnych, szczególnie w scenariuszach klinicznie istotnych. Proteomika oparta na spektrometrii mas (MS) jest potężnym narzędziem analitycznym do profilowania i kwantyfikacji białek w próbkach biologicznych. Niniejsza praca doktorska skupia się na rozwijaniu i poszerzaniu zastosowania technologii proteomiki opartej na MS w celu pogłębienia zrozumienia biologii nowotworów i dalszego udoskonalenia diagnostyki i leczenia raka.

Pierwsza część pracy opisuje zastosowanie metody proteomiki mikroskalowej, przetwarzania mikrokropelek w jednym naczyniu dla próbek śladowych (microPOTS), która może być użyta do identyfikacji i kwantyfikacji białek z małej liczby komórek. Metoda microPOTS została zastosowana do identyfikacji zmian proteomicznych w komórkach przelyku Barretta po ekspozycji na stres fizjologiczny i radiacyjny. Metoda microPOTS osiągnęła stosunkowo wysokie pokrycie proteomu dla niewielkiej populacji komórek, identyfikując więcej niż 1500 grup białkowych i uzyskując wysoką powtarzalność kwantyfikacji.

Druga część pracy koncentruje się na rozwijaniu i stosowaniu toku pracy w proteomice MS opartym na pozyskiwaniu niezależnym od danych (DIA) do identyfikacji i określania ilościowego białek z tkanek utrwalonych w formalinie i zatopionych w parafinie (FFPE). Metodyka ta została zastosowana do mikrodysekcji tkanki glejaka wielopostaciowego (GBM) FFPE, gdzie wykryto ponad 1700 białek, a ponad 1400 białek zostało skwantyfikowanych. Białka istotne dla GBM (np. GFAP, FN1, VIM i MBP) zostały określone ilościowo z dużą precyzją (średni współczynnik zmienności <12%). Ponadto, białka powiązane z odpornością (np. ILF2, MIF i CD38) były konsekwentnie wykrywane i kwantyfikowane. Strategia ta kryje w sobie wielki potencjał ustandaryzowania szacowania białek w próbkach tkanek FFPE.

Trzecia część pracy przedstawia badanie wpływu niedotlenienia na prezentację antygenów w GBM, wykorzystując zintegrowane podejście łączące proteomikę opartą na MS i strategię pochodzącą z immunopeptydomiki. Stres hipoksyczny wywołuje znaczące zmiany w proteomie GBM. Kluczowe dla aparatu przetwarzania i prezentacji antygenów (APM) enzymy (ERAP1 i ERAP2) miały zmniejszoną ekspresję podczas stanu niedotlenienia. Ponadto, wyniki ujawniły, że zasób peptydów antygenowych związanych z HLA-I był zredukowany w analogicznych warunkach. Wyniki te ukazują nowe perspektywy dla dalszego badania współzależności między niedotlenieniem i prezentacją antygenów.

Podsumowując, niniejsza praca doktorska prezentuje serię badań, które posunęły naprzód dziedzinę proteomiki i jej zastosowanie w biologii nowotworów, ze szczególnym uwzględnieniem guzów GBM. Rozwój nowych technologii proteomiki i ich zastosowanie w badaniach nad rakiem ma potencjał poszerzenia zrozumienia biologii nowotworów oraz rozwoju nowych strategii diagnostycznych i terapeutycznych w immunologii.

Abbreviations

| | |
|----------|---|
| ABC | Ammonium bicarbonate |
| ACN | Acetonitrile |
| AGC | Automatic gain control |
| AGR2 | Anterior Gradient 2 |
| APC | Antigen presenting cells |
| APM | Antigen processing and presentation machinery |
| ARNT | Aryl hydrocarbon receptor nuclear translocator |
| ATRX | Alpha thalassemia/mental retardation syndrome X-linked |
| BASP1 | Brain acid soluble protein 1 |
| BPE | Bovine pituitary extract |
| BRAF | v-raf murine sarcoma viral oncogene homolog B1 |
| CAR | Chimeric antigen receptor |
| CDKN2A/B | Cyclin-dependent kinase inhibitor 2A/B |
| CI | Chemical ionization |
| CNS | Central nervous system |
| CPA4 | Carboxypeptidase A4 |
| CPTAC | Clinical Proteomic Tumor Analysis Consortium |
| CTLA-4 | Cytotoxic T-lymphocyte-associated protein 4 |
| CV | Coefficient of variation |
| DDA | Data-dependent acquisition |
| DDM | N-Dodecyl β -d-maltose |
| DEP | Differentially enrichment analysis |
| DEqMS | Differential expression analysis of quantitative mass spectrometry data |
| DIA | Data-independent acquisition |
| DIA-NN | Data-Independent Acquisition by Neural Networks |
| DNA | Deoxyribonucleic acid |
| DT | Direct trypsinization |
| DTT | Dithiothreitol |
| EGFR | Epidermal growth factor receptor |
| EGFR | Epidermal growth factor receptor |
| EI | Electron ionization |
| ER | Endoplasmic reticulum |
| ERAP | Endoplasmic reticulum aminopeptidase |
| ESI | Electrospray ionization |
| FA | Formic acid |
| FACS | Fluorescence-activated cell sorting |
| FASP | Filter-aided sample preparation |
| FDR | False discovery rate |

| | |
|---------------|--|
| FF | Freshly frozen |
| FLNA | Filamin-A protein |
| FN1 | Fibronectin 1 |
| FT-ICR | Fourier-Transform Ion-Cyclotron Resonance |
| GBM | Glioblastoma |
| GO | Gene ontology |
| GRAVY | Grand average of hydropathy |
| GTR | Gross total resection |
| HAIR | Heat-induced antigen retrieval |
| HCD | High energy dissociation |
| HIF | Hypoxia-inducible factor |
| HIF1 α | Hypoxia-inducible factor 1 α |
| HLA | Human leukocyte antigen |
| HRAM | High-resolution accurate-mass |
| IAA | Iodoacetamide |
| ICAT | Isotope-coded affinity tag |
| ICI | Immune checkpoint inhibitors |
| IDH | Isocitrate dehydrogenase |
| IDT | Integrated DNA Technologies |
| ILF2 | Interleukin enhancer-binding factor 2 |
| ISD | In-solution digestion |
| iTRAQ | Isobaric tags for relative and absolute quantification |
| LC-MS | Liquid chromatography–mass spectrometry |
| LCA | Lithocholic acid |
| LIT/FT | Linear ion Trap/Fourier-Transform |
| LOH | Loss of heterozygosity |
| MALDI | Matrix-assisted laser desorption/ionization |
| MAPK | Mitogen-activated protein kinase |
| MBP | Myelin basic protein |
| MGMT | O ⁶ -methylguanine DNA methyltransferase |
| MHC | Major histocompatibility complex |
| microPOTS | Microdroplet processing in one pot for trace samples |
| MIF | Macrophage migration inhibitory factor |
| MS | Mass spectrometry |
| MW | Molecular weight |
| nanoPOTS | Nanodroplet processing in one pot for trace samples |
| NK | Natural killer |
| OS | Overall survival |
| PCA | Principal component analysis |

| | |
|---------|--|
| PCR | Polymerase chain reaction |
| PD-1 | Programmed cell death protein 1 |
| PFS | Progression free survival |
| PHD | Prolyl hydroxylases |
| PPI | Protein-protein interaction |
| PRM | Parallel-reaction monitoring |
| PSM | Peptide-to-spectrum matches |
| PTEN | Phosphatase and tensin homolog |
| PTM | Post-translational modifications |
| Q/ToF | Quadrupole/Time-of-Flight |
| QqQ | Triple quadrupoles |
| QTRAP | Quadrupole ion trap |
| RCPR | Rare cell proteomic reactor |
| rEGF | Recombinant epidermal growth factor |
| RNA | Ribonucleic acid |
| SCX | Strong cation exchange |
| SILAC | Stable isotope labelling with amino acids in cel culture |
| SISPROT | Simple and integrated spintip-based proteomics technology |
| SOPs-MS | Surfactant-assisted one-pot sample processing at the standard volume coupled with MS |
| SP3 | Single-pot, solid phase-enhanced sample preparation |
| SPE | Solid-phase extraction |
| SRM/MRM | Selected/multiple-reaction monitoring |
| STR | Subtotal resection |
| TAM | Tumor-associated macrophages |
| TAP | Transport-associated with antigen processing |
| TERT | Telomerase reverse transcriptase |
| TFA | Trifluoroacetic acid |
| TME | Tumor microenvironment |
| TMT | Tandem mass tag |
| TMZ | Temozolomide |
| ToF | Time-of-Flight |
| VHL | von Hippel-Lindau |
| VIM | Vimentin |
| VSN | Variance stabilizing normalization |
| WHO | World health organization |
| XIC | Extracted ion chromatogram |

List of papers

The following original articles have been included in this doctoral thesis

1. **Weke, K.**, Singh, A., Uwugiaren, N., Alfaro, J. A., Wang, T., Hupp, T. R., O'Neill, J. R., Vojtesek, B., Goodlett, D. R., Williams, S. M., Zhou, M., Kelly, R. T., Zhu, Y., and Dapic, I. MicroPOTS Analysis of Barrett's Esophageal Cell Line Models Identifies Proteomic Changes after Physiologic and Radiation Stress. *J. Proteome Res.* **20**, 2195–2205 (2021).
2. **Weke, K.**[†], Kote, S.[†], Faktor, J., Al Shboul, S., Uwugiaren, N., Brennan, P. M., Goodlett, D. R., Hupp, T. R., and Dapic, I. DIA-MS proteome analysis of formalin-fixed paraffin-embedded glioblastoma tissues. *Anal. Chim. Acta* **1204**, 339695 (2022). [†] Equal contribution as first authors.
3. **Weke, K.**, Dziubek, K., Papak, I., Muszczek, M., Hupp, T.R., Kote, S. The interplay of hypoxia and antigen presentation in glioblastoma tumor: Insights from proteomics and immunopeptidomics. *Unpublished manuscript*.

The following paper is not included in this doctoral thesis but is of relevance to the field:

1. Bedran G., Gasser H-C., **Weke K.**, Wang T., Pankanin D., Laird A., Battail C., Zanzotto F.M., Pesquita C., Axelson H., Rajan A., Harrison D.J., Palkowski A., Pawlik M., Parys M., O'Neill R., Brennan P.M., Symeonides S., Goodlett D.R., Litchfield K., Fahraeus R., Hupp T.R., Kote S., & Alfaro J.A. The Immunopeptidome from a Genomic Perspective: Establishing the Noncanonical Landscape of MHC Class I-Associated Peptides. *Cancer Immunol Res.* 2023 Jun 2. PMID: 36961404.

List of figures

Figure 1-1. *Malignancy development, and cancer hallmarks.*

Figure 1-2. *Simple illustration of the three basic components of a mass spectrometer — ion source, mass analyzer, and detector.*

Figure 1-3. *Illustrating the process of electrospray ionization.*

Figure 2-1. *Overview of the Workflow of MicroPOTS for the Identification of Proteomic Changes in ~200 CP-A Cells after Treatment with Either LCA or X-ray.*

Figure 2-2. *Barplots and Venn diagram plots respectively showing the number of identified protein groups and overlap for CP-A WT, CP-A KO, and CP-A dKO samples of ~200 cells.*

Figure 2-3. *Qualitative and quantitative comparison of the identified proteins between replicates.*

Figure 2-4. *Assessment of the physicochemical features and subcellular localization of the identified proteins.*

Figure 2-5. *Volcano plots illustrating the fold difference in the expression levels of proteins.*

Figure 3-1. *Method evaluation and DIA-MS analysis of GBM FFPE tissues workflow.*

Figure 3-2. *Quantitative reproducibility assessment.*

Figure 3-3. *Sequence coverage and ranked abundance across the methods.*

Figure 3-4. *DIA-MS analysis of FFPE tissues derived from GBM patients (n = 5).*

Figure 3-5. *Heatmap and STRING diagram for a few selected proteins.*

Figure 3-6. *Exploring expression levels of immune-related genes.*

Figure 4-1. *Schematic workflow for proteomics and immunopeptidomics analysis.*

Figure 4-2. *Analysis of U-87 MG GBM cells' responsiveness to hypoxia and determination of surface expression of HLA-I molecules.*

Figure 4-3. *Hypoxia-dependent differential regulation of U-87 MG cells' proteome.*

Figure 4-4. *Enrichment analysis of differentially regulated proteins.*

Figure 4-5. *Hypoxia-dependent dysregulation of APM components.*

Figure 4-6. *Hypoxia alters HLA-I peptide repertoire.*

Supplementary Figure 2-1. *Venn diagram illustrating the shared number of identified proteins among different sample types, as well as proteins that are unique to each sample type.*

Supplementary Figure 2-2. *The number of missed cleavages is shown for all replicates and is expressed in percentage.*

Supplementary Figure 2-3. *Qualitative assessment of the reproducibility of the microPOTS system.*

Supplementary Figure 2-4. *Quantitative assessment of the reproducibility of the*

microPOTS system. The scatter plots illustrate the pairwise comparison of the logarithmically (base 10) transformed LFQ.

Supplementary Figure 2-5. *Scatter plot with associated Pearson's correlation coefficient ($R = 0.625$) between microPOTS and bulk proteomics for all overlapping 1066 proteins that we identified.*

Supplementary Figure 2-6. *Assessment of physicochemical characteristics.*

Supplementary Figure 2-7. *Assessment of physicochemical characteristics.*

Supplementary Figure 2-8. *Subcellular localization. Distribution of the identified proteins in various subcellular locations for each sample type.*

Supplementary Figure 3-1. *Evaluation of different protein extraction methods.*

Supplementary Figure 3-2. *Evaluation of digestion efficiency.*

Supplementary Figure 3-3. *Comparison of protein identifications using different approaches.*

Supplementary Figure 3-4. *Quantitative comparison of DIA-NN library and library-free approaches.*

Supplementary Figure 3-5. *Comparison of sequence coverage by DDA-MS and DIA-MS.*

Supplementary Figure 3-6. *Global correlation matrix of protein relative expression levels between samples.*

Supplementary Figure 3-7. *Boxplot showing the interpatient CV distribution for the top 20 proteins.*

Supplementary Figure 3-8. *Scatter plot showing Pearson's correlation coefficient ($R = 0.42$) between DIA-MS (this study) and DDA-MS (CPTAC - Clinical Proteomic Tumor Analysis Consortium) proteomics data.*

Supplementary Figure 4-1. *Overview of proteomics data and quality control.*

List of tables

Table 1-1. *Key molecular changes in high-grade gliomas.*

Table 3-1. *A brief highlight of tissue proteomics studies using DIA-MS or DDA-MS.*

Supplementary Table 2-1. *Median coefficient of variation (CV) for quantile normalized protein LFQ values for each sample type.*

Supplementary Table 2-2. *List of differentially expressed proteins between CP-A dKO cells with or without LCA treatment.*

Supplementary Table 2-3. *List of differentially expressed proteins between CP-A dKO cells with or without X-ray treatment.*

Table of contents

| | |
|---|-------------|
| Dedication..... | i |
| Acknowledgment..... | ii |
| Abstract..... | iii |
| Streszczenie..... | v |
| Abbreviations..... | vii |
| List of papers..... | x |
| List of figures..... | xi |
| List of tables..... | xiii |
| Table of contents..... | xiv |
| CHAPTER 1: Introduction..... | 1 |
| 1.1. Background information on glioblastoma tumor..... | 3 |
| 1.1.1. Evolution and classification of glioma tumors..... | 3 |
| 1.1.2. Pathogenesis of glioma tumors..... | 4 |
| 1.1.4. Epidemiology, clinical presentation, and diagnosis of GBM..... | 6 |
| 1.1.5. Treatment and management of GBM..... | 7 |
| 1.2. Immuno-oncological dynamics..... | 8 |
| 1.2.1. Antigen processing and presentation..... | 9 |
| 1.2.2. Tumor hypoxia..... | 10 |
| 1.2.3. Hypoxia-immune interface..... | 11 |
| 1.3. Mass spectrometry-based proteomics..... | 11 |
| 1.3.1. Sample preparation for bottom-up proteomics..... | 12 |
| 1.3.2. Emerging sample preparation techniques for low-input samples..... | 13 |
| 1.3.3. Mass spectrometry principle and instrumentation..... | 15 |
| 1.3.4. MS data acquisition schemes..... | 17 |
| 1.3.5. Quantification techniques..... | 19 |
| 1.4. Aims of this doctoral thesis..... | 20 |
| CHAPTER 2: MicroPOTS analysis of Barrett's oesophageal cell line models identifies proteomic changes after physiologic and radiation stress..... | 21 |
| 2.1. Abstract..... | 21 |
| 2.2. Introduction..... | 21 |
| 2.3. Experimental Section..... | 24 |
| 2.3.1. Materials..... | 24 |
| 2.3.2. Cell culture..... | 25 |
| 2.3.3. Cells treatment with LCA and X-ray..... | 25 |
| 2.3.4. Proteomic sample preparation in microwells..... | 26 |
| 2.3.5. LC-MS/MS analysis..... | 26 |
| 2.3.6. Data and statistical analysis..... | 27 |
| 2.3.7. Protein annotation and assessment of physicochemical aspects..... | 28 |
| 2.4. Results and discussion..... | 28 |
| 2.4.1. Analysing Barrett's oesophageal cell samples by microPOTS..... | 28 |
| 2.4.2. Protein identification from different CP-A genotypes and stress conditions..... | 29 |
| 2.4.3. Evaluation of protein extraction efficiency and reproducibility..... | 31 |
| 2.4.4. Comparison of microPOTS data to bulk proteomics dataset..... | 34 |

| | |
|--|-----------|
| 2.4.5. Comparison of physicochemical aspects of identified proteins..... | 34 |
| 2.4.6. Effect of stress on protein expression..... | 36 |
| 2.5. Conclusion..... | 37 |
| CHAPTER 3: DIA-MS proteome analysis of formalin-fixed paraffin-embedded glioblastoma tissues..... | 39 |
| 3.1. Abstract..... | 39 |
| 3.2. Introduction..... | 39 |
| 3.3. Materials and methods..... | 43 |
| 3.3.1. FFPE tissue samples..... | 43 |
| 3.3.2. Sample collection..... | 43 |
| 3.3.3. Sample processing..... | 43 |
| 3.3.4. Liquid chromatography-mass spectrometry (LC-MS/MS) analysis..... | 45 |
| 3.3.5. MS data analysis..... | 46 |
| 3.3.6. Data availability..... | 46 |
| 3.4. Results..... | 47 |
| 3.4.1. Comparative evaluation of different lysis buffer compositions..... | 47 |
| 3.4.2. Quantitative analysis and evaluation of precision..... | 49 |
| 3.4.3. Comparison of sequence coverage..... | 51 |
| 3.4.4. DIA-MS reliably detects and quantifies putative GBM biomarkers in FFPE tumor tissues..... | 54 |
| 3.4.5. Screening of immune-related genes in a DIA-MS dataset..... | 58 |
| 3.5. Discussion..... | 59 |
| CHAPTER 4: The interplay of hypoxia and antigen presentation in glioblastoma tumor: Insights from proteomics and immunopeptidomics..... | 62 |
| 4.1. Abstract..... | 62 |
| 4.2. Introduction..... | 63 |
| 4.3. Materials and methods..... | 64 |
| 4.3.1. Cultivation of cells under hypoxia or normoxia..... | 64 |
| 4.3.2. Protein isolation for Western Blot (WB)..... | 64 |
| 4.3.3. Immunoblotting..... | 65 |
| 4.3.4. Flow cytometry analysis..... | 65 |
| 4.3.5. Immunopeptidomics sample preparation..... | 66 |
| 4.3.7. LC-MS/MS sample measurement..... | 68 |
| 4.3.8. MS data processing and analysis..... | 69 |
| 4.4. Results and discussion..... | 70 |
| 4.4.1. Experimental design and study workflow..... | 70 |
| 4.4.2. HIF1 α is markedly expressed at 0.1% hypoxia..... | 71 |
| 4.4.3. Hypoxia-induced proteome changes in U-87 GBM cell lines..... | 72 |
| 4.4.4. Metabolic reprogramming is a key adaptation to extreme hypoxic stress..... | 76 |
| 4.4.5. Hypoxia-dependent dysregulation of APM components..... | 77 |
| 4.4.6. Hypoxia alters HLA-I peptide repertoire..... | 80 |
| 4.5. Conclusion..... | 81 |
| CHAPTER 5: Conclusions and future directions..... | 83 |
| Appendices..... | 84 |
| Appendix I: Supplementary material for chapter 2..... | 84 |

| | |
|--|------------|
| Appendix II: Supplementary material for chapter 3..... | 96 |
| Appendix IV: Supplementary material for chapter 4..... | 102 |
| References..... | 104 |

CHAPTER 1: Introduction

Normal cells have a complex genetic regulatory system that tightly controls cell proliferation. Damage to this control apparatus renders the cells incapable of regulating their cell cycle¹. Stochastic processes or environmental stresses spark aberrant changes within this regulatory system, leading to mutations such as deletions, insertions, and duplications. These mutational events modify cellular activities, leading to the inactivation of tumor suppressor genes, activation of oncogenes, deregulation of DNA (deoxyribonucleic acid) repair genes, downregulation of apoptotic genes, and widespread epigenetic alterations¹⁻⁴. Consequently, the modified cellular programs form the hallmarks of cancer development and progression⁵⁻⁸ (**Figure 1a-b**). While there have been remarkable steps to improve cancer diagnosis and therapy, the disease remains the second leading cause of death globally, with an incidence of 19.3 new cases and about 10 million deaths⁹. According to recent studies, the number of new cancer cases is predicted to be 28.4 million by the year 2040, which would reflect a 47% increase^{10,11}. The overall burden of cancer on healthcare is burgeoning, and more research is needed to improve patient clinical outcomes.

Clinical management for cancer patients is improving with research efforts aimed at early detection, accurate triaging, and treatment efficacy. Most of the research works that have yielded dramatic impact have mainly been conducted at the molecular level using a systems biology approach to characterize tumors¹²⁻¹⁶. Genomic and transcriptomic approaches have been the mainstays of tumor classification and bio-target discovery. Whereas these -omics studies have enabled the comparison and differentiation of various cancers, most of the results have not translated into clinics. Aberrant DNA changes and deregulated RNA (ribonucleic acid) expression profiles do not always reflect protein expression levels¹⁷⁻²². As a result, this limits our comprehensive understanding of this disease. Proteins are the effector biomolecules responsible for driving cellular functions; hence, they present the level where clinical diagnostic tests and targeted therapies work. Mass spectrometry (MS)-based proteomics has emerged as a powerful tool that can overcome the limitations of genomics.

MS-based proteomics can identify, quantify, and characterize almost all the proteins in a cell or tissue sample, including post-translational modifications and protein complexes²³. This information can be used to gain a deeper understanding of cancer biology, and cancer immunology and to identify new diagnostic and therapeutic targets.

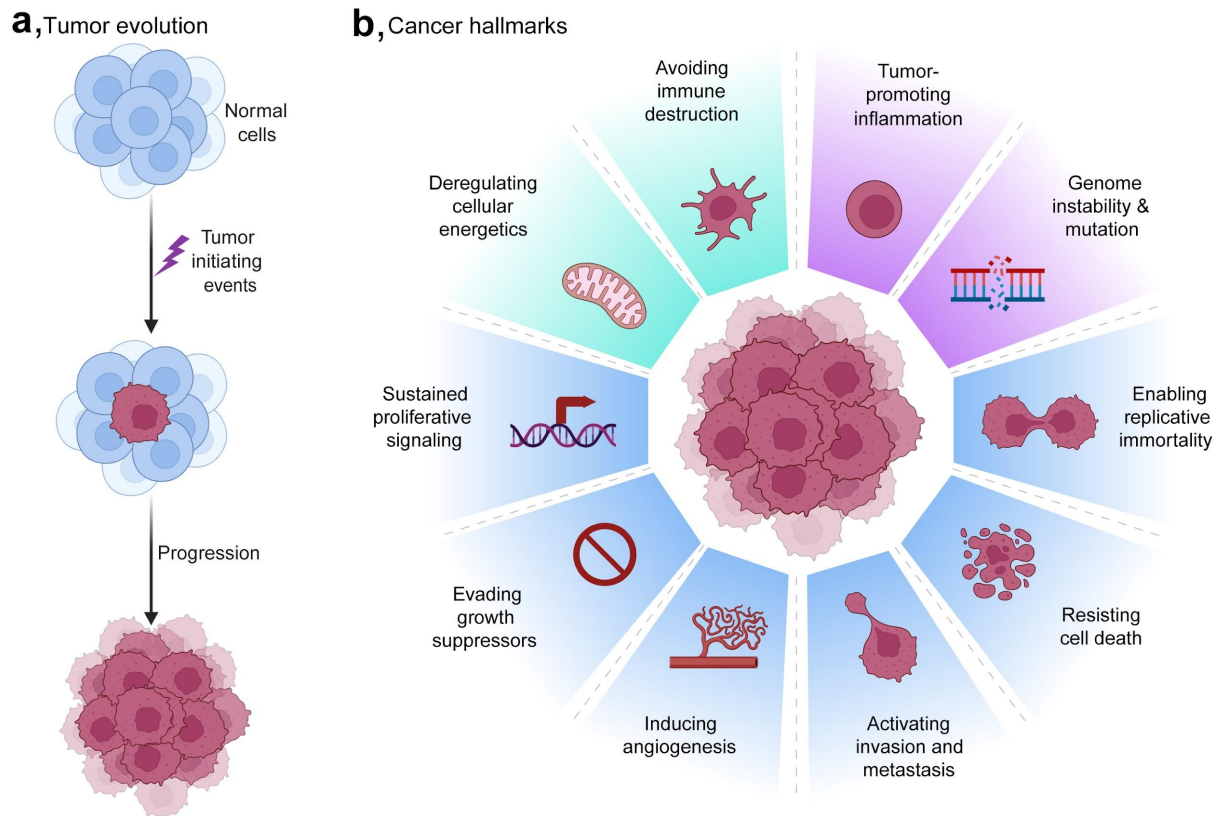


Figure 1-1. Malignancy development. a, Tumor initiation and development. Blue circles represent normal cells, and dark pink color display tumor cells. **b, Cancer hallmarks** showing the distinct biological phenotypes driving tumor growth and progression. Gradient colors, blue, green, and violet, show original hallmarks, enabling factors, and emerging hallmarks, respectively. The figure was created with [BioRender.com](https://www.biorender.com) and arranged in Adobe Illustrator.

1.1. Background information on glioblastoma tumor

1.1.1. Evolution and classification of glioma tumors

Glioblastoma (GBM) is a type of brain tumor that belongs to a larger group of tumors called gliomas. Gliomas are the most common type of malignant tumor of the brain and central nervous system (CNS)^{24,25}. They are characterized by high molecular and cellular heterogeneity and display varying grades of malignancy¹². Gliomas arise when normal glial cells, such as astrocytes, oligodendrocytes, and ependymal cells, undergo a neoplastic transformation and develop into astrocytomas, oligodendrogliomas, and ependymomas tumors, respectively. There have been profound changes to glioma nomenclature, which has seen a shift from histological-based classification to one anchored on the molecular landscape of the tumor²⁶⁻²⁸. The current classification by the World Health Organization (WHO) is an integrative approach involving histomorphologic features and molecular data to distinguish between the tumor subtypes. To discriminate between tumors that affect pediatrics and adult populations, the new approach groups diffuse gliomas into adult-type, pediatric-type diffuse low-grade, and pediatric-type high-grade gliomas.

The adult-type diffuse gliomas are classified according to the mutational status of the isocitrate dehydrogenase (IDH) gene, together with many other key molecular genetic aberrations. Based on this classification, there are three distinct tumor types, including IDH-mutant astrocytoma, IDH-mutant, and 1p/19q-codeleted oligodendroglioma, and finally, IDH-wildtype glioblastoma. IDH-mutant astrocytoma forms as a result of heterozygous point mutations of the IDH1 or IDH2 (less common) gene and mutations of alpha thalassemia/mental retardation syndrome X-linked (ATRX) and tumor protein P53^{29,30}. These tumors usually exhibit a diffuse growth pattern ranging from low-grade to high-grade malignancy. Within the IDH-mutant astrocytoma, three grades, namely WHO grades 2, 3, and 4, have been recognized. Tumor grades that harbor a homozygous deletion of cyclin-dependent kinase inhibitor 2A/B (CDKN2A/B) have been shown to have poor clinical outcomes and are assigned grade 4, irrespective of the histological findings³¹. IDH-mutant

and 1p/19q-codeleted oligodendroglioma is characterized by the presence of an IDH1 or IDH2 mutation and the codeletion of chromosomes 1p and 19q³². Different subsets of oligodendrogliomas exist and range from slow-growing to rapidly-growing malignant tumors. According to the latest grading scheme, the existing grades include WHO grade 2 or 3²⁷. IDH-wildtype glioblastoma is the most common malignant adult-type diffuse glioma. These tumors have features that include high cellularity, diffuse infiltrating, and pleomorphic with mitotic activity as well as either microvascular proliferation or necrosis. GBM is frequent among older adults; however, the tumor can still develop at any age.

1.1.2. Pathogenesis of glioma tumors

The advancement in -omics technologies has significantly improved the understanding of the key molecular drivers that initiate and support the progression of glioma tumors. These pathogenic molecular alterations (**Table 1-1**) have become extremely useful in subtyping glioma tumors. ATRX mutation—the protein encoded by the ATRX gene, situated on Xq21.1, plays a crucial role in the chromatin-rearrangement pathway by facilitating the integration of histone H3.3 into heterochromatin. ATRX loss is more common in IDH-mutant than IDH-wildtype glioma and correlates to TP53 mutation²⁹. In a large-scale study of glial tumors, ATRX mutations were strongly linked to IDH1/2 and H3F3A mutations³⁰. Telomerase reverse transcriptase (TERT) promoter mutation—the TERT gene is critical for encoding the telomerase enzyme, adding the missing 3' end of the DNA strand during replication. TERT mutation is a prognostic marker and has been demonstrated to occur often in IDH-wildtype compared to IDH-mutant²⁶.

TP53 mutation—the TP53 gene on chromosome 17p13.1 plays a critical role in the genome regulatory network, modulating cell proliferation and survival, among other cellular functions. Mutation in the TP53 gene is associated with GBM progression³³. Chromosome 1p/19q codeletion—the complete deletion of the short and long of chromosomes 1 and 19, respectively, is a defining characteristic of oligodendroglial tumors. A phase III clinical study and another report showed that 1p/19q codeletion is a powerful predictor of a favorable

treatment outcome and survival for patients with diffuse gliomas^{32,34}. BRAF (v-raf murine sarcoma viral oncogene homolog B1) is a proto-oncogene on chromosome 7q34, encoding the B-Raf protein, which is part of the Raf protein family. Raf proteins are key components of the mitogen-activated protein kinase (MAPK) signaling pathway, which controls most cell activities, such as motility, proliferation, differentiation, and apoptosis³⁵. V600E is the most common oncogenic driver mutation in the BRAF gene and is present in a variety of glioma tumors.

The epidermal growth factor receptor (EGFR) gene on chromosome 7p12 is crucial in promoting cell proliferation by activating MAPK and PI3-Akt pathways. EGFR amplification occurs more frequently in IDH-wildtype than IDH-mutant and is associated with poor prognosis²⁶. In the recent CNS WHO classification, EGFR amplification, TERT promoter mutation, and/ or the concurrent gain and loss of whole chromosomes 7 and 10 (+7/-10 genotypes), respectively, were shown to adequately define IDH-wildtype GBM^{36,37}. O⁶-methylguanine DNA methyltransferase (MGMT) gene encodes DNA repair enzyme that reverses the genotoxic impacts of alkylating agents. MGMT promoter methylation is a useful molecular prognosticator in glioma tumors³⁸. Phosphatase and tensin homolog (PTEN) gene on chromosome 10q23 plays a critical role in regulating cell growth and suppressing the tumor. Loss of PTEN function as a result of mutation or loss of heterozygosity (LOH) impairs control of cell proliferation, growth, and survival.

Table 1-1. Key molecular changes in high-grade gliomas

| Type of molecular alteration | | | | |
|------------------------------|-----------------------------|-----------|----------------|-----------|
| Methylation | Chromosomal aberrations | Mutations | Amplifications | Deletions |
| MGMT | Gain: 1q, 19q, 20q | PTEN | MET | CDKN2A |
| PTEN | Loss: 6q, 9p, 13q, 14q, 22q | ATRX | EGFR | CDKN2B |
| RB1 | Concomitant: +7/-10 | TP53 | PIK3CA | CDKN2C |

| | | | | |
|----------|--|--------|--------|--------|
| TP53 | | RB1 | PDGFRA | PTEN |
| CDKN2A | | IDH1/2 | CCND2 | RB1 |
| PDGFB | | NF1 | MDM2/4 | NFKB1A |
| EMP3 | | EGFR | | |
| SOCS1 | | | | |
| PCDHGA11 | | | | |
| OLIIG1/2 | | | | |

1.1.4. Epidemiology, clinical presentation, and diagnosis of GBM

As previously stated, GBM is the most prevalent primary malignancy of the brain. Its incidence has not been clearly defined; however, various reports have estimated the age-adjusted incidence rate at an annual average of between 3.19 to <5 cases per 100,000 persons^{24,25}. In the pediatric population, the incidence rate is approximated at 0.85 per 100,000 person-years, and GBM accounts for 3-15% of all primary brain tumors^{24,39}. The median age of diagnosis is 64 years, and the prevalence rises with age and finally peaks at 75-84 years and then starts to drop at 85 years of age. It occurs more in men than women, with an incidence rate of 1.57% in men compared to that in women⁴⁰. Primary GBMs are frequent in men, while secondary ones occur more in women. Ancestrally, the prevalence of GBM is higher among those of Caucasian ancestry⁴¹.

The clinical presentation of GBM is often linked to the region where the tumor is located. Patients harboring supratentorial tumors present with headaches as the tumor presses on the surrounding normal tissues⁴². Nausea, vomiting, or seizures are other common presentations that may occur. In some cases, the onset of symptoms may be insidious and patients may only present with behavioral or cognitive changes⁴³. Patients may complain of gait disturbances, cranial deficits, and vertigo when they have infratentorial tumors. Besides, patients can also present with non-specific symptoms including fatigue, malaise, and weight loss.

1.1.5. Treatment and management of GBM

GBM tumor is extremely challenging to treat, and efforts to achieve improved treatment have not dramatically changed the therapeutics landscape for patients. Surgery remains the primary treatment for GBM patients. Surgical resection aims to remove as much of the tumor as possible without causing damage to the surrounding healthy brain tissue. The extent of resection has been demonstrated as an essential prognostic factor in GBM. For instance, a meta-analysis of 37 studies, including 41,117 patients, reported that gross total resection (GTR) was associated with improved overall survival (OS) compared to subtotal resection (STR) or biopsy alone⁴⁴. Following surgical resection, radiation therapy is typically administered to the tumor bed and surrounding brain tissue to eliminate any remaining cancer cells. Standard radiation therapy involves delivering a total dose of 60 Gy in 30 fractions over six weeks⁴⁵.

In a landmark phase III trial, Stupp *et al.* demonstrated that the addition of temozolomide (TMZ) to radiation therapy improved median OS from 12.1 months to 14.6 months and increased the 2-year survival rate from 10.4% to 26.5% in patients with newly diagnosed glioblastoma⁴⁶. TMZ is an oral alkylating agent that is often used in the treatment of GBM. It works by methylating the O⁶ position of guanine, which leads to DNA damage and apoptosis in tumoral cells. The standard adult dosing regimen for TMZ is usually 75 mg/m² per day for five days during each cycle, which is repeated every 28 days. The optimal duration of TMZ therapy is not well defined, but most clinicians continue TMZ for six cycles. Immunotherapy is an emerging treatment modality for GBM. Immune checkpoint inhibitors (ICIs) target inhibitory receptors on T cells, such as programmed cell death protein 1 (PD-1) and cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), which can enhance antitumor immunity⁴⁷⁻⁴⁹.

The efficacy of ICIs in GBM is limited, partly due to the immunosuppressive tumor microenvironment (TME) and the lack of reliable biomarkers to predict response. Other immunotherapy approaches being investigated in GBM include chimeric antigen receptor (CAR) T cell therapy, tumor vaccines, and ICIs targeting other immune checkpoints⁵⁰⁻⁵².

Targeted therapy refers to the use of drugs that target specific molecular pathways or mutations that are important for cancer cell growth and survival. Several targeted therapies have been investigated in GBM, but they have not demonstrated significant clinical benefit⁵³⁻⁵⁵. For example, inhibitors of vascular endothelial growth factor (VEGF), such as bevacizumab, have been shown to improve progression-free survival (PFS) but not OS^{56,57}. Inhibitors of the EGFR, such as erlotinib and gefitinib, have also been investigated, but their efficacy in GBM is limited⁵⁸⁻⁶⁰.

1.2. Immuno-oncological dynamics

The TME is cellularly and molecularly complex, which greatly influences the clinical outcome of cancer patients⁶¹. Various cell types, including tumor parenchyma cells, stromal cells, and tumor-infiltrating immune cells, contribute to the TME diversity. These constituents of TME interact at different levels and scales and exhibit nonlinear dynamics. Two immune components of TME exist—those involved in anti-tumor activity and the component that promotes tumor progression. Several studies have documented the interaction between the tumor and the immune system^{62,63}. A significant hypothesis that has emerged to describe this interaction at the cellular level is the concept of immunoediting⁶⁴. Immunoediting describes the dynamic relationship between the immune system and tumor cells, suggesting this interplay can be divided into three phases: elimination, equilibrium, and escape. In the first phase, elimination, the immune system recognizes and seeks to eradicate emerging tumor cells. At this stage, both innate and adaptive immune responses are activated — cytotoxic CD8+ T cells, natural killer (NK) cells, and macrophages target and eliminate nascent tumor cells that display antigens deemed foreign or altered.

As some tumor cells manage to elude total eradication, they devise mechanisms to withstand immune detection. These cells enter a state of dormancy, and equilibrium, where the immune response prevents their proliferation but does not eliminate them completely. This phase is characterized by the continuous sculpting of the tumor, where the immune system unintentionally selects for tumor cells that are less immunogenic and, therefore, less

recognizable. Thus, there is a balance between the rate of tumor cell proliferation and elimination. Production of pro-tumor cytokines, e.g., TGF- β , and anti-tumor cytokines, such as IL-12 and IFN- γ is at equal levels. Both CD4+T and CD8+T cells continue to mount immune responses, with NK and T_{reg} cells getting involved. The phase can last for several years, with the emergence of non-immunogenic tumor cells. In the escape phase, tumor cells that have adapted to evade immune detection or suppress immune responses become dominant. These immune-resistant cells multiply, resulting in tumor growth and potential metastasis. These cells implement different strategies that make them immune-resistant, e.g., they downregulate antigen-presenting cells and proteins involved in antigen processing and presentation machinery, upregulate the expression of co-inhibitory signals such as CTLA-4 and PD-1 that bind to B7-1/B7-2 or PD-L1/PD-L2 and increase the production of anti-apoptotic molecules⁶⁵.

1.2.1. Antigen processing and presentation

For the adaptive arm of immunity to function effectively, antigens must be processed and presented to immune cells. This process is complex and involves different cells and major histocompatibility complex (MHC) molecules, which for humans is the human leukocyte antigen (HLA). Cells that mediate this process are called antigen-presenting cells (APC), and they include dendritic cells, macrophages, and B cells. HLA classes I and II are the two molecules expressed on cell surfaces that mediate antigen presentation^{66,67}. Both HLA-I and HLA-II bind and display short antigenic peptides, which CD8+ and CD4+, respectively, recognize. HLA-I typically binds to peptides that are 8-11 amino acids long, with a dominant 9 mer, whereas HLA-II binds peptides that are 13-25 amino acids long⁶⁸. In terms of antigenic peptide sources, HLA-I-bound peptides are mainly endogenous or intracellular, while those associated with HLA-II originate exogenously or extracellularly. However, there are instances, where endogenous peptides are presented by HLA-II and vice versa, in a phenomenon referred to as cross-presentation.

The endoplasmic reticulum (ER) is crucial for packaging and assembling HLA-I molecules. HLA-I has two major components—the β 2-microglobulin chain and a polymorphic heavy chain, which form a complex. Initial steps in antigen presentations by HLA-I involve antigen generation by a ubiquitin–proteasome system, which degrades proteins into peptides specific for HLA-I binding. Peptide-loading complex comprising of transport associated with antigen processing (TAP), tapasin, HLA-I, ERp57, and calreticulin is critical for translocating the generated peptides into the ER. Aminopeptidases, ERAP (ER aminopeptidase 1) 1 and 2, further process the peptides by trimming them in the ER. The peptide-loading complex enables the binding of the trimmed peptides by HLA-1. Next, the peptide-HLA-I complex is transported to the cell surface for antigen presentation CD8+ T cells.

1.2.2. Tumor hypoxia

Tumor hypoxia, a condition in which tumor cells are deprived of oxygen, is a common feature of solid tumors. It is caused by the rapid growth of tumor cells, which leads to dysfunctional vasculature and a reduced oxygen supply. Physiological oxygen concentration in most normal tissues is between 4% and 10% (40-60 mmHg oxygen tension), depending on the tissue type or organ^{69,70}. In tumors, oxygen levels are below physiological concentrations (<10 mmHg oxygen tension), establishing a pathological state called hypoxia. Hypoxia is linked to increased tumor aggression, progression, and resistance to therapeutic interventions. There are two main types of tumor hypoxia: acute and chronic. Acute hypoxia is caused by limited perfusion due to aberrant blood vessels shutting down. Chronic hypoxia is caused by limited diffusion and occurs as a result of increased diffusion distances.

The cellular response to hypoxia is principally modulated via the hypoxia-inducible factor (HIF) pathway⁷¹. HIF is a heterodimeric transcription factor encompassing an oxygen-sensitive α -subunit (HIF-1 α , HIF-2 α , or HIF-3 α) alongside a constitutively expressed β -subunit, HIF-1 β or the aryl hydrocarbon receptor nuclear translocator (ARNT). Under conditions of normal oxygen levels (normoxia), the HIF- α subunit undergoes hydroxylation mediated by prolyl hydroxylases (PHDs). This hydroxylation facilitates the binding of the von

Hippel-Lindau (VHL) protein to HIF- α , which recruits E3-ubiquitin ligases, targeting HIF- α for proteasomal degradation. Conversely, during hypoxia, the activity of PHDs is inhibited, resulting in the preservation and stabilization of HIF- α . Once stabilized, HIF- α translocates to the nucleus, where it associates with HIF- β and cofactors, including CBP/p300, thereby driving the transcription of various target genes. These genes play pivotal roles in processes such as angiogenesis, the reprogramming of cellular metabolism towards glycolysis, and the survival and proliferation of tumor cells, among other crucial tumorigenic pathways.

1.2.3. Hypoxia-immune interface

The hypoxia-immune interface in tumors is characterized by profound changes in the function and behavior of various immune cells. Cytotoxic CD8+ T cells and NK cells are inhibited in their function and activity in hypoxic tumor regions. This suppression can arise directly from low oxygen conditions or indirectly due to the altered secretome of the TME that contains immune-suppressive molecules. Tumor-associated macrophages (TAMs) accumulate in the hypoxic zones of tumors⁷². Within these regions, TAMs tend to adopt a phenotype that is skewed toward supporting tumor growth, enhancing angiogenesis, and further suppressing the immune attack against the tumor. Under hypoxic conditions, dendritic cell differentiation and antigen presentation are impaired, weakening the subsequent T-cell-mediated immune response against tumor cells. Additionally, hypoxia increases T_{reg} recruitment and function⁷³.

1.3. Mass spectrometry-based proteomics

Proteins are the effector biomolecules whose functions alter the physiological states of cells within a biological system. A comprehensive understanding of distinct features of proteins, including their expression levels, distribution, post-translational modifications (PTMs), and interactions, is imperative to gaining insights into cancer evolution, progression, and therapeutic treatments. The term proteome was first coined by Wilkins in 1994⁷⁴. Different approaches, including affinity-based such as immunofluorescence, Western blotting, enzyme-linked immunosorbent assay, immunoprecipitation, immunohistochemistry, and flow

cytometry have been employed in the identification and characterization of proteins⁷⁵. However, affinity-based methods suffer many limitations, e.g., expensive antibodies and lack of multiplexing, and as a result, MS-based methods have evolved to overcome these challenges and have become the analytical technique of choice for proteomics analysis⁷⁶.

MS-based proteomics can be carried out in different approaches, including top-down, bottom-up, and middle-down depending on the study^{77,78}. In top-down proteomics analysis, intact proteins are analyzed directly without breaking them down into peptides. It is useful in detecting PTMs; hence, providing detailed information about a protein's modification state. However, it is relatively less sensitive compared to the bottom-up strategy. In a bottom-up approach, the proteins are digested enzymatically (trypsin in most cases) into peptides, which are then analyzed by MS. It is highly sensitive with well-established workflows; however, decomplexing the resulting data is sometimes a challenge. For the middle-down approach, proteins are partially digested, leading to larger peptide fragments in comparison to bottom-up peptides. It does offer a balance between preserving PTMs and sensitivity. Bottom-up proteomics is the most popular approach, and work on this study is chiefly based on it.

1.3.1. Sample preparation for bottom-up proteomics

Initial steps in sample preparation commence with the retrieval of proteins from biological samples, e.g., cells, tissues, or biofluids. Depending on the type of the biological specimen, samples can be first pretreated, including the use of protease inhibitors to inactivate enzymes. Subsequent processes include homogenization, protein extraction, and quantification. Extracted proteins are quantified to determine the extraction yield/concentration, which informs the amount of enzyme necessary for efficient digestion. The next step in the workflow after extracting proteins is converting them into peptides in a series of steps, including reduction to break the disulfide bridges and alkylating to cap the reduced cysteines, enzymatic digestion, peptide fractionation (optional), and desalting.

Currently, there are several strategies for performing protein digestion including in-solution, in-gel, and filter-aided sample preparation (FASP)^{79,80}.

1.3.2. Emerging sample preparation techniques for low-input samples

The standard bottom-up proteomic workflows are inefficient for single-cell or microscale proteomics analysis⁸¹. Usually, a classical bottom-up proteomics requires a large amount of starting material, often ranging from several hundreds of thousands to slightly over one million cells. Sample processing is invariably accompanied by adsorptive losses due to the multiple steps that expose the samples to many surfaces. Losses attributed to the surface adsorption; however, become negligible with large input material. In addition, bulk methods involve the use of large volumes of liquid, sometimes up to 1.5 mL. Without downscaling such huge liquid volumes for single-cell analysis, the consequence is inefficient enzymatic digestion. These bottlenecks have been the impetus for emerging techniques that focus on miniaturizing conventional sample preparation workflows and improving separation to fit the need for single-cell analysis.

Some of these emerging techniques include the single POT systems (nano- and microdroplet processing in one pot for trace samples—nanoPOTS and microPOTS)^{82,83}. The -POTS system has been enabled hugely by microfluidics technology. All sample processing steps, including cell lysis, reduction, alkylation, and digestion, take place in one pot/nanowell. Furthermore, the processing volume in each nanowell is minimized to <200 nL. With each nanowell having a surface area of about 0.8 mm², adsorptive sample losses due to surface exposure are massively reduced. MS-compatible buffer components are used for cell lysis as there is no further clean-up step. Initially, nanoPOTS processed as few as 10 HeLa cells using 0.1% RapiGest and applied a two-step protease reaction leading to the identification of >1500 proteins⁸².

Single-tube sample processing formats aim to eliminate multiple reaction vessels that inevitably result in adsorptive losses due to multi-surface exposures. For instance, a

single-tube sample processor eliminated the need for a clean-up step by using MS-compatible organic solvents to extract and denature proteins⁸⁴. Using a surfactant (NP-40) to lyse the cells and subsequently precipitate proteins by acetone, Wang *et al.*⁸⁵, demonstrated the potential of a single tube to process 500 to 5000 cells. With this method, the authors reported on average 167, 237, and 491 protein identifications from 500, 1000, and 2500 MCF-7 breast cancer cells, respectively. Li *et al.*⁸⁶ developed an integrated sample processing platform with the capability of isolating a few hundred cells and subsequent preparation in a mini-tube. The platform reported enhanced detection sensitivity with an increased number of identified proteins, e.g., ~4000 proteins were identified with an injection of the equivalent of only 100–200 cells per analysis of rare MCF-7 cells spiked into whole blood.

A modified version of FASP—called micro-FASP—has emerged to detect and analyze proteins in low-input samples⁸⁷. Martin *et al.*⁸⁸ reported SOPs-MS (surfactant-assisted one-pot sample processing at the standard volume coupled with MS), which is an easy-to-implement technique and resulted in a reliable label-free quantification of about 1200 to 2700 protein groups from 50 to 1000 MCF10A cells, respectively. Furthermore, using a single low-binding polymerase chain reaction (PCR) tube with the right surfactant e.g. n-Dodecyl- β -D-maltoside can accelerate single-cell analysis as reported recently⁸⁹.

In on-line processing formats, samples can either be processed using the principle of strong cation exchange or immobilized enzyme reactor systems. The strong cation exchange (SCX) approach was demonstrated by Ethier *et al.*⁹⁰ when they developed a sample processing system termed a proteomic reactor. Other methods based on SCX to address proteomics analysis of low sample input have emerged. For instance, a system called rare cell proteomic reactor (RCPR) has since been developed to improve sensitivity⁹¹. An automated SCX-based fractionation system was shown to achieve high digestion efficiency with over 7000 proteins identified from 20 μ g protein extracts of HEK 293T cells. Simple and integrated spin tip-based proteomics technology (SISPROT) consisting of SCX beads and

C18 disk in a single pipette enabled the identification of 1270 proteins from about 2000 HEK 293 cells and achieved good reproducibility⁹².

1.3.3. Mass spectrometry principle and instrumentation

MS is an analytical technique that is used to determine the molecular weight, elemental composition, and structural information of a molecule. MS has many areas applications in several areas of research, including proteomics, metabolomics, drug discovery, and forensic science.^{93–95} In proteomics, which is a large-scale study of proteins, MS has increasingly become the preferred method for protein analysis. MS-based proteomics generates data that is enough to provide detailed information about proteins, including amino acid sequence, post-translational modifications (PTMs), protein-protein interactions, and spatial and temporal distributions within cells^{96–98}. MS detects and analyzes components of a sample based on mass-to-charge ratio (m/z) values. Advancement in technology has seen the emergence of different MS platforms. While many instrumentation configurations exist, all mass spectrometers share three basic components—an ion source, a mass analyzer, and an ion detector (**Figure 1-2**). The ion source generates ions, which the mass analyzer separates and measures their m/z , and the detector detects and records the number of ions at each m/z value.

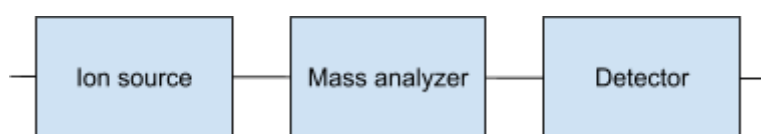


Figure 1-2. Simple illustration of the three basic components of a mass spectrometer—ion source, mass analyzer, and detector.

In the ionization process, the sample is transformed into charged ions in a gaseous phase. Currently, different ion sources are available with the capability of generating either positively or negatively charged ion species. During ionization, a sufficient amount of energy is supplied to the peptide molecule, which resultantly removes one or more electrons to generate charged ions that are singly- ($[M]^+$, $[M + H]^+$) or multiply-charged ($[M + nH]^{n+}$).

Initially, ion generation occurred in a vacuum; however, with advancements in technology, instruments with ion sources that operate at atmospheric pressure are now available, but analysis and detection of ions still take place in a vacuum environment. Different ionization techniques exist, including chemical ionization (CI), electron ionization (EI), matrix-assisted laser desorption/ionization (MALDI), and electrospray ionization (ESI) (**Figure 1-3**)^{99–101}. In proteomics analysis, there is a huge preference for MALDI and ESI over other ionization methods as they have varied separation modes, high analytical sensitivity and accuracy, and ease of implementation. All data in this study rely entirely on ESI. In MALDI-facilitated ionization, samples and organic matrix are co-crystallized on a metal plate, followed by a pulsed laser that excites the matrix, thereby thermally heating the molecules, and finally desorption of ions into the gaseous phase. Principally, ESI relies on spraying a fine mist of ions into the MS at atmospheric pressure. ESI is invariably interfaced with liquid separation systems as it applies a voltage charge to a solution to ionize molecules.

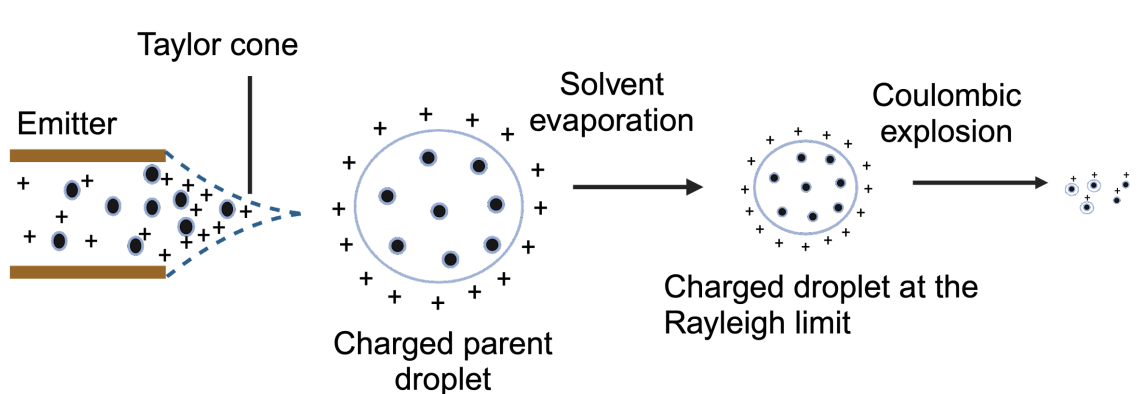


Figure 1-3. *Illustrating the process of electrospray ionization. Created with [BioRender.com](https://www.biorender.com).*

Once the molecules are ionized, ions are channeled to the mass analyzer for separation on the basis of their m/z ratios using a magnetic or electrical field. Depending on the need, sample, and type of analysis, different mass analyzers can be used either as standalone or in combination^{102–105}. Mass analyzers fall into two broad classifications—beam analyzers, in which ions move from the ion source in the form of a beam and travel through the analyzing

field to the detector, and trapping analyzers, in which ions are first trapped within the analyzing field prior to detecting them by a detector. Under beam analyzers, common ones include Time-of-Flight (ToF), and quadrupole (Q) analyzers, while popular trap analyzers are quadrupole ion trap (QTRAP), orbitrap, and Fourier-Transform Ion-Cyclotron Resonance (FT-ICR). Mass analyzers can be combined together in tandem or hybrid fashion, which is more beneficial in providing detailed information on the analyte. Tandem platforms basically have similar kinds of analyzers, e.g., Time-of-Flight/Time-of-Flight (ToF/ToF) and Triple Quadrupoles (QqQ). On the other hand, hybrid MS uses different kinds of analyzers to analyze and detect peptide ions. Linear ion Trap/Fourier-Transform (LIT/FT) and Quadrupole/Time-of-Flight (Q/ToF) are examples of hybrid instruments¹⁰⁵. After the mass analyzer separates the ions, they are channeled to the detector, which detects the signal in the form of ion currents, consequently amplifying them and determining and registering their abundance values. Most detectors are multipliers, e.g., electron multiplier, and basically amplifies the signal by generating a cascade of electrons as ions hit its surface.

1.3.4. MS data acquisition schemes

In bottom-up or peptide-based experiments, MS data acquisition schemes fall into two broad classes—untargeted/discovery and targeted. Untargeted/discovery approaches include data-dependent acquisition (DDA) and data-independent acquisition (DIA), while targeted strategies consist of selected/multiple-reaction monitoring (SRM/MRM) and parallel-reaction monitoring (PRM)^{106,107}.

In DDA, the instrument acquires a full/survey scan mass spectrum known as MS1, which has all the precursor ions collected within the set time or once the automatic gain control (AGC) target value is reached^{108,109}. Following this, a pre-defined number of the most abundant precursor ions—“topN”—are selected and subjected to fragmentation. The resultant fragment ions are then analyzed in a second mass spectrometer scan, referred to as MS2. This process repeats cyclically, ensuring the most intense ions are continually selected for fragmentation and analysis. Less abundant peptide ions can be targeted by

setting a dynamic exclusion window to ensure the same m/z is not repeatedly fragmented. One advantage of DDA is the immediate usability of MS2 spectra because every spectrum is derived by fragmenting a narrow m/z isolation window (usually around 1 Da), typically representing fragments of a singular chemical entity. However, DDA suffers from a small number of ions that can be fragmented in a single injection. Also, fragmentation is inherently stochastic, leading to relatively low quantitative reproducibility.

DIA can be viewed as a hybrid approach that merges PRM's strengths with DDA's extensive protein coverage. In DIA, precursors are segmented into successive m/z windows (typically 5–25 Da). The instrument then cycles within each window, fragmenting all peptide ions. Analyzing these complex MS2 spectra often requires searching a DDA-established peptide spectral library, although new bioinformatics tools are making direct analyses possible^{110–113}. Study shows that the quantification level of DIA schemes is on par with targeted methods¹¹¹. As DIA continues to evolve, it is being recognized for its potential in large-scale protein quantification; hence, impacting precision medicine and drug discovery. However, the width of MS1 isolation windows in DIA impacts the results. Generally, narrower windows enhance selectivity and sensitivity, but they lengthen the cycle time, which can reduce data points over a chromatographic peak. Besides, the choice of spectral library matters. While comprehensive organism-level libraries are convenient, the risk of false discoveries can be high. Therefore, a project-specific library generated from a DDA study of the same samples, is often preferable.

SRM/MRM approach operates on a triple quadrupole mass spectrometer (QqQ MS). In this mode, targeted peptide precursors pass through Q1, get fragmented, and then only the selected ions are monitored in Q3¹¹⁴. This consequently produces multiple transition ion chromatograms; thus, enabling precise detection and quantification. On the other hand, PRM is an evolution of the SRM/MRM approach. Instead of a third quadrupole, PRM uses a high-resolution accurate-mass (HRAM) mass analyzer like an Orbitrap or a TOF analyzer.

Both SRM/MRM and PRM are renowned for their sensitivity, precision, and consistency, but they have low proteome coverage.

1.3.5. Quantification techniques

A comprehensive understanding of variation in protein levels is critical to distinguishing healthy and disease phenotypic features, which enables the uncovering of underlying pathological mechanisms. In MS-based proteomics, proteins can be quantified to determine changes between conditions. Depending on the study question and sample, several techniques exist for both proteome-wide as well as quantification of individual proteins. These techniques broadly fall into two categories—absolute and relative quantification^{115–117}. The absolute quantification method is used when determining the absolute abundance or number of individual molecules of particular proteins in a sample. However, this approach is complex to perform as it requires instruments with the capability of measuring the number of copies of the given protein in a sample mixture; hence, it is not common. In principle, the relative quantification technique compares the abundance of a particular protein in each of two or several samples and then determines the ratio of their occurrence between samples.

Relative quantification methods can be further divided into label-based and label-free quantification¹¹⁸. Label-based strategies consist of *in vitro* or *in vivo* (metabolic) labeling of proteins or peptides with heavy, non-radioactive, isotopes. This methodology enables multiplexing of samples; thus, achieving high quantitative accuracy and precision. The incorporation of heavy isotopes results in a *m/z* shift of the MS spectrum. In this manner, the respective signal intensities from labeled and unlabeled peptides are compared to achieve quantification. Popular chemical labeling methods include tandem mass tags (TMT), isobaric tags for relative and absolute quantification (iTRAQ), and isotope-coded affinity tags (ICAT). Stable isotope labeling with amino acids in cell culture (SILAC) is a metabolic labeling approach that labels proteins *in vivo* during synthesis. While these label-based techniques have more quantitative benefits, including high reproducibility and data completeness, they remain expensive, have several sample preparation steps, and only a small number of

samples can be analyzed in a single experiment. On the other hand, label-free quantification can further be subdivided into two classes—measuring of peptide signal intensity using extracted ion chromatogram (XIC) or counting the number of peptide-to-spectrum matches (PSM, spectral counting) observed for each protein. In the first approach, the signal intensity of the area under the chromatographic peak of peptide ion is measured—quantification is achieved by integrating the ion intensities of a particular peptide over its chromatographic elution profile. For spectral counting, quantification is accomplished by comparing the number of MS/MS fragment ion spectra of given peptides between distinct samples. Label-free relative quantification, in comparison to other methodologies, is simple and cost-efficient; however, it is low in quantitative accuracy, and oftentimes one requires many technical replicates to circumvent data missingness.

1.4. Aims of this doctoral thesis

The overall aim of this thesis was to develop, apply, and integrate MS-based proteomics and immunopeptidomics technologies to study cancer, with a special focus on Glioblastoma (GBM), aiming to enhance the diagnosis and treatment modalities of cancer.

1. To apply a microscale MS-based proteomics method for the analysis of limited numbers of cells.
2. To develop a workflow for proteome analysis of formalin-fixed paraffin-embedded glioblastoma tissues
3. To investigate the impact of hypoxia on antigen presentation in glioblastoma using an integrated proteomics and immunopeptidomics approach.

CHAPTER 2: MicroPOTS analysis of Barrett's oesophageal cell line models identifies proteomic changes after physiologic and radiation stress

This work was originally published in the *Journal of Proteome Research* and is presented with a few modifications:

Weke, K., Singh, A., Uwugiaren, N., Alfaro, J. A., Wang, T., Hupp, T. R., O'Neill, J. R., Vojtesek, B., Goodlett, D. R., Williams, S. M., Zhou, M., Kelly, R. T., Zhu, Y., and Dapic, I. MicroPOTS Analysis of Barrett's Esophageal Cell Line Models Identifies Proteomic Changes after Physiologic and Radiation Stress. *J. Proteome Res.* **20**, 2195–2205 (2021).

2.1. Abstract

Moving from macroscale preparative systems in proteomics to micro- and nanotechnologies offers researchers the ability to deeply profile smaller numbers of cells that are more likely to be encountered in clinical settings. Herein a recently developed microscale proteomic method, microdroplet processing in one pot for trace samples (microPOTS), was employed to identify proteomic changes in ~200 Barrett's esophageal cells following physiologic and radiation stress exposure. From this small population of cells, microPOTS confidently identified >1500 protein groups and achieved a high reproducibility with a Pearson's correlation coefficient value of $R > 0.9$ and over 50% protein overlap from replicates. A Barrett's cell line model treated with either lithocholic acid (LCA) or X-ray had 21 (e.g., ASNS, RALY, FAM120A, UBE2M, IDH1, ESD) and 32 (e.g., GLUL, CALU, SH3BGRL3, S100A9, FKBP3, AGR2) overexpressed proteins, respectively, compared to the untreated set. These results demonstrate the ability of microPOTS to routinely identify and quantify differentially expressed proteins from limited numbers of cells.

2.2. Introduction

Mass spectrometry (MS) has emerged as the most powerful technology for the analysis and discovery of proteins¹⁰³. Since the term proteome was coined in 1994,⁷⁴ researchers have used MS to comprehensively define the molecular mechanisms that underpin cellular

functions. To achieve this goal, there are two main approaches to proteomics: top-down and bottom-up, with the latter being more often applied today^{77,119}. Significant gains have been made by applying these approaches in large-scale studies¹²⁰ to fully profile protein expression and their post-translational modifications. However, standard proteomic analysis demands substantial amounts of starting material to exhaustively characterize a proteome. For instance, about 10⁵ to millions of cells have typically been used to achieve a high proteome coverage¹²¹. Historically, utilization of such large amounts of starting material often precluded the ability of proteomics to compete with genomics in the analysis of small numbers of cells. This is because genomics allows sample material to be amplified through polymerase chain reaction (PCR)¹²². Proteomics, on the other hand, has had to pay special attention to sample preparation of small numbers of cells in order to avoid adsorptive losses of low-abundance proteins. Therefore, this limitation has hampered the application of proteomics to study samples of limited availability, such as human tissues from, e.g., biopsies.

Against the backdrop of the shortcomings of traditional macroscale sample preparation, mostly inherited from the field of protein chemistry, methods for working with limited numbers of cells have recently been reported, including laser capture microdissection, immobilized enzyme reactors, fluorescence-activated cell sorting (FACS), and microfluidics formats^{85,123–128}. For example, using laser capture microdissection, Clair *et al.*¹²⁴ identified >3400 proteins from 4000 cells. However, none of these recent developments allowing the characterization of proteomes from fewer than 1000 cells by MS-based proteomics can yet be said to be the method of choice for exploring proteomes. Thus, there is burgeoning interest in the community to develop and optimize highly sensitive and specific microscale proteomic workflows to interrogate protein changes in both health and disease. Consequently, reports of micro- and nanoscale MS-based proteomics have dramatically increased in number lately because they facilitate new opportunities to explore trace levels of samples previously out of reach to researchers^{92,129–131}. For instance, an ultrasensitive

nanoscale method, which used gold nanoparticles, identified 650 proteins from a proteomic analysis of 80 cells with a detection limit of proteins reaching 50 zmol¹³². Additionally, many integrated proteome methods optimized for single-cell analysis are increasingly becoming commonplace^{133,134}. Some newly assembled proteome analysis devices have reported high numbers of identified proteins, e.g., 328 proteins identified from analysis of 10 single HeLa cells, and with a detection limit approximated to be between 1.7–170 zmol¹³⁵.

One new microfluidics-based platform termed nanodroplet processing in one pot for trace samples (nanoPOTS) that was developed recently has demonstrated remarkable results from the proteomic analysis of small samples⁸². By applying a bottom-up proteomic approach, nanoPOTS proved to be capable of processing samples in nanowells with volumes of less than 200 nL.⁸² This method was applied in the analysis of about 10 to 140 cells, and over 1500 proteins were confidently identified. Recently, nanoPOTS was also integrated with a top-down proteomic workflow, and ~170 to ~620 proteoforms from ~70 to ~770 HeLa cells were quantitatively identified with high confidence¹³⁶. An adaptation of nanoPOTS that utilizes conventional micropipettes and operates in low-microliter range called microdroplet processing in one pot for trace samples (microPOTS) has also been developed to address a few bottlenecks such as the demands for nanoliter pipetting platform and highly skilled personnel to run nanoPOTS. Initially, microPOTS was applied to ~25 cultured HeLa cells and 50 μm square mouse liver tissue thin sections, and about 1800 and 1200 unique proteins were generated from HeLa cells and mouse liver, respectively¹³⁷. Additionally, high reproducibility was reported based on pairwise Pearson's correlation coefficient values of 0.96–0.98, and with median CVs of $\leq 12.4\%$ from the results of the previously mentioned analysis.

In this study, we applied the microPOTS to characterize proteomes of ~200 cells used as Barrett's esophagus cell model following various perturbations. Barrett's esophagus is a premalignant condition thought to arise in the lower esophagus due to chronic reflux of gastric acid and bile leading to genotoxic stress and mutation of the gatekeeper genes TP53

and SMAD4¹³⁸. Barrett's confers an approximately 100-fold greater risk of development of esophageal adenocarcinoma, and understanding the molecular changes during carcinogenesis may be useful to guide preventative therapy¹³⁹. It is known that gut bacteria modify bile acids derived from cholic acid and chenodeoxycholic acid to deoxycholic acid and lithocholic acid (LCA)¹⁴⁰. In turn, deoxycholic acid and lithocholic acid are conjugated to yield a variety of conjugated bile salts that can exist in wide-ranging concentrations in patients being monitored for acid-reflux disease. It is not known whether there is a specific role for bile acids in the selection for specific genetic mutations in esophageal adenocarcinoma progression. Recent work has demonstrated a novel sponge-device (Cytosponge) can sample small numbers of surface cell populations from the esophagus without endoscopy and can determine the presence of Barrett's esophagus¹⁴¹. This device can also be used to triage patients with Barrett's for more intensive endoscopic surveillance according to the presence of markers of progression to esophageal adenocarcinoma¹⁴². As these devices become integrated into clinical practice the molecular changes during the progression from Barrett's to esophageal adenocarcinoma need to be identified from the small numbers of cells retrieved during sampling. Identifying protein markers of progression that can be tested by immunohistochemistry will aid in improving the sensitivity and translation of this technology. The findings of this study reveal that microPOTS allowed for the identification of >1500 proteins from fewer than 200 cells, and radiologic and physiologic stress induced proteomic changes in cell models.

2.3. Experimental Section

2.3.1. Materials

MicroPOTS chips were fabricated in-house as described previously¹³⁷. The microwell chips were designed with a diameter of 2.2 mm and a well-to-well spacing of 4.5 mm. LC-MS grade water and acetonitrile, formic acid (FA), iodoacetamide (IAA), and dithiothreitol (DTT) were purchased from Thermo Fisher Scientific (Waltham, MA). N-Dodecyl β -d-maltose

(DDM) was a product of Sigma-Aldrich (St. Louis MO). Both Lys-C and trypsin were purchased from Promega (Madison, WI).

2.3.2. Cell culture

CP-A cells were cultured in keratinocyte media (ThermoFisher) supplemented with human recombinant epidermal growth factor (rEGF), bovine pituitary extract (BPE), 1% penicillin/streptomycin (Invitrogen) and incubated at 37 °C with 5% CO₂. All other chemicals and reagents were obtained from Sigma unless otherwise mentioned. The guide RNAs targeting the p53 and smad4 genes to generate isogenic gene knockout cells are described in a separate manuscript. Briefly, the guide RNAs were either cloned into lentiCRISPRv2 transfer plasmid or procured as custom synthetic crRNAs from Integrated DNA Technologies (IDT), USA; tracrRNA was also manufactured by IDT. Cells were either transfected using attractene transfection reagent (Qiagen) or electroporated using Nucleofector Kit V (Amaxa, Lonza). After the electroporation, cells were transferred into a 6-well plate and allowed to recover for 3–5 days. The bulk population was single-cell isolated using flow cytometry and individual cells were deposited into 96-well plates using BD FACSJazz cell sorter. Individual colonies obtained were later replicated into 96-well plates and screened for successful gene deletion using immunocytochemistry against the p53 or Smad4. The clonal lines that stained negative for corresponding proteins were further expanded, and the loss of functional p53 and Smad4 protein was confirmed by immunoblotting against respective antibodies and Sanger sequencing for selected knockout clones using the primers flanking the gRNA cleavage site confirmed the genetic editing.

2.3.3. Cells treatment with LCA and X-ray

CPA wildtype, p53 null, and p53 Smad4 double null cell lines were cultured in triplicate and treated with 10 μM lithocholic acid (LCA) or DMSO as a control for 24 h or irradiated with 2 Gy X-ray and cultured for a further 24 h. Following the treatment, the cells were trypsinized, washed with PBS, counted and samples were prepared with 2 × 10⁵ cells in 50 μL PBS for each condition.

2.3.4. Proteomic sample preparation in microwells

To prepare samples for LC-MS analysis, 5 μL of 1% DDM and 0.5 μL of 500 mM DTT were added to 50 μL of sample, followed by incubation at 65 $^{\circ}\text{C}$ and 600 rpm for 1 h to lyse the cells and denature proteins. The cell lysates were diluted to 200 cells/500 nL with 50 mM ABC (pH 8.5), and 500 nL of cell suspension was pipetted into the microwells. Next, 500 nL of 10 mM IAA was added, and the samples were allowed to incubate in the dark at room temperature for 45 min. Two-step enzymatic digestion was applied by sequentially adding 500 nL of 10 ng/ μL Lys-C and 500 nL of 20 ng/ μL trypsin in Ammonium bicarbonate buffer, followed by incubation at 37 $^{\circ}\text{C}$ for 3 and 10 h, respectively. Thereafter, 500 nL of 5% FA was added, followed by incubation at RT for 1 h. The chips were stored in the humidified box sealed in a Ziploc bag at 4 $^{\circ}\text{C}$ until analysis.

2.3.5. LC-MS/MS analysis

A nanoPOTS autosampler was employed to introduce samples in microwells into LC-MS¹⁴³. The samples in microwell chips were extracted and loaded into a solid-phase extraction (SPE) column (4 cm long, 150 μm i.d., packed with 3 μm , 300 \AA C18 particles, Phenomenex, Torrance, CA, USA) using 100% buffer A (0.1% formic acid) delivered by a Dionex UltiMate NCP-3200RS pump. After sample loading, the concentrated peptides were separated using a 50 cm long, 50 μm i.d. nanoLC column with an integrated electrospray emitter (PicoFrit column, New Objective, Woburn, MA, USA). The LC column was packed in-house with the same C18 particles used for the SPE column described above. The LC flow rate was 150 nL/min. A 50 min linear gradient from 8% to 22% buffer B (0.1% formic acid in ACN) was used for peptide elution, followed by raising the gradient to 35% buffer B in 10 min to elute hydrophobic peptides. The column was then washed by flushing the column with 80% buffer B for 5 min. Finally, the column was equilibrated using 2% buffer B for 15 min before the next injection.

An Orbitrap Fusion Lumos Tribrid MS (ThermoFisher, San Jose, USA) operating in data-dependent acquisition mode was employed for peptide signal collection. To trigger

electrospray, a high voltage of 2200 V was applied at the metal union (Valco, Houston, USA) between the SPE column and LC column. The ion transfer tube was set at 200 °C for desolvation and the radio frequency of the ion funnel was set at 30% for optimal peptide transmission. For MS1 acquisition, an Orbitrap resolution of 120 000, an MS scan range from 375 to 1600, an AGC level of 1×10^6 , and a maximum injection time of 100 ms were used. Precursor ions with charges between +2 and +7, and intensity values over 1×10^4 were selected for HCD fragmentation and MS2 scanning. Precursors were isolated with an m/z window of 2 and fragmented by high energy dissociation (HCD) set at 30%. The fragment ions were transferred to Orbitrap for MS2 acquisition at a scan resolution of 60 000 and a maximum injection time of 118 ms. To reduce repeated sampling, an exclusion duration of 30 s and m/z tolerance of ± 10 ppm was applied.

2.3.6. Data and statistical analysis

The .raw files from LC-MS/MS were loaded in the MaxQuant software (v1.6.7.0) for analysis. Identification of peptides was performed using the built-in Andromeda to search against the reviewed UniProt human proteome database (2019 release with a total of 42 427 entries, where 20 350 were reviewed and 22 077 were unreviewed). All of the search parameters were used in their default setting. The enzyme was set to trypsin, and maximum missed cleavages set to 2, while fixed modification was set to carbamidomethylation of cysteines and a false discovery rate (FDR) at peptide-to-spectrum matches (PSM) and protein levels set to 0.01. The resulting .txt output files from MaxQuant were loaded into R statistical environment (v4.0.2) and preprocessed before analyzing with DEP (differentially enrichment analysis of proteomic data) package^{144,145}. The DEP package offers a robust and reproducible analysis workflow for MS-based proteomics data when determining differentially enriched/expressed proteins. DEP filters out contaminant and reverse protein sequences, logarithmically transforms the data, and then normalizes the data by variance stabilizing normalization (VSN) method. Subsequently, it imputes missing values and runs statistical tests to determine proteins with significantly altered expression levels. The latter step is

made feasible by the *test_diff* function, which performs a differential enrichment test based on protein-wise linear models and empirical Bayes statistics using limma^{146,147}. Data visualization was carried out using the BPG (v6.0.1)¹⁴⁸ and BioVenn (v1.0.2)¹⁴⁹ packages. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE¹⁵⁰ partner repository with the data set identifier PXD020741.

2.3.7. Protein annotation and assessment of physicochemical aspects

The sequences of the identified proteins and gene ontology (GO) for protein annotation regarding cellular components were accessed and retrieved from UniProtKB¹⁵¹. To calculate the grand average of hydropathy (GRAVY) value for the protein sequences, an online GRAVY calculator was used¹⁵².

2.4. Results and discussion

2.4.1. Analysing Barrett's oesophageal cell samples by microPOTS

The field of proteomics is rapidly changing with new technologies and methods advancing the capacity to examine a larger proportion of the proteome from small numbers of cells even down to the single-cell¹⁵³ level with a high degree of certainty. Over the past five years, the quest to design and develop highly sensitive and specific proteomic methodologies for this purpose has become intense, and many such methods have now been made available^{154–156}. Moving away from macroscale preparative techniques in proteomics to micro- and nanotechnologies offers researchers the ability to characterize smaller numbers of cells that are more likely to be encountered in clinical settings. The recently unveiled microPOTS system is one of the methodologies that are promising and poised to widen the window of possibilities in proteomics research¹³⁷. This study aimed at applying the microPOTS separation system to identify proteome signatures of either physiological stress or radiation in 200 Barrett's esophageal cells. Initially, CP-A p53 single null (CP-A KO) and CP-A p53-SMAD4 double null cells (CP-A dKO) were generated, and together with the parental CP-A wild-type cells (CP-A WT), were subsequently subjected to either LCA or

X-ray treatments. The microPOTS system was then applied for protein extraction and a nano-LC-MS workflow followed as depicted in **Figure 2-1**.

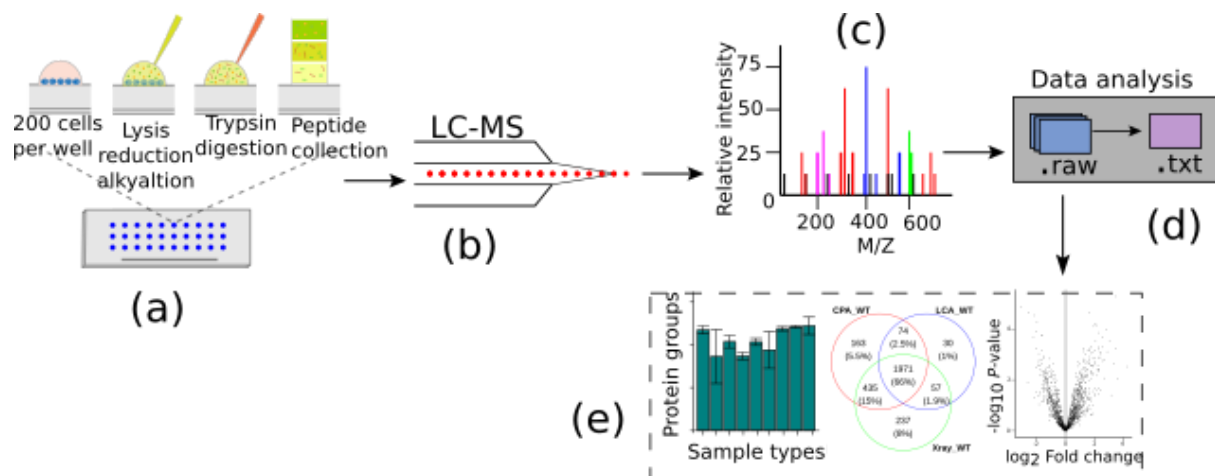


Figure 2-1. Overview of the Workflow of MicroPOTS for the Identification of Proteomic Changes in ~200 CP-A Cells after Treatment with Either LCA or X-ray. First, cells were grown in keratinocyte media, and the generated CP-A null cells and wild-type were subjected to different stressors as previously described. **a**, Samples were then processed for protein extraction and further digested into peptides using the microPOTS system. **b**, The collected peptides were subsequently subjected to LC-MS/MS analysis. **c**, A spectrum showing the relative intensity and mass-to-charge ratio (m/z) of the ions being analyzed. **d**, The resulting files were loaded into MaxQuant for peptide identification, after which the output files of this step were next imported into the R statistical environment and analyzed using the DEP package. **e**, The results of the analysis were then visualized using the previously mentioned R packages.

2.4.2. Protein identification from different CP-A genotypes and stress conditions

All the samples reported a high number of identified protein groups, with each replicate having over 1500 protein groups from only 200 cells. Replicates from each sample type were averaged and the mean value of protein groups was visualized using a bar graph (**Figure 2-2a**). The average number of confidently identified protein groups for each sample type was

as follows with the average number of proteins identified in parentheses: CP-A WT (2273), CP-A KO (1673), CP-A dKO (2008), CP-A LCA-WT (1685), CP-A LCA-KO (2004), CP-A LCA-dKO (1821), CP-A X-ray-WT (2300), CP-A X-ray-dKO (2345), and CP-A X-ray-KO (2367). The total number of proteins obtained from three replicates increased to over 2000 for all sample types when results were pooled from three wells for each sample type (**Figure 2-2b**). Next, an intersection analysis was performed to assess the number of shared and uniquely identified proteins. From this comparison, it was shown that over 1800 (60%) proteins could be identified in all sample types. Comparison of the total proteins from three replicates of CP-A LCA-WT, CP-A LCA-KO, and CP-A LCA-dKO produced a high protein overlap of 67% with 1867 proteins (**Figure 2-2c**). Out of these 1867 proteins, 93 (3.4%), 185 (6.7%), and 195 (7%) were unique to LCA-WT, LCA-KO, and LCA-dKO, respectively. The X-ray treatment equally reported a high overlap of 2354 (77%) proteins between the groups (**Figure 2-2d**). Also, the WT, KO, and dKO samples were overlapped with their corresponding treatment groups (LCA and X-ray) to determine the number of shared proteins and proteins that were unique to each sample. There was a high overlap of greater than 60% for the number of identified proteins for each sample type (**Supplementary Figure 2-1**). The number of identified proteins in this work is comparable to a previous study that performed proteomic analysis on small numbers of cells (~100) on an LTQ-Orbitrap system and identified ~1500 proteins¹⁵⁷.

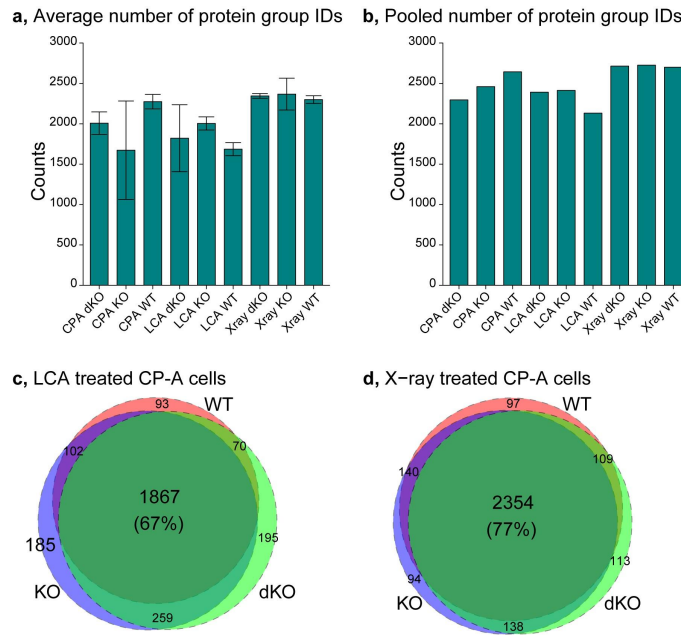


Figure 2-2. Barplots and Venn diagram plots respectively showing the number of identified protein groups and overlap for CP-A WT, CP-A KO, and CP-A dKO samples of ~200 cells. **a**, Average numbers of identified protein groups with each bar representing the mean and standard deviation (error bars) of replicates from individual sample types. **b**, Total numbers of identified protein groups from three replicates for each sample type. **c**, A 67% overlap of the identified proteins for samples treated with LCA. **d**, Samples treated with X-ray show a 77% overlap for the identified protein groups.

2.4.3. Evaluation of protein extraction efficiency and reproducibility

The results show that over 1500 proteins were confidently identified from 200 cells in all sample types. This high number of identified proteins from only 200 cells indicates the high extraction efficiency of the microPOTS system for microscale proteomic analysis. A look at the mean and standard deviations across the sample types reveals a consistency in the number of identified proteins. As illustrated in **Supplementary Figure 2-2**, almost all replicates had over 75% fully tryptic cleavage sites with fewer than 25% missed cleavages, indicating good tryptic digestion, which also translates into a high extraction efficiency.

The evaluation of microPOTS reproducibility in this study was predicated on two approaches—qualitative and quantitative analysis. Just like in the case of method comparison studies^{158–160} that often use this kind of approach, we decided to use the strategy to assess the performance of microPOTS. The reproducibility of the measurements is important in evaluation of the results and correct identification of proteins is critical to discovering new proteomic signatures with high certainty and specificity^{161–164}. Qualitative reproducibility was achieved by comparing the overlap of the identified proteins between sample types and illustrating the shared proteins in an area-proportional Venn diagram. In **Figure 2-3a**, results for LCA-treated CP-A dKO cells showed over 52% protein overlap between the replicates. Whereas for the X-ray-treated CP-A dKO cells, an overlap of 73% was reported between replicates. Additionally, protein overlap between replicates for the rest of the samples was assessed, and a high protein overlap of over 50% was reported for almost all comparisons (**Supplementary Figure 2-3**). Further, quantitative reproducibility between the replicates was assessed using the LFQ values of replicates to perform a pairwise Pearson's correlation coefficient analysis. As shown in **Figure 2-3b**, the quantitative assessment of reproducibility demonstrated a high Pearson's correlation coefficient value ($R > 0.9$). A high correlation coefficient ($R > 0.9$) was observed for almost all the pairwise comparisons that were conducted (**Supplementary Figure 2-4**). CP-A KO replicates reported a lower correlation coefficient ($R = 0.83$) relative to the rest of the samples. Reproducibility of the MicroPOTS system was further assessed by computing the coefficient of variation (CV) for individual protein intensities in each sample condition (**Figure 2-3c**). All samples except CP-A KO showed little variation with a median CV that is less than 50% (**Supplementary Table 2-1**).

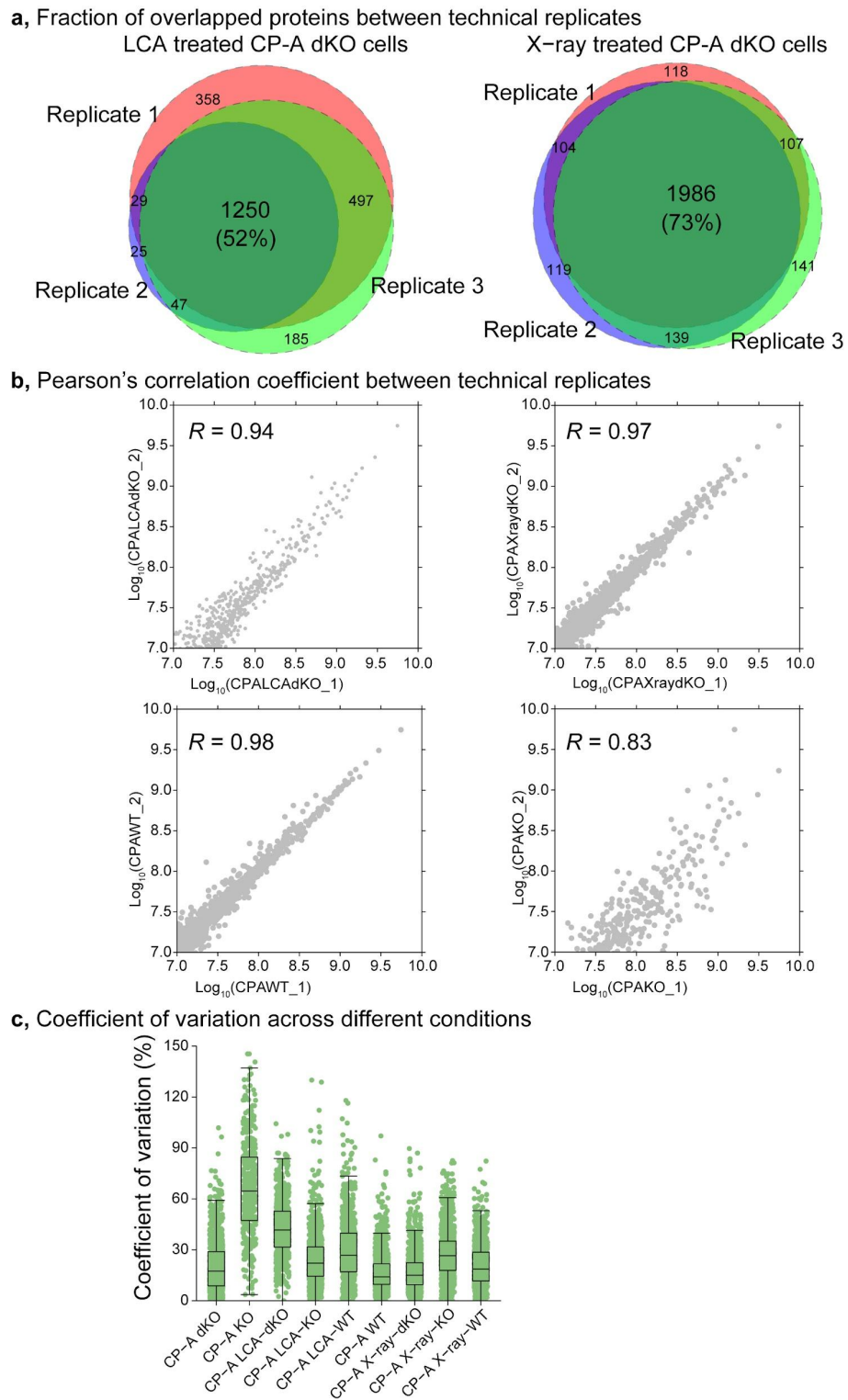


Figure 2-3. Qualitative and quantitative comparison of the identified proteins between replicates. **a**, The area-proportional Venn diagrams illustrate the proportion of shared and unique proteins between technical replicates. **b**, Scatter plots showing Pearson's correlation coefficient between logarithmically (\log_{10}) transformed quantile normalized

LFQ values of the corresponding replicates. *c*, A boxplot representing the distribution of coefficient of variation for quantile normalized protein intensities across different experimental conditions. Each dot represents the CV of an individual protein intensity within each sample type.

2.4.4. Comparison of microPOTS data to bulk proteomics dataset

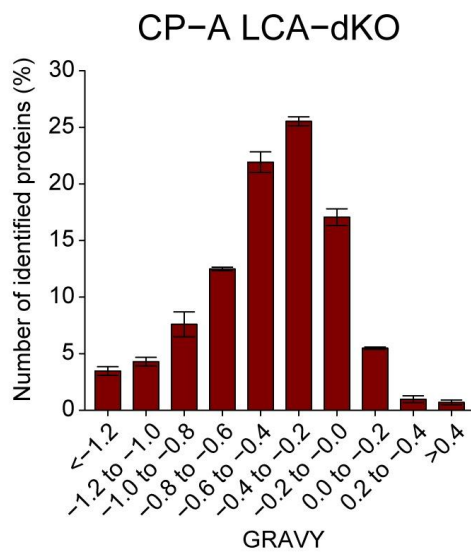
The microPOTS data were compared to an existing data set of the same cell line (CP-A). The samples that generated the bulk data were prepared and analyzed according to standard proteomic workflows as described in **Notes 2-1**. The resulting data were processed and analyzed using bioinformatics tools as stated in **Notes 2-2**. Next, we correlated all the overlapping proteins that were confidently identified between the two data sets using Pearson correlation method. As illustrated in the scatter plot in **Supplementary Figure 2-5**, there was a high positive correlation of $R = 0.625$.

2.4.5. Comparison of physicochemical aspects of identified proteins

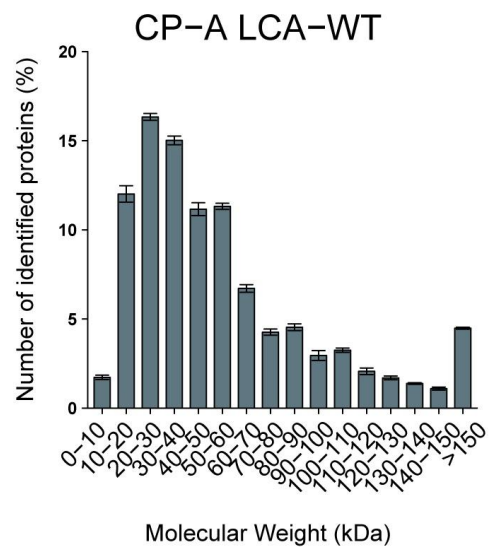
It has been established that high molecular weight (MW) and basic proteins are often challenging to extract due to their propensity to undergo intra- and intermolecular interactions^{165,166}. As such, we were interested in exploring the MW, and GRAVY distribution patterns of the identified proteins. The GRAVY distribution scores for LCA-treated CP-A dKO cells indicated that a high number of hydrophobic proteins were detected with microPOTS (**Figure 2-4a**). A similar distribution trend of GRAVY was noticed from all the sample types with the majority of identified protein IDs within the 0.4 to -0.2 range (**Supplementary Figure 2-6**). LCA-treated CP-A WT cells showed that the most abundant proteins had a MW between 20 and 30 kDa in all sample types (**Figure 2-4b**). Additionally, a similar MW distribution pattern was observed across all sample types, which demonstrates that the proteins show no major difference regarding their physicochemical characteristics (**Supplementary Figure 2-7**). Next, the distribution of identified proteins according to their subcellular localization was explored, which showed that over 50% of the identified proteins resided within the cytosolic region, and this was observed across all sample types. Nearly

20% of confidently identified proteins for all the sample types were located within the cytosol (Figure 2-4c). All samples revealed that few ribosomal proteins could be detected by the microPOTS system, and this finding is consistent with Zhu *et al.*⁸² reported in their study. Ribosomal and cytoskeleton-derived proteins were the least likely to be identified with a percentage of less than 5%. The distribution pattern for subcellular proteins was almost the same across all sample types (Supplementary Figure 2-8).

a, GRAVY distribution



b, Molecular weight distribution



c, Subcellular localization of identified proteins

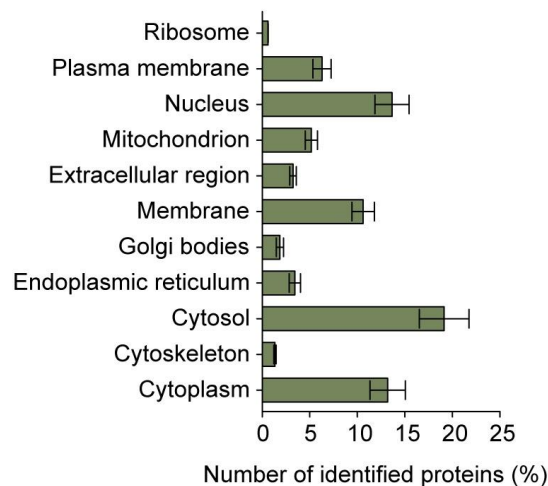


Figure 2-4. Assessment of the physicochemical features and subcellular localization of the identified proteins. Distribution of, **a**, GRAVY in CP-A LCA-dKO, **b**, molecular weight

in CP-A LCA-WT, and **c**, subcellular localization of identified proteins in CP-A X-ray-KO.

2.4.6. Effect of stress on protein expression

Differential expression levels of the identified proteins were determined between CP-A WT, CP-A KO, and CP-A dKO as well as between their corresponding treatment set (LCA and X-ray treated groups). A pairwise comparison for all sample types was carried out to evaluate differentially expressed proteins between cells treated with different stresses. The findings from the present study indicate that LCA and X-ray induced changes in the proteome of CP-A cells, which is also consistent with the findings of Proungvitaya *et al.* that reported bile acids-induced alteration of protein expression in model cells system¹⁶⁷. Significant alterations were observed in CPA-dKO cells following LCA and X-ray treatment. Specifically, a pairwise comparison between CP-A LCA-dKO and CP-A dKo revealed that 21 proteins were upregulated, and 13 proteins were downregulated, some of which include ASNS, RALY, CSRP1, CTSD, FAM120A, ESTD, and GSR (**Figure 2-4a**). Also, CP-A X-ray-dKO and CP-A dKO comparison reported 32 upregulated proteins with 14 downregulated proteins, including NONO, SAR1A, HNRL2, PLEC.1, FARSA, S100P, FKBP3, and AGR2 (**Figure 2-4b**). **Supplementary Table 2-2** and **Supplementary Table 2-3** represent the complete list of proteins that were significantly differentially expressed for LCA and X-ray-treated CP-A dKO cells, respectively. Anterior Gradient 2 (AGR2) is a member of the protein disulfide isomerase family, and its overexpression has been associated with many human cancers including neoplasia of the esophagus^{168,169}. This evidence is consistent with the present study, which has shown that AGR2 is overexpressed in Barrett's esophageal cells (CP-A X-ray-dKO). Having detected ~1500 proteins from fewer than 200 cells, and capturing differential expression, signify a potential use of the microPOTS-LC-MS method to explore subcellular populations within a tissue/tumor microenvironment including T-cells, fibroblasts, and macrophages.

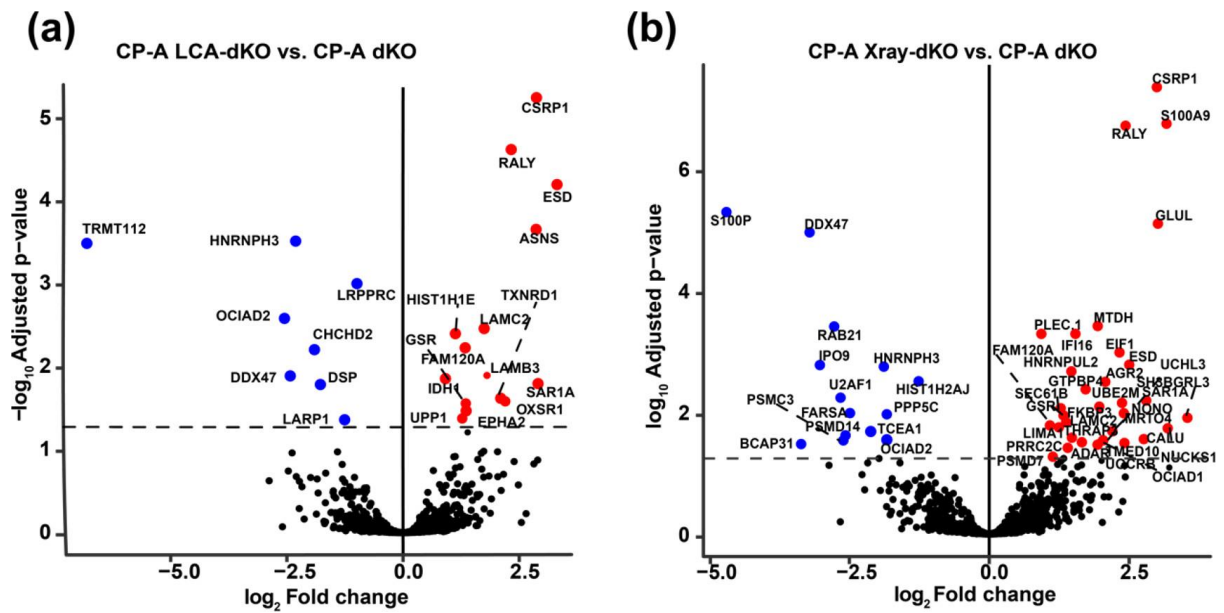


Figure 2-5. Volcano plots illustrating the fold difference in the expression levels of proteins in, **a**, CP-A LCA-dKO vs CP-A dKO and, **b**, CP-A X-ray-dKO vs CP-A dKO. The horizontal coordinate represents the log₂ fold change difference, while the vertical coordinate represents the $-\log_{10}$ p-value. Black color indicates proteins with nonsignificant changes in abundance, blue color represents significantly downregulated proteins, and red color represents significantly upregulated proteins.

2.5. Conclusion

In this study, a recently developed proteomic method called microPOTS was applied to identify proteins and determine the changes in the proteome of ~200 cells (including an isogenic cell panel being used for the Barrett's esophageal studies) following radiation and physiological stress treatment. The results show that the microPOTS method is applicable for use in qualitative and quantitative proteomic studies where only low cell numbers are available. Ionizing radiation was one stress used since it is DNA damaging and can activate p53 function. LCA was used since it is a component of bile acids that can impact on acid reflux disease and cancer progression in this tissue. The results were highly reproducible ($R > 0.9$) between replicates, allowing us to investigate confidently the effect of stress on the cells important in biological applications. With the microPOTS method ~1500 unique proteins

were quantified in all the samples. Moreover, results for the cells treated with LCA revealed differential expression analysis of 21 upregulated proteins and 13 downregulated proteins (CP-A LCA-dKO vs CP-A dKO), some of which include RALY, CSRP1, ASNS, ESTD, and FAM120A. Also, a comparison set between CP-A X-ray-dKO and CP-A dKO reported 33 significantly overexpressed proteins and 15 underexpressed proteins, including NONO, SKP1, HNRL2, PLEC.1, S100P, FKBP3, and AGR2. The results of the present study offer a basis for further studies to deeply interrogate in the future the molecular mechanisms that underpin LCA induction of proteome changes using clinical biopsies, which could aid in uncoupling the distinct role of bile acids in the selection for specific genetic mutations in esophageal adenocarcinoma progression. Importantly, the use of microPOTS, while not as sensitive as its companion technique nanoPOTS, was implemented and will be used in future studies.

CHAPTER 3: DIA-MS proteome analysis of formalin-fixed paraffin-embedded glioblastoma tissues

This work was originally published in *Analytica Chimica Acta* and is presented with a few modifications:

Weke, K.[†], Kote, S.[†], Faktor, J., Al Shboul, S., Uwugiaren, N., Brennan, P. M., Goodlett, D. R., Hupp, T. R., and Dapic, I. DIA-MS proteome analysis of formalin-fixed paraffin-embedded glioblastoma tissues. *Anal. Chim. Acta* **1204**, 339695 (2022). [†] equal contribution

3.1. Abstract

Developments in quantitative proteomics and data-independent acquisition (DIA) methodology are enabling quantification of proteins in biological samples. Currently, there are a few reports on DIA mass spectrometry (MS) approaches for proteome analysis of formalin-fixed paraffin-embedded (FFPE) tissues. Therefore, to facilitate detection and quantification of immune- and glioblastoma (GBM)-relevant proteins from FFPE patient materials, we established a simple and precise DIA-MS workflow. We first evaluated different lysis buffers for their efficiency in protein extractions from FFPE GBM tissues. Our results showed that more than 1700 proteins were detected and over 1400 proteins were quantified from GBM FFPE tissue microdissections. GBM-relevant proteins (e.g., GFAP, FN1, VIM, and MBP) were quantified with high precision (median coefficient of variation <12%). In addition, immune-related proteins (e.g., ILF2, MIF, and CD38) were consistently detected and quantified. The strategy holds great potential for routinizing protein quantification in FFPE tissue samples.

3.2. Introduction

Mass spectrometry (MS)-based proteomics is currently the last frontier for comprehensive characterization of proteins and their functional changes. The data-dependent acquisition (DDA) strategy has been primarily used to detect and quantify proteins for a long time. However, due to insufficient proteome coverage and reproducibility occasioned by the

stochastic sampling of the most intense peptide ions, the quantitative proteomics field is now shifting to data-independent acquisition (DIA) methodology^{107,110,170,171}. DIA-MS is based on a full scan of all precursors that are subsequently isolated and fragmented within a defined mass-to-charge (m/z) window until the scanning covers the entire m/z range of the initial full scan^{172–174}. DIA-MS has demonstrated high reproducibility and precision in protein quantification¹⁷⁵. It has been successfully applied so far on freshly frozen (FF) and formalin-fixed paraffin-embedded (FFPE) tissue samples (**Table 3-1**). Thus, DIA-MS is increasingly becoming popular in quantifying proteins in clinical samples^{111,176–179}.

Table 3-1. A brief highlight of tissue proteomics studies using DIA-MS or DDA-MS

| Reference | Tissue origin and amount | Buffer component | Column type | Mass spectrometer | # of identified proteins |
|--|--|---|-------------------------|-----------------------------|--------------------------|
| Marchione <i>et al.</i> ¹⁷⁹ | human FFPE liver, 1 mm thick | 5% SDS 100 mM Tris pH 8.5 | 20 cm x 75 μ m i.d | Q Exactive HF-X, (DIA-MS) | 3,297 |
| Hou <i>et al.</i> ¹⁷¹ | human FF esophageal squamous cell carcinoma | 2% SDS, 7 M urea, 10 mM EDTA, 0.1 M Tris-HCl, pH 7.6 | 50 cm x 75 μ m i.d | TripleTOF 5600+, (SWATH-MS) | 1,758 |
| Kim <i>et al.</i> ¹¹¹ | human FF colorectal cancer tissue, 1 μ g | Liquid Tissue buffer | 10 cm x 100 μ m i.d | TSQ Quantiva, (DIA-MS) | 3,713 |
| Gao <i>et al.</i> ¹⁷⁸ | human hepatocellular carcinoma tissue, 0.2 g | 50 mM Hepes, 6 M urea, 2 M thiourea, 1 x protease inhibitor | 15 cm x 75 μ m i.d | TripleTOF 5600+, (SWATH-MS) | 4,216 |
| Pirog <i>et al.</i> ¹⁸⁰ | human GBM, 15 μ m thick | 100 mM ABC, 30% ACN (PCT) | 21 cm x 75 μ m i.d | TripleTOF 5600+, (DDA-MS) | 998 |
| | | 1% SDS, 0.1 M Tris pH 7.6 (FASP) | | | 700 |
| | | 8 M urea, 30% ACN, 100 mM ABC (ISD) | | | 1,233 |

| | | | | | |
|-----------------------------------|---------------------------|--|-------------------|------------------------------------|-------|
| Föll <i>et al.</i> ¹⁵⁹ | mouse kidney, 10 µm thick | 0.1% RapiGest, 0.1 M HEPES pH 8, 1 mM DTT (DT) | 15 cm x 50 µm i.d | Orbitrap Q-Exactive plus, (DDA-MS) | 1,841 |
| | | 4% SDS, 0.1 M HEPES pH 7.5, 0.05 M DTT (FASP) | | | 1,857 |

HEPES - 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid.
Tris - Tris(hydroxymethyl)aminomethane.

However, the application of DIA-MS on FFPE is still limited compared to DDA-MS. FFPE samples from diagnostic laboratories are of particular interest for retrospective studies since they have considerable follow-up data useful for correlating observed clinical outcomes with patient-specific proteome. In addition, formalin fixation and paraffin embedding is a routine method for tissue preservation, and regular tissue collection and long-term stability make them easily accessible for large cohort studies¹⁸¹. Therefore, effective proteomics analysis of FFPE is critical to the study of clinically-relevant proteins and the discovery of novel protein biomarkers for early disease detection and prediction of clinical outcomes. Recently, the use of DIA-MS on clinical FFPE samples facilitated tumor stratification and enabled the discovery of potential biomarkers for prostate cancer¹⁸². While FFPE specimens remain more appealing for retrospective investigations than FF tissues, they present a challenge for proteomics analysis due to introduced protein crosslinks during the fixation process. Thus, effective protein extraction from the FFPE tissues is crucial for successful analysis.

Effective extraction of proteins from FFPE samples requires reversing the formaldehyde-induced crosslinks. Some of the methods described used heat-induced antigen retrieval (HAIR)^{183,184} to achieve this. Numerous protocols have been published for protein extraction from human tissues for subsequent LC-MS analysis. Extraction buffers are used with the main goal of successfully solubilizing and denaturing proteins. Some of the buffers as primary amine-containing buffers, such as ammonium bicarbonate (ABC) are

often supplemented with chaotropes (e.g., urea or guanidine hydrochloride), organic solvents (e.g., acetonitrile (ACN) or trifluoroethanol), and/or detergents (e.g., RapiGest, sodium dodecyl sulfate or sodium deoxycholate) all to achieve good protein extraction^{125,158,159,180,185,186}. In Table 1, we have briefly highlighted several protein extractions and digestion protocols, including in-solution digestion (ISD)^{79,187}, direct trypsinization (DT)^{158,159,188}, filter-aided sample preparation (FASP)^{185,189}, and single-pot, solid phase-enhanced sample preparation (SP3)¹⁹⁰ for FFPE tissue proteomics. To date, protein retrieval with ISD protocols and optimized chemical buffers is a typical workflow in tissue proteomics. Therefore, a critical assessment of chemical buffers is crucial prior to establishing the most effective workflow for MS-based proteome analysis of FFPE tissue samples.

Herein we present a study focused on establishing a simple DIA-MS workflow for detecting and quantifying immune- and GBM-relevant proteins in patient-derived FFPE microdissected tissue samples. GBM is cancer with unmet clinical needs with a median survival rate of less than two years⁴⁶. In addition, it exhibits high molecular heterogeneity^{12,191}, which complicates its biology. Therefore, to ramp up studies on complex tumors such as GBM, simple, highly precise, and easily implementable proteomics methods are necessary. Here, we compared the efficiency in protein extraction of three different buffers to find the best conditions for protein isolation and digestion. Consequently, we determined if single-shot DIA-MS using ~10 mm² (15 μm thick) GBM FFPE tissue sections would precisely quantify the expression levels of relevant proteins. As a result, we detected and quantified GBM-relevant proteins such as fibronectin 1 (FN1), vimentin (VIM), and myelin basic protein (MBP) from the DIA-MS data. Additionally, immune-relevant proteins such as C3, CD47, CD38, interleukin enhancer-binding factor 2 (ILF2), and macrophage migration inhibitory factor (MIF) were also quantified. Thus, this approach offers novel opportunities for accelerating biomarker discovery using patient-derived FFPE samples.

3.3. Materials and methods

3.3.1. FFPE tissue samples

Tissue samples used in this study were acquired from human GBM patients (n = 5) according to local ethical approvals (University of Edinburgh, Edinburgh Cancer Research Centre, reference number 06/S1101/16). All ethical requirements regarding using human specimens for research were taken into account. Informed consent was obtained from patients, and the Royal Edinburgh Hospital Ethics Review Committee approved the study.

3.3.2. Sample collection

GBM FFPE tissue samples were frozen in liquid nitrogen within 60 min of surgical removal and stored at $-80\text{ }^{\circ}\text{C}$. Each FFPE sample was evaluated by a neuropathologist to confirm the GBM diagnosis. A set of five GBM tissues were further processed. Three $4\text{ }\mu\text{m}$, $10\text{ }\mu\text{m}$, and $15\text{ }\mu\text{m}$ thick FFPE tissue sections were retrieved from each block of GBM tissue.

3.3.3. Sample processing

Glass-mounted FFPE tissue sections were initially deparaffinized and rehydrated prior to applying the various extraction protocols used in the study. FFPE tissue was deparaffinized by using two xylene washes for 2 min each. ACN, ethanol, and water were acquired from VWR-Avantor (PA, USA). Other chemicals, unless otherwise stated, were obtained from Sigma (MO, USA). Tissue rehydration was performed in ethanol series from 100% to 85% and a final 70%, each for 2 min, and followed by placing the glass slide in LC-MS grade water twice for 2 min each. The deparaffinized and rehydrated tissue sections were then scratched off from the glass slide using a clean razor and put into a clean LoBind Eppendorf tube (Merck, Darmstadt, Germany).

Method 1 - (ACN-containing buffer with ISD). Here, the extraction buffer consisting of $50\text{ }\mu\text{L}$ of 30% ACN and 100 mM ABC was added to the samples. The sample-extraction buffer mixture was incubated at $95\text{ }^{\circ}\text{C}$ for 90 min followed by cooling to $37\text{ }^{\circ}\text{C}$. Disulfide bonds were reduced by adding $2.8\text{ }\mu\text{L}$ of 700 mM dithiothreitol (DTT) to the sample-buffer mixture and

then incubated at 37 °C for 30 min. Alkylation was performed by adding 9.2 µL of 700 mM iodoacetamide (IAA) to the mixture and incubating for a further 30 min at 37 °C in the dark. After incubation, samples were diluted by adding 120 µL of 100 mM ABC and 880 µL of LC-MS grade water. Protein digestion was carried out by adding trypsin (Promega, MA, USA) to each tissue section in a ratio of 5 ng/mm², and incubating at 37 °C for 17 h. Digestion was quenched by adding 50 µL of 5% trifluoroacetic acid and vortexed for 30 s.

Method 2 - (ACN-urea-containing buffer with ISD). For this protocol, 50 µL of 30% ACN, 100 mM ABC were added to the samples, and the mixture was incubated for 90 min at 95 °C. Subsequently, the sample-buffer mixture was cooled down to 37 °C after which, 50 µL of 8 M urea was added. Next, 5.6 µL of 700 mM DTT was added to the mixture, followed by a 30-min incubation at 37 °C. For alkylation, 18.4 µL of 700 mM IAA was added, and the sample was incubated for 30 min at 37 °C in the dark. Samples were diluted by adding 120 of 100 mM ABC and 880 µL of LC-MS grade water. Protein digestion was conducted as previously mentioned.

Method 3 - (ACN-urea-RapiGest-containing buffer with ISD). Here, 50 µL of 30% ACN, 100 mM ABC, 45 µL of 8 M urea, and 5 µL of 1% RapiGest were added to the samples, and the mixture was incubated for 90 min at 95 °C. Reduction of disulfide bonds was achieved by adding 5.6 µL of 700 mM DTT to the samples and subsequently incubating for 30 min at 37 °C. Alkylation was accomplished by adding 18.4 µL of 700 mM IAA followed by a 30-min incubation at 37 °C in the dark. Samples were diluted by adding 120 of 100 mM ABC and 880 µL of LC-MS grade water. Trypsinization was carried out as already mentioned.

The resulting peptides were desalted on Waters Sep-Pak C18 cartridges (Fisher Scientific, NH, USA). Peptides were desalted using 5% methanol with 0.1% TFA and were subsequently eluted using 50% ACN with 0.1% TFA. Peptide eluates were vacuum dried using a SpeedVac concentrator (Thermo Scientific, MA, USA) and were then stored at -80 °C until LC-MS/MS analysis.

3.3.4. Liquid chromatography-mass spectrometry (LC-MS/MS) analysis

LC-MS/MS analysis was performed using Ultimate 3000 RSLCnano (ThermoFisher Scientific, MA, USA) coupled to Orbitrap Exploris 480 mass spectrometer (ThermoFisher Scientific, MA, USA). The dried peptides were reconstituted in 30 μ L of loading buffer (0.08% TFA in 2.5% ACN). Afterward, the peptide concentration was measured using NanoDrop 2000 (Thermo Scientific, MA, USA) absorbance at 280 nm. About 800 ng of peptides for each sample were injected into the LC system and analyzed in 3 technical replicates. First, peptides were trapped and concentrated on an Acclaim™ PepMap™ 100C18 trap column (ThermoFisher Scientific, MA, USA) of 300 μ m i.d x 5 mm, and packed with 5 μ m, 100 Å particles. Subsequently, 75 μ m i.d x 15 cm Acclaim™ PepMap™ RSLC 100 reverse phase C18 analytical column packed with 2 μ m particles having 100 Å pore size, was used to separate peptides at a flow rate of 300 nL/min. The analytical column was first equilibrated for 10 min at 2.5% solvent B (0.1% (v/v) TFA in acetonitrile) and 97.5% of solvent A (0.1% (v/v) TFA in water). Following, to separate the peptides, solvent B was increased in a 90-min gradient between 2.5% and 40%. A post-gradient cycle of 99% B was followed by a post-run equilibration at 2.5% B. The analytical column was coupled to the Nanospray Flex Ion Source. Recording of all MS measurements and spectra was done in resolution positive ion mode with a high voltage of 2500 V. For DDA mode, the ion transfer tube was heated to a temperature of 250 °C. MS1 data was acquired by setting the Orbitrap resolution to 120000, an MS scan range of 350–1200 m/z, an AGC (automatic gain control) level to custom mode, and a maximum injection time mode to auto. Precursor ions with charges between +2 and +6 and intensity values over 5.0e3 were selected for subsequent higher-energy C-trap dissociation (HCD) fragmentation and MS2 scanning. Precursors were isolated with an m/z window of 2 and fragmented by HCD set at 30%. The product/fragment ions were channeled to the Orbitrap for the MS2 acquisition at a scan resolution of 15000 and a maximum injection time mode set to 60 ms. Repeated sampling was avoided by setting the dynamic exclusion mode to custom, exclusion after n times to 1, exclusion duration to 20, and mass tolerance to 10 parts per million (ppm). For data-independent acquisition (DIA) mode,

precursor mass range was 400–1100, isolation window (m/z) of 8, window overlap (m/z) of 1, number of scan events set 87, collision energy of 30%, and in a fixed mode, orbitrap resolution of 15000, and the scan range mode was set to auto. In addition, the AGC target was set to custom mode, the maximum injection time set to 25 ms, and microscan was set to 1.

3.3.5. MS data analysis

Mass spectrometer DDA data were processed with MaxQuant¹⁹² (v1.6.17.0). A reviewed UniProt human proteome database with 20381 protein entries (download on March 4, 2021) was used. MaxQuant data processing was done as previously described¹⁹³, with a few modifications. The fixed modification was set to carbamidomethylation of cysteines and the false discovery rate (FDR) for both protein and peptide-to-spectrum matches (PSM) levels was set at 0.01. A spectral library was generated from DDA-MS data using the FragPipe (v14.0) workflow that comprises MSFragger¹⁹⁴ (v3.1.1), Philosopher (v3.4.13), and Python (3.8.5). The generated library consisted of 2058 protein isoforms, 2058 protein groups, and 17001 precursors in 14702 elution groups. DIA-MS RAW files were processed with Data-Independent Acquisition by Neural Networks (DIA-NN) software (v1.7.15)¹¹³. DIA-NN was operated in both library and library-free modes. The gene-centric output matrix was used for further analysis in an R statistical environment (v4.1.0). Protein intensities values for the processed DDA and DIA data were normalized by the median method followed by logarithmic (base 10) transformation. Custom R scripts were used for creating visualizations of the processed data.

3.3.6. Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE¹⁵⁰ partner repository with the data set identifier PXD029497.

3.4. Results

3.4.1. Comparative evaluation of different lysis buffer compositions

To establish a robust, simple, and precise FFPE tissue DIA-MS proteomics workflow, we first evaluated the efficiency of three ISD-based protocols (Method 1, Method 2, and Method 3) for protein extraction and digestion. Firstly, lysis buffers were used to analyze the samples of GBM FFPE tissue sections of varying thicknesses (4 μm , 10 μm , and 15 μm) using DDA-MS (Figure 3-1)

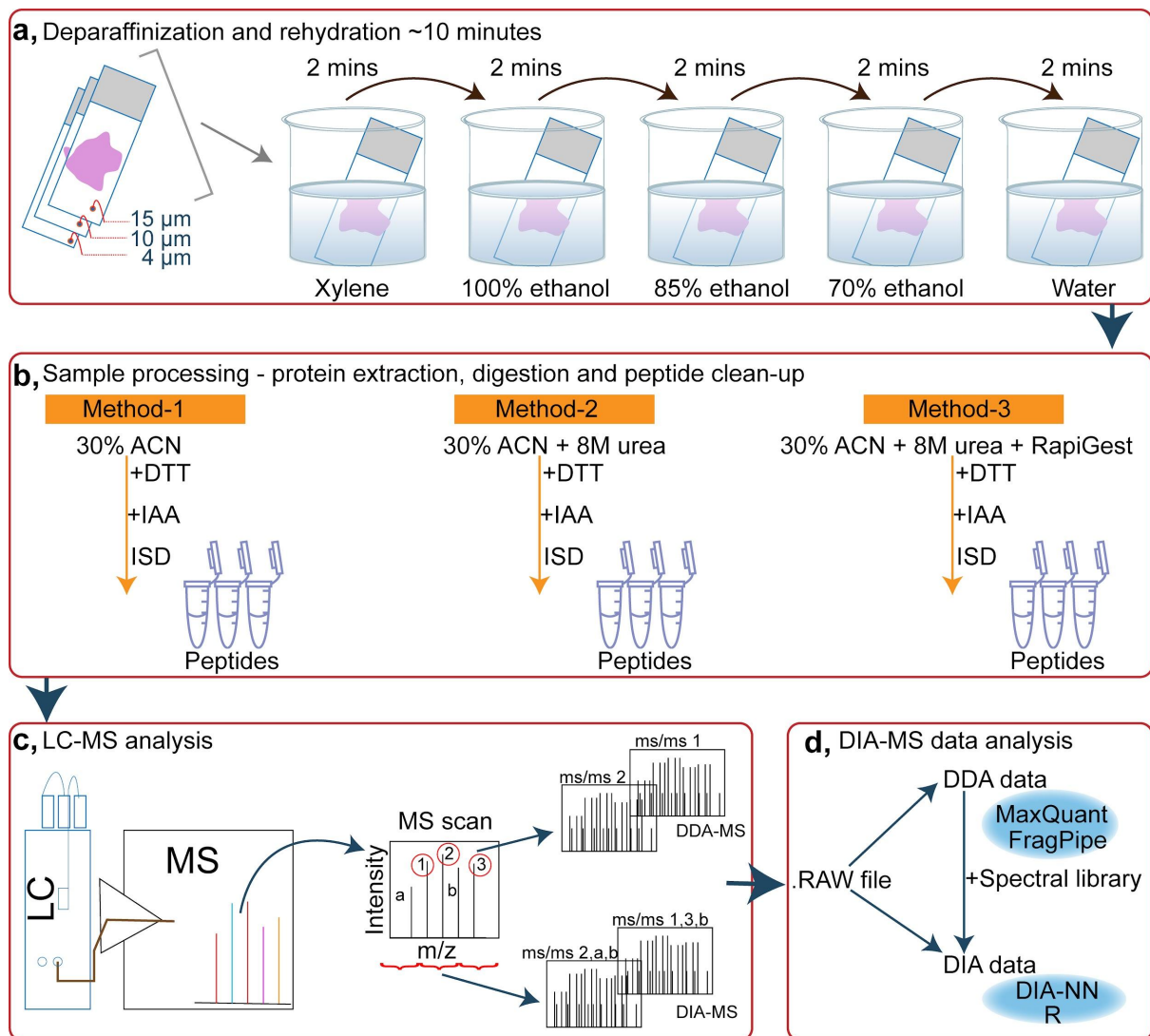


Figure 3-1. Method evaluation and DIA-MS analysis of GBM FFPE tissues workflow. *a*, Deparaffinization and rehydration of FFPE tissue sections. The 4 μm , 10 μm , or 15 μm thick GBM tissue sections were first deparaffinized by two xylene washes

followed by rehydration using a gradient of ethanol washes (100%, 85%, and 70%) and a final wash with LC-MS grade water. b, Protein extraction, digestion, and peptides clean-up. Tissues were scrapped into the Eppendorf tubes and three methods were each applied to process the deparaffinized and rehydrated tissue sections. c, LC-MS analysis of the resulting peptides were analyzed by Ultimate 3000 RSLCnano coupled to Orbitrap Exploris 480 mass spectrometer. d, DIA-MS data analysis using DIA-NN with MSFragger-generated spectral library.

The number of identified proteins was compared between the methods and the results are shown in **Supplementary Figure 3-1a**. All the methods showed to be sensitive enough to detect more than 1000 proteins from 4 μm tissue microdissection. More specifically, with Method 1, Method 2, and Method 3, we have detected 1274 ± 159 , 1405 ± 15 , and 1311 ± 171 proteins, respectively (results are shown as mean \pm standard deviation [SD]). Moreover, we observed a trend showing urea-containing buffers had higher numbers of protein identifications than the ACN only buffer, which agrees with our previous study on FF uterus tissue¹⁹⁵. For 10 μm thick sections, 1356 ± 38 , 1303 ± 75 and 1347 ± 6 proteins were identified by Method 1, Method 2, and Method 3, respectively. A similar trend was witnessed with 15 μm thick sections. The average number of detected proteins by all methods for 4 μm (1330 ± 68) thick tissue was comparable to 10 μm (1335 ± 28) or 15 μm (1337 ± 10) thick sections implying that small tissue thickness as 4 μm can be used. However, a somewhat higher SD value for 4 μm tissue compared to 10 μm and 15 μm tissue microdissections might be related to the decreasing amounts of input material for ISD proteomic analysis¹⁹⁵. In our further work, we focused on protein quantification from DDA measurements, and results showed that slightly less than 1000 proteins on average in any of the measurement groups were quantified (**Supplementary Figure 3-1b**).

Next, we overlapped proteins identified by individual methods for all tissue thicknesses to determine the proportion of common proteins. Results from 3 technical replicates were first combined and unique IDs in each group were overlapped. A high percentage (>90%)

overlap was observed with 4 μm , 10 μm , and 15 μm reporting, 91%, 95%, and 95% of shared proteins, respectively (**Supplementary Figure 3-1c**). This demonstrates that almost all the detected proteins in each case were similar, indicating a similarity in the performance of tested methods. Consequently, the high percentage of shared proteins among the methods has shown a similar distribution pattern for gene ontology (GO) terms and physicochemical characteristics of the identified proteins (**Supplementary Figure 3-1d**). Further, we selected a few (20) GO terms previously shown to have differential regulation in GBM^{196,197}. The highest percentage of the top five cellular components are cytosolic, followed by the extracellular exosome, cytoplasmic, nuclear, and plasma membrane. As illustrated in the inset in **Supplementary Figure 3-1d**, a median molecular weight (MW) of ~42 kDa was observed across the comparison groups. Further, in the evaluation of missed cleavage sites, methods showed similarity in the number of missed cleavages, and over 60% of cleavage sites were fully cleaved in all the comparison groups (**Supplementary Figure 3-2**). Overall, Method 3 (25.1%) showed a slightly smaller proportion of missed cleavages relative to Method 1 (25.3%) and Method 2 (27.5%) — indicating a beneficial effect of including RapiGest.

3.4.2. Quantitative analysis and evaluation of precision

Quantitative analysis is an important aspect of biomarker discovery and presents a particular analytical challenge in DDA strategies due to missing intensities values. Thus, we focused on evaluating the precision of quantification methods in each tissue thickness category. The data was quantile-normalized before performing logarithmic transformation (base 10). We correlated technical replicates using a pairwise Pearson's correlation coefficient to determine quantitative reproducibility between sections processed with the same method. We observed a high Pearson's correlation ($R > 0.90$) between technical replicates in all the methods (**Figure 3-2a**). We subsequently carried out a coefficient of variation (CV) analysis to discern methods that allowed minimal fluctuations in protein quantification. CV analysis was applied to the log-transformed data and the resulting values were expressed as percentages. From

this, we noticed less variability in protein quantification in almost all the methods as we could observe CV values of <20% (**Figure 3-2b**). For 4 μm thick sections, we could observe 17.3%, 16.2%, and 23.8% CV values in Method 1, Method 2, and Method 3, respectively. In 10 μm thick sections, 11.5%, 14.0% and 10.5% CVs corresponded to Method 1, Method 2 and Method 3. As for the 15 μm thick sections, we found 17.5%, 12.0%, and 15.3% with Method 1, Method 2, and Method 3, respectively. High precision was witnessed with Method 2 and 3 where 10 μm or 15 μm thick sections were used. We then calculated the mean CVs in each thickness category by summing up CVs for the three methods and averaging. Overall, the 4 μm thick sections comparatively reported the highest variation with a mean CV of 19.1%, which is likely to be related to the low input material. Mean CV values of 12% and 14.9% were recorded for 10 μm and 15 μm thick sections, respectively.

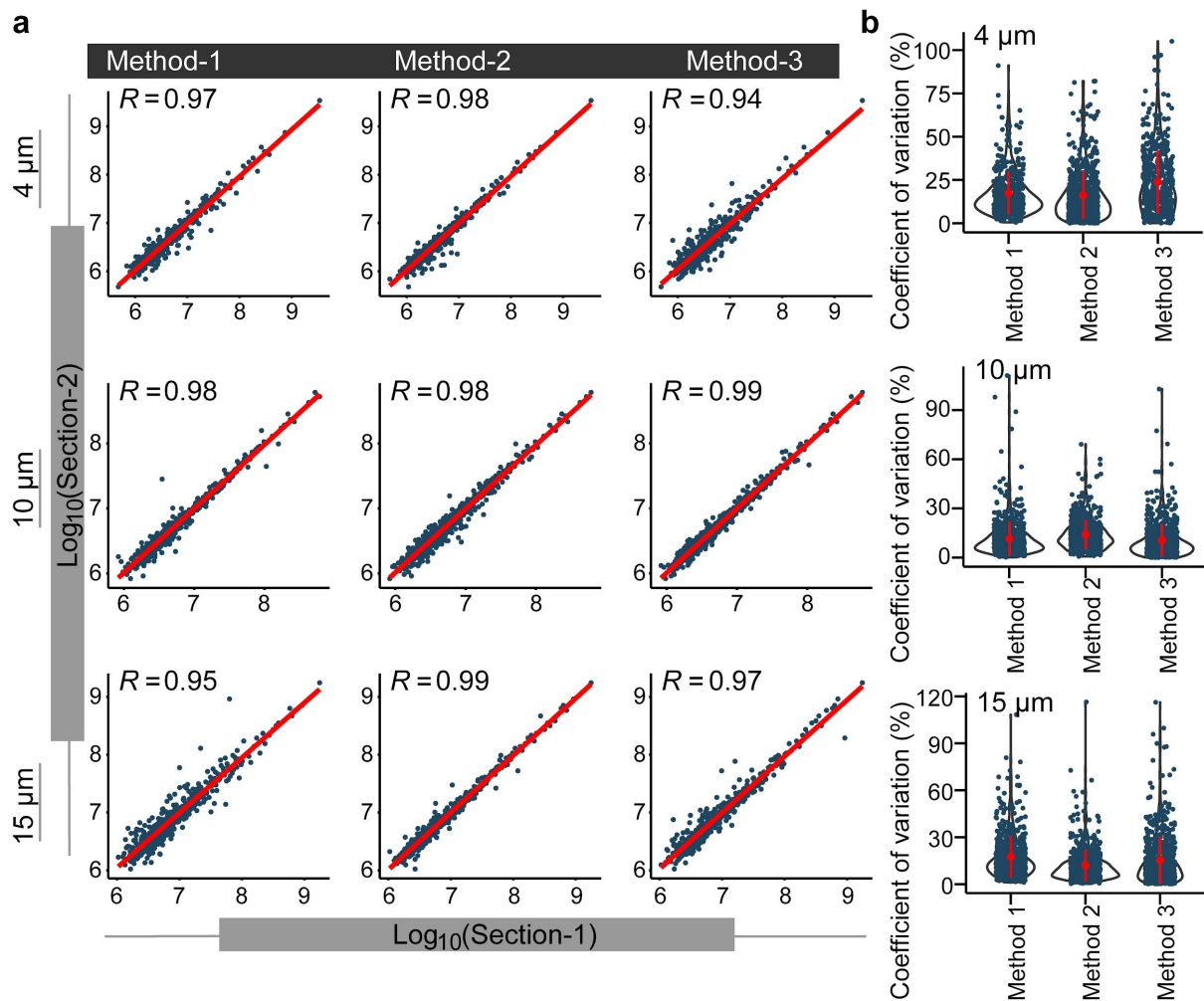


Figure 3-2. *Quantitative reproducibility assessment. a, Scatter plots illustrating the correlation between technical replicates processed by each of the three methods. The LFQ values were normalized using the quantile method, and then log (base 10) transformed. Each dot in the plot represents a protein, and the red line indicates the regression line. The Pearson's correlation coefficient value is indicated by R. Each column displays the scatter plot for the individual method e.g., from left-to-right: Method 1, Method 2, and Method 3. Each row displays the scatter plot for different sizes of tissue sections e.g., from top-to-bottom: 4 μm , 10 μm , and 15 μm . b, Violin plots showing CV for each method using different tissue sizes.*

3.4.3. Comparison of sequence coverage

Sequence coverage represents the average percentage of amino acids covered in the identified protein. The type of strategy applied to prepare samples for mass spectrometric proteomic analysis can influence the sequence coverage of the identified proteins¹⁹⁸. We compared the sequence coverage as a function of protein abundance for the covered proteome by each method (**Figure 3-3a**). The sequence coverage and intensity values for technical replicates analyzed by the same method were averaged to obtain mean values. In 4 μm sections, all methods performed similarly in terms of sequence coverage of the proteins (Method 1 = 14.3%, Method 2 = 14.1% and Method 3 = 14.8%). We, however, observed an increasing trend in the percentage sequence coverage in the case of 10 μm and 15 μm thick tissue compared to 4 μm thick sections. For 10 μm thick sections, the coverage values were 18.4%, 16.5%, and 17.5% for Method 1, Method 2, and Method 3, respectively. On the other hand, the 15 μm thick tissue sections reported the following coverage values; 17.5%, 18.4%, and 17.9% for Method 1, Method 2, and Method 3, respectively. The difference in the sequence coverage between relatively smaller sections (4 μm) and larger (10 μm or 15 μm) is likely attributable to the amounts of the input material. Interestingly, this suggestion was perfectly demonstrated with the urea-RapiGest-containing buffer (Method 3) as we could see an increase in the percentage of sequence coverage with

14.8%, 17.5%, 17.9% reported for 4 μ m, 10 μ m and 15 μ m thick sections, respectively. Next, we were interested in comparing the number of proteins having percentage sequence coverage greater than 20%. From this comparison, we could observe that Method 2 and Method 3 showed the number of proteins having a sequence coverage greater than 20% increase with increasing tissue size from 4 μ m to 15 μ m.

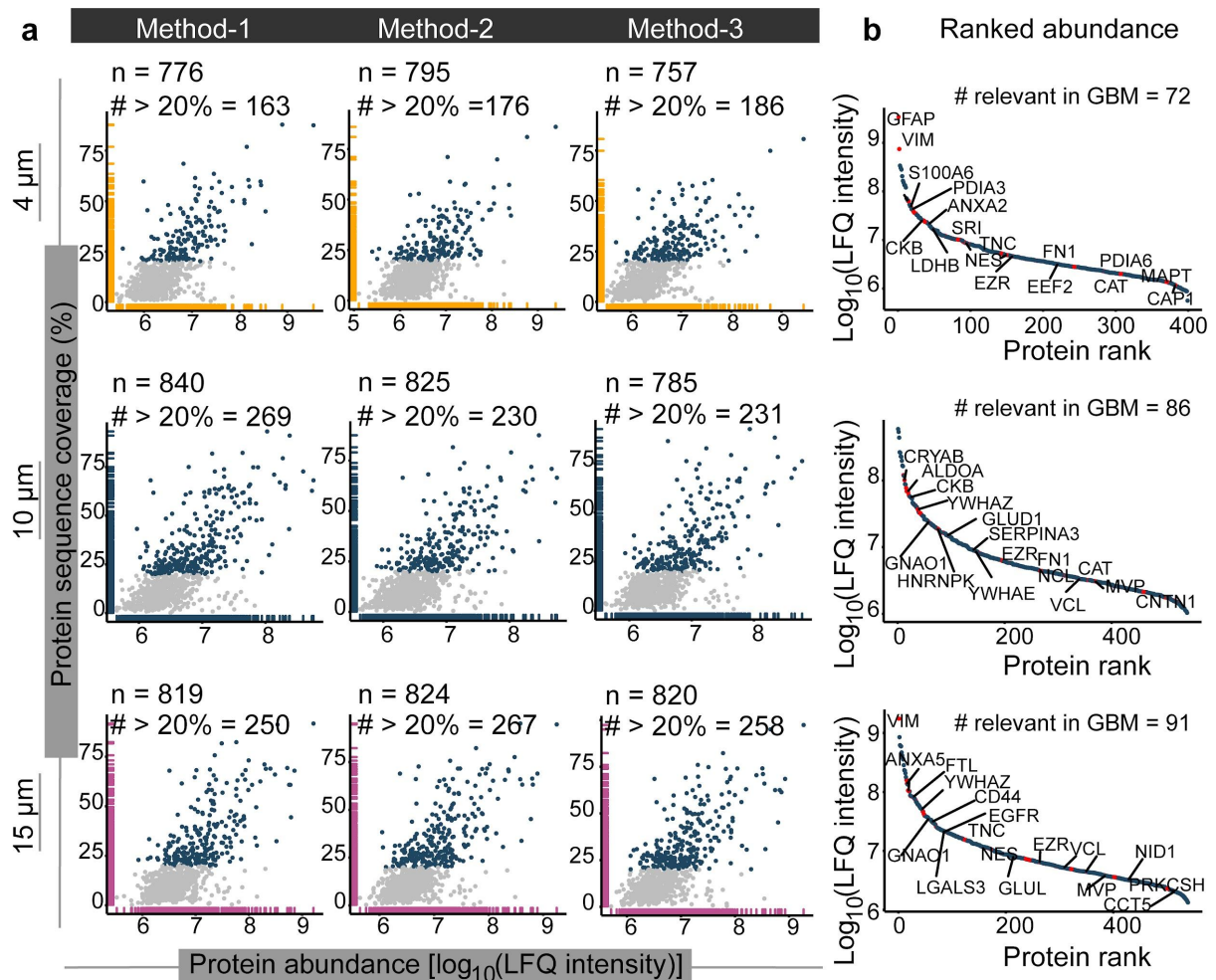


Figure 3-3. Sequence coverage and ranked abundance across the methods. **a**, Scatter plots of the percent of mean protein sequence coverage (y-axis) and quantile-normalized and logarithmically (base 10) transformed mean LFQ values (x-axis). Each column represents the method type (Method 1, Method 2, and Method 3 successively) and each row represents the size of the tissue section (4 μ m, 10 μ m, and 15 μ m successively). Median protein sequence coverage averaged at 20%, and thus, proteins detected with sequence coverage above this value were considered highly

*covered by peptides. The gray color indicates proteins with sequence coverage less or equal to 20% while cello color indicates proteins with sequence coverage above 20%. A density trace marked with a rug on both axes shows the spread of values along the coordinates, and the color codes of the rug correspond to the different tissue sizes. The number of complete observations in all replicates for each scatter is shown by n . The number of proteins with sequence coverage greater than 20% is shown by the # symbol. **b**, Dynamic range plot based on quantile normalized and logarithmically (base 10) transformed mean LFQ values of consistently identified proteins in all methods for each size of tissue section (4 μm from top, 10 μm middle and 15 μm bottom) used. The x-axis shows the ranked order distribution. The red color indicates proteins that have been reported as potential biomarkers in GBM.*

For the three thickness categories, we examined them based on proteins relevant to GBM that were covered. At first, we cataloged most proteins that have been reported as clinically relevant or candidate biomarkers for GBM. The raw (LFQ intensities) protein abundance values were quantile-normalized and subsequently, log (base 10) transformed, and the outputs were ranked. Finally, the ranked lists were intersected with the list containing GBM-relevant proteins. We noticed an increase in the number of GBM-relevant proteins with increasing tissue thickness, e.g., 72, 86, and 91 for 4 μm , 10 μm , and 15 μm , respectively (**Figure 3-3b**).

Having performed a comparative evaluation of the three methods, we selected Method 3 (ACN-urea-RapiGest-containing buffer) and applied it to analyze five patient samples. Even though all the methods demonstrated comparable performance we could witness some consistency in protein identification and quantification with Method 3, which informed the decision to select it for the next step in our DIA-MS analysis workflow.

3.4.4. DIA-MS reliably detects and quantifies putative GBM biomarkers in FFPE tumor tissues

We optimized a DIA-MS workflow with an 8 Da window to perform an in-depth analysis of FFPE tissues to capture and quantify GBM-relevant proteins. Patient-derived GBM tissue samples (n = 5) with 15 µm thick sections were processed in three biological replicates using Method 3 (ACN-urea-RapiGest-containing buffer). With our single-shot DIA-MS workflow, we were able to identify >1700 proteins across the samples. This was a huge improvement in proteome coverage compared to what we observed with the DDA analysis workflow (**Figure 3-4a**). We compared different functionalities of DIA-NN by analyzing the data in either library-free or spectral library mode (**Supplementary Figure 3-3a and b; Supplementary Figure 3-4a, b, c, d**).

We also compared DDA-MS and DIA-MS based on the sequence coverage. DIA-MS achieved a higher mean sequence coverage of 28.6% than 24.7% by DDA-MS (**Supplementary Figure 3-5a, b, c**). DIA-MS has been demonstrated in several studies^{111,112,199,200} to outperform DDA-MS in terms of sensitivity and proteome coverage. Therefore, this result corroborates such earlier findings. To interrogate the protein expression profile, we proceeded with the gene-centric matrix output of DIA-NN. Using protein intensities, we performed Pearson's cross-correlations of all LC-MS runs (n = 15) and technical replicate values were first averaged before correlating the samples. The global correlation matrix revealed strong agreement among replicates of the sample with a high Pearson's correlation coefficient of $R > 0.8$ (**Supplementary Figure 3-6**). Further, we explored the CV for all quantified proteins in each patient sample. All the samples recorded low median CV values of <20%, with an average of 16.2% and a median of 11.9% (**Figure 3-4b**). The low CV values indicate the high precision of our DIA-MS workflow for in-depth protein quantification. To explore the differences in the patient proteome profiles, we performed a principal component analysis (PCA) of quantified proteins (n = 1498) in all samples.

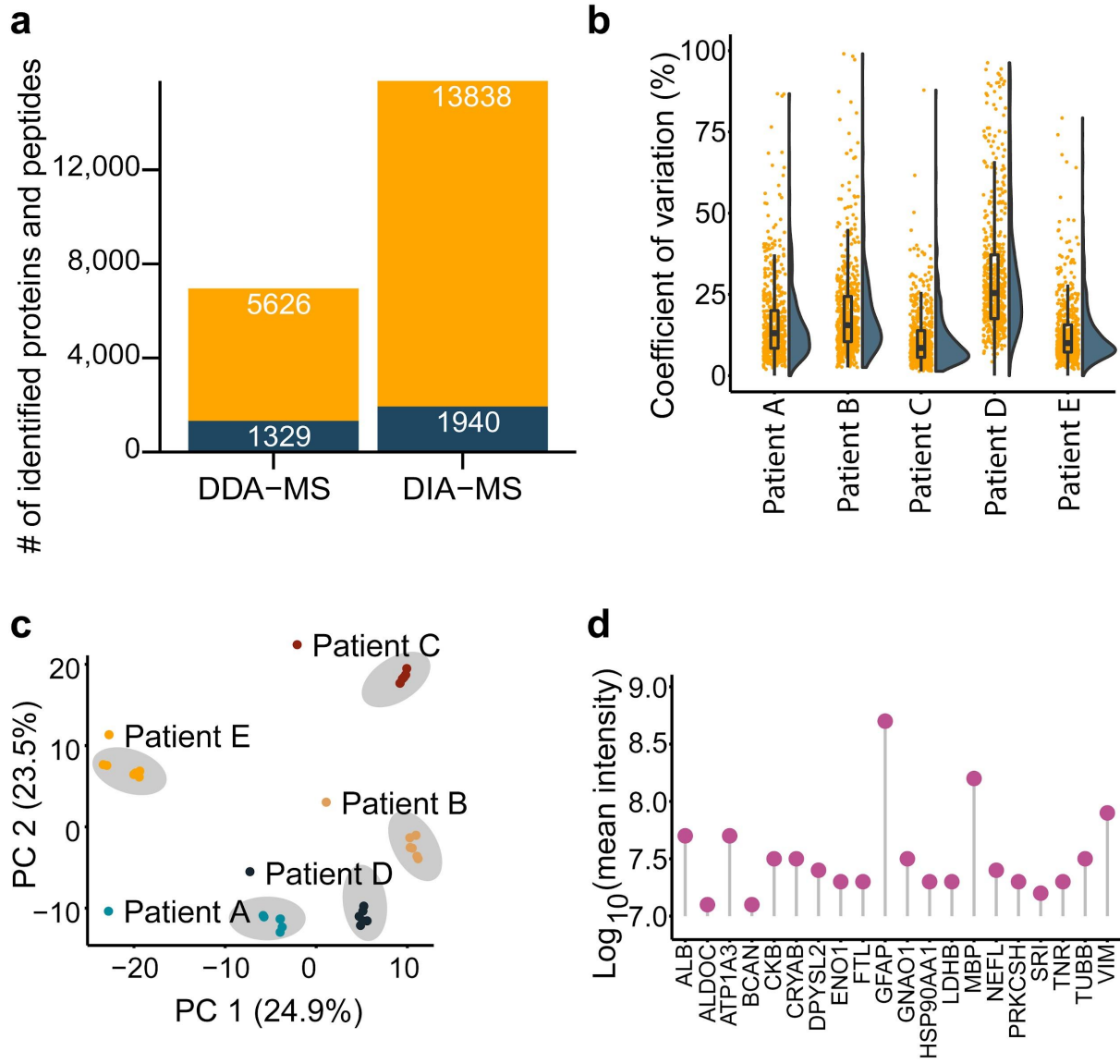


Figure 3-4. DIA-MS analysis of FFPE tissues derived from GBM patients ($n = 5$). **a**, Bar plots displaying the number of identified peptides and proteins with either DDA-MS or DIA-MS. **b**, Raincloud plot illustrates the CV of quantile-normalized protein intensities in all the sections for each patient e.g., A, B, C, D, and E. The midline of the boxplot in the raincloud plot indicates the median CV, the half violin illustrates the density of values. Each dot in the plot corresponds to the calculated CV for the individual protein in all sections per patient. **c**, Principal component analysis of proteins detected in all patient FFPE tissue samples. The color corresponds to different patients. **d**, Lollipop plot showing the top 20 GBM-relevant proteins. The ranking was based on quantile-normalized log (base 10) transformed mean intensity.

The PCA clearly classified the samples according to the patient where they were derived (**Figure 3-4c**). This finding demonstrates that the separation is driven by patient-specific proteome signatures. We further examined the expression values of proteins across samples. First, a mean intensity value was calculated and then ranked from highest to lowest. A screen of the top twenty proteins in terms of intensity values revealed the presence of GBM-relevant markers such as glial fibrillary acidic protein (GFAP)^{201–203} (**Figure 3-4d**). Similarly, VIM and MBP, which showed high-intensity values relative to other proteins, have been reported as valuable biomarkers in GBM^{204,205}. The confirmation of the presence of the proteins unique to glioma- and specifically GBM in the analyzed data demonstrates our DIA-MS workflow's high sensitivity and robustness in detecting and quantifying disease-relevant proteins (e.g., FN1 and BASP1). Next, we set out to explore the variation in quantifying top twenty proteins based on intensity values by calculating their CVs using mean expression intensities. The CV analysis recorded low values of 11.4% and 10.5% for mean and median, respectively (**Supplementary Figure 3-7**). This result further supports the robustness of our pipeline for protein quantification.

In order to investigate the observed proteome differences in the PCA (**Figure 3-4c**) further, we performed an unsupervised hierarchical clustering on the relative expression levels of 78 proteins with promising clinical utility in the context of GBM. As illustrated in **Figure 3-5a**, each patient displayed a distinct gradient on the heatmap, concordance with the PCA result. Next, we selected a few clusters of proteins focusing on those having high relative expression levels, and determined the variation in quantification. CV was calculated for replicates of the proteins for each cluster. We observed a lower CV at the cluster level (**Figure 3-5a**) compared to the global CV witnessed in **Supplementary Figure 3-7**. Again, all this data supports the robustness of the pipeline that we used in this analysis. Next, we leveraged STRING software to query the nature of the interaction of the 78 GBM-relevant proteins (**Figure 3-5b**). The network reported significant interactions with protein-protein interaction (PPI) enrichment p-value of $<1.0e-16$ and an average local clustering coefficient

of 0.527. This result demonstrates the ability of the DIA-MS workflow in this study to capture the PPI of interest that could be interrogated further to get some exciting biological insights concerning GBM tumor evolution and progression.

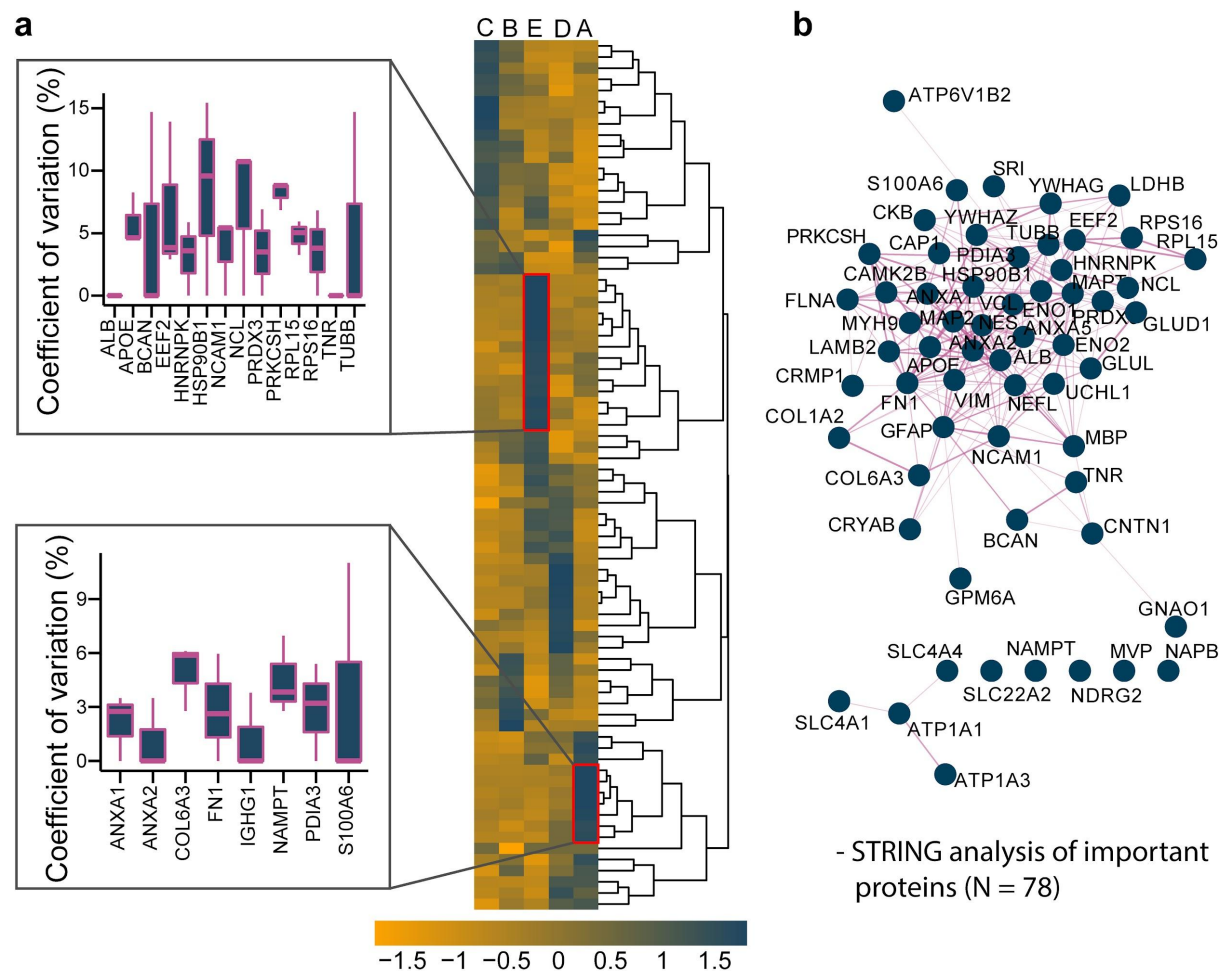


Figure 3-5. Heatmap and STRING diagram for a few selected proteins. **a**, Heatmap of relative expression levels of all proteins detected in all DIA-MS runs. Unsupervised hierarchical clustering is based on the scaled protein expression values. Clusters with high expression values are marked with cello color while those with low expression values are indicated with orange color. The boxplots show the CV distribution for proteins in the selected clusters. **b**, STRING diagram displaying the interaction of the 78 GBM-relevant proteins.

3.4.5. Screening of immune-related genes in a DIA-MS dataset

With the intensified interest in developing novel immunotherapeutics, a comprehensive understanding of how immune cells infiltrate GBM is necessary. We screened our DIA-MS dataset to identify the detected immune-related genes. Out of the 1498 quantified proteins, 348 were associated with the immune system. Next, we leveraged the recently published CPTAC (Clinical Proteomic Tumor Analysis Consortium) mRNA data²¹ to correlate with the protein expression levels in our dataset. When we intersected the CPTAC mRNA dataset with the cataloged 348 genes, we found 320 common immune-related genes. A low positive Pearson's correlation coefficient of $R = 0.37$ (**Figure 3-6a**) was reported when we compared the DIA readouts to CPTAC's mRNA expression levels (FPKM values). A low correlation between transcriptome and proteome has been reported in previous studies^{206,207}, and this is consistent in our study. The transcripts expression levels do not completely reflect the protein abundance since proteins undergo various post-transcriptional modifications. This phenomenon possibly contributes to the low positive correlation between transcriptome and proteome. A pairwise protein-level correlation between data in this study and CPTAC reported a positive Pearson's correlation coefficient value of $R = 0.42$ (**Supplementary Figure 3-8**). Some of the proteins e.g. MIF, which are currently being pursued as novel drug targets, showed elevated relative expression levels (**Figure 3-6b**).

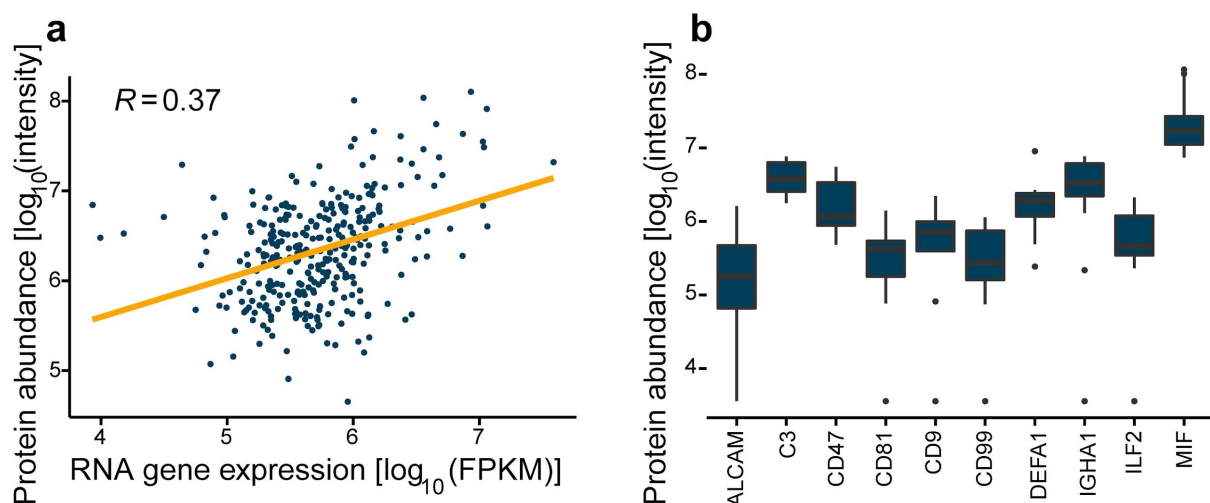


Figure 3-6. *Exploring expression levels of immune-related genes. a, Scatter plot showing the correlation between protein and mRNA expression levels. The yellow line represents the line of best fit with the shaded area showing the 95% confidence interval. R and p-value are based on Pearson's coefficient correlation. b, Boxplot showing the abundance of top immune-relevant proteins.*

3.5. Discussion

FFPE tissue samples have shown superiority over other sample types since they are stable over long storage; hence, providing more opportunities for researchers to discover clinically-relevant proteins¹⁸². Thus, efficient protein extraction from FFPE samples is crucial for targeted and discovery-based proteomics research. Presently, several methods applying DDA-MS for FFPE tissue sample proteomics analysis have been reported^{208–210}. DIA-MS is increasingly being applied in tissue proteomics. However, there are currently few studies on GBM where DIA-MS has been performed. Therefore, we aimed to establish a simple and highly precise DIA-MS workflow able to detect and quantify both GBM-relevant and immune-related proteins in the GBM tumor.

The analytical superiority of DIA-MS over DDA-MS has been demonstrated in several studies^{110,200}. Besides, it was recently shown that DIA-MS could yield results comparable to those generated using selected reaction monitoring (SRM)¹¹¹. Our study's observed increment in the number of detected proteins with DIA-MS relative to DDA-MS was consistent with earlier works²¹¹. The random sampling of the most abundant ion species in the DDA-MS has been cited as one of the contributing factors in lower protein identification in comparison to DIA-MS. Moreover, the stochastic sampling nature in DDA-MS often generates a high number of missing protein intensity values¹¹⁰. We achieved high precision in protein quantification, indicated by the low CV values.

The DIA-MS detected most of the proteins relevant to GBM pathology, including FN1, VIM, MBP, and GFAP^{18,212–215}. VIM and GFAP are co-expressed in human glioma²¹⁶, and

overexpression of VIM has been associated with poor survival outcomes for GBM patients²¹⁷. The co-expression of these two proteins is evident in our DIA-MS data, as depicted by their relatively elevated expression intensity values (**Figure 3-4d**). Taken together, this evidence is a demonstration of how sensitive and precise this DIA-MS workflow is for the quantification of proteins in FFPE tissue samples. The detection of these proteins in glioma samples has been performed mainly by using relatively less sensitive techniques. A recent study using a Q-Exactive mass spectrometer reported a significant elevation of VIM in GBM patient specimens²¹².

Overexpression of brain acid soluble protein 1 (BASP1) has been shown to promote the proliferation of cancer cells²¹⁸. In pancreatic cancer tissue, quantitative proteomics analysis using Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer identified BASP1 as a candidate prognostic biomarker²¹⁹. BASP1 plays a role in the organization of the cytoskeleton elements. FN1, a glycoprotein with cell adhesive features, has been implicated in tumor cell migration and invasion by interacting with integrin proteins²²⁰. Additionally, its overexpression in glial tumors has been associated with an increased migration rate of tumor cells²²¹. Some of the immune-related proteins (e.g., MIF, CD47, and TTYH1) that we detected and quantified with high precision are currently being interrogated as potential targets for immune-based therapy. MIF has been found differentially expressed in a subpopulation of GBM myeloid-derived suppressor cells and could be potentially targeted to repress immunosuppression in tumor microenvironment²²². Elevated expression of MIF with poor survival was recently reported in GBM samples in patients with O-6-Methylguanine-DNA Methyltransferase (MGMT) methylation²²³.

To conclude, the quantitative MS proteomics field is quickly gravitating towards simple and highly sensitive workflows that enable in-depth coverage and precise quantification of proteome-wide expression. In this regard, DIA-MS has emerged top and is increasingly being applied to quantify proteins in tissues. Therefore, accurate and precise quantification of proteins in patient-derived FFPE samples is expected to provide new leads to discover

novel therapeutic targets. In this study, a simple DIA-MS workflow was established and was successfully applied to patient-derived GBM tissue samples (n = 5). This DIA-MS approach enabled precise detection and quantification of proteins, and over 1400 proteins were quantified. Importantly, immune-related (e.g. MIF) and GBM-relevant proteins (e.g., GFAP) were detected and quantified. Finally, we anticipate that our approach would be useful for researchers interested in quantitative MS proteomic analysis of FFPE tissue specimens. Notably, it facilitates the detection and quantification of protein candidates for new therapeutic strategies for glioma and other malignant tumors.

CHAPTER 4: The interplay of hypoxia and antigen presentation in glioblastoma tumor: Insights from proteomics and immunopeptidomics

This work is from unpublished manuscript.

Weke, K., Dziubek, K., Papak, I., Muszczek, M., Hupp, T.R., Kote, S. The interplay of hypoxia and antigen presentation in glioblastoma tumor: Insights from proteomics and immunopeptidomics. *Unpublished manuscript*.

4.1. Abstract

Glioblastoma (GBM) is a highly aggressive and deadly brain tumor. The tumor microenvironment (TME) of GBM is characterized by hypoxia, which can suppress antitumor immune responses. Antigen presentation via the human leukocyte antigen class I (HLA-I) molecules is a critical process for T cells to recognize and eliminate tumor cells. However, the impact of hypoxia on antigen presentation in GBM is not fully understood. In this study, we used an integrated approach—combining mass spectrometry (MS)-based proteomics and immunopeptidomics strategies to investigate the interplay between hypoxia and antigen presentation. We found that severe hypoxic stress (0.1% O₂) induces significant changes in the GBM proteome. Key enzymes (ERAP1 and ERAP2) in the antigen processing and presentation machinery (APM) were downregulated in hypoxia. Moreover, the findings revealed that HLA-I-associated antigen peptide repertoire was reduced under hypoxia. Overall, these results open a newer horizon to explore the interplay between hypoxia and antigen presentation, and we propose that carboxypeptidase A4 (CPA4) has a potential role in shaping the tumor immunopeptidome landscape under a hypoxic microenvironment.

4.2. Introduction

Glioblastoma (GBM) is the most aggressive and common primary brain tumor in adults, leading to substantial morbidity and mortality. GBM accounts for 14.9% of all primary brain tumors and has a median survival of just 15 months post-diagnosis^{24,25,46}. The aggressive nature of GBM coupled with its resistance to conventional radiation and chemotherapy contribute to its dismal prognosis²²⁴. Therefore, the urgent need for developing novel therapeutic strategies against GBM has never been clearer. One emerging avenue of research focuses on leveraging the immune system to target GBM-specific antigens, offering a unique therapeutic approach^{225–228}. For such strategies to be efficacious, it is paramount to understand the tumor microenvironment (TME), as it heavily influences immune responses within the tumor. A significant player within the TME, with profound consequences on tumor biology, is hypoxia^{229–231}. GBM tumors, given their rapid proliferation rate, often exhibit regions of hypoxia, where the oxygen concentration falls below 2%, due to the tumor outpacing its vascular supply^{232,233}. This hypoxic environment stabilizes the transcription factor hypoxia-inducible factor 1 α (HIF1 α), which modulates numerous cellular pathways and impacts immune cell functions, thereby aiding tumor evasion from immune recognition^{234,235}.

In order for CD8+ T cells to recognize and mount an immune response against tumor cells, the tumor antigens must first be processed and then presented to them. Effective presentation of tumor antigens is through the antigen processing and presentation machinery (APM) in complex with the human leukocyte antigen class I (HLA-I) molecule^{236–238}. The impact of hypoxia on antigen presentation remains unresolved with some studies showing enhanced HLA-I expression, while others depict its downregulation. For instance, using murine tumor cells, Kajiwara *et al.*²³⁹ reported that hypoxia augments the expression of MHC (major histocompatibility complex) class I. Contrarily, Sethumadhavan *et al.*²⁴⁰ and Mohan *et al.*²⁴¹ demonstrated hypoxia-induced downregulation of HLA-I. Therefore,

the interplay between hypoxia and antigen presentation in tumors is a complex and multifaceted phenomenon, requiring an in-depth study to fully understand.

Mass spectrometry (MS)-based proteomics and immunopeptidomics are two powerful tools that can be used to investigate this interplay. MS-based proteomics allows for the identification and characterization of the proteins that are expressed by tumor cells under hypoxic conditions. A proteomics study to explore the effects of hypoxia on antigen presentation showed that the APM components did not change significantly under a hypoxic environment²⁴¹. Immunopeptidomics can be used to identify and quantify the antigenic peptides associated with HLA-I (pHLA-I). In a large-scale study, applying immunopetidomics strategies, GBM tumor antigens were identified²⁴². In this study, we aim to delve deeper into hypoxia's effect on antigen presentation. We integrated MS-based proteomics and immunopeptidomics approaches to investigate the interplay between hypoxia and antigen presentation in U-87 GBM cell lines.

4.3. Materials and methods

4.3.1. Cultivation of cells under hypoxia or normoxia

U-87 MG GBM cell lines were obtained from ATCC, cat. HTB-14. Cells were cultured in DMEM high glucose medium (Gibco, cat. 41966029) supplemented with 10% heat-inactivated Premium Plus FBS (Gibco, cat. A4766801) and 100UI/ml Penicillin-Streptomycin (Gibco, cat. 15140122). Cultures were maintained in the humidified atmosphere supplemented with 5% CO₂ at 37°C. Depending on the experimental conditions, cells were supplemented with varying levels of oxygen including 21%, 5%, 1%, or 0.1% of oxygen. Cultures requiring reduced oxygen levels (below 21%) were maintained in InVivo2 1000 Dual Chamber Hypoxia Workstation, Baker Ruskinn.

4.3.2. Protein isolation for Western Blot (WB)

Cells were washed two times with PBS and stored at -80°C until lysed directly on a cell culture plate with ice-cold CellLytic™ M (Sigma-Aldrich, cat. C2978) mixed with protease

inhibitors (Sigma-Aldrich, cat. P8340-1ML). Subsequently, lysates were gently mixed on ice for 20 minutes and centrifuged for 15 minutes at 14,000 x g at 4°C. Protein concentration in the supernatant was quantified by Bradford assay. Lysates were mixed with 4X reducing sample buffer and samples were denatured by boiling for 5 minutes at 95°C.

4.3.3. Immunoblotting

Protein lysates were resolved by 8% SDS-PAGE using 50µg of total protein lysate per well. Resolved proteins were then transferred to a nitrocellulose blotting membrane (Amersham Protran®) using Trans-Blot® Turbo™ Transfer System (Bio-Rad). Subsequently, membranes were blocked overnight in 5% non-fat dry milk in 0.1% Tween 20 Phosphate buffered saline (PBST) at 4°C. The next day, membranes were washed three times with 0.1% PBST and incubated overnight at 4°C with anti-HIF1A polyclonal antibody (cat. SAB1405933, Sigma-Aldrich) diluted in 2% blocking buffer. Restore Plus Western Blot Stripping Buffer (Thermo Scientific™, cat. 46430) was used for membrane re-probing with anti-β-actin antibody (Abcam, cat. ab6276). Following primary antibody incubation, membranes were washed three times in PBST and incubated for 1 hour at RT with HRP-conjugated rabbit anti-mouse secondary antibody, dilution 1:5000 (abcam, cat. ab6728). Finally, membranes were washed three more times with PBST, visualized by ECL substrate (Westar Antares, CYANAGEN), and ChemiDoc imaging system (Bio-Rad).

4.3.4. Flow cytometry analysis

Cells were dissociated from a cell culture plate with 0.25% Trypsin (Thermo Scientific™, cat. 25200072). Subsequently, 1,000,000 cells were aliquoted into flow cytometry tubes and centrifuged for 5 minutes at 1500rpm. Next, cell pellets were washed in PBS and centrifuged for 5 minutes at 1500 rpm. Subsequently, cells were stained for HLA I expression with 5µl of anti-HLA I antibody (clone W6/32, Thermo Scientific™ cat. 11-9983-42) for 30 minutes at RT, protected from light, and washed two times with PBS. Cell pellets were resuspended with 250µl of PBS and analyzed with BD FACSAria II cell sorter (BD Biosciences). Results were analyzed with FlowJo v10.8.1 flow cytometry analysis software (BD Biosciences).

4.3.5. Immunopeptidomics sample preparation

pHLA-I were isolated by antibody-free method. Cells were taken out of the incubators and immediately placed on ice. Culture medium was subsequently discarded and cells were washed three times with ice-cold PBS. Next, 2 mL of citrate-phosphate buffer at pH 3.3—(prepared using 0.131 M citric acid, 0.066 M Na₂HPO₄, and 150 mM NaCl, all from Sigma, PL)—was added to the cells and incubated for 3 min while continuously swirling the plate to ensure maximum contact of the buffer with cells and dislodgement of peptides bound to HLA-I molecules. Replicates were subsequently pooled and transferred to a clean 15 mL falcon tube and centrifuged at 1000 rpm for 5 min at 4°C to remove potential cell debris and media components. Next, peptides were purified using Oasis HLB VAC RC 30 mg cartridges (Waters, Germany). The cartridge was conditioned by adding of 1 mL of 0.2% formic acid/methanol, followed by an equilibration with 0.2% formic acid/water. Next, samples was loaded and washed three times with water/5% methanol/0.2% formic acid. Peptides were eluted with 1 mL water/80% methanol/0.2% formic acid and diluted to water/40% methanol/0.2% formic acid. Eluted peptides were subjected to an ultrafiltration step using an Ultracel-3kDa centrifugal filter unit (Merck, Poland). Initially, 1 mL of LC-MS grade water was added to the filters and centrifuged at 14000 xg for 30 min to rinse them before use. Filters were inserted in new collection tubes and eluted peptides were added, followed by centrifugation at 14000 xg for 30 min. Finally, the extracted peptides were evaporated using vacuum centrifugation and then stored at -20°C until MS analysis.

4.3.6. Proteomics sample preparation

After collecting pHLA-I, cells were washed three times with ice-cold PBS to remove any citrate-phosphate buffer left. Next, cells were gently scratched off the culture plate and transferred into a 1.5 mL Eppendorf tube, followed by a 5-minute centrifugation at 1000 rpm. The supernatant was aspirated out and the pellet was immediately snap-frozen in liquid nitrogen until the following day. Cell lysis was performed by adding 200 µL of 8 M urea in 0.1 M Tris/HCl—lysis buffer—to the cell pellet. The mixture was vortexed for a minute and

subsequently sonicated softly for 15 minutes. Samples were frozen overnight at -80°C to enhance lysis. Subsequently, samples were subjected to three successive cycles of snap-freezing in liquid nitrogen and thawing. Afterward, centrifugation was performed at $17,000\text{ xg}$ for 30 minutes at room temperature, and the clear supernatant was transferred to a clean LowBinding Eppendorf tube. Protein quantification was done using Pierce™ BCA Protein Assay Kits (Thermo Scientific™ cat. 23225). The next steps were carried out following the filter-aided sample preparation (FASP)⁸⁰ protocol with some modifications. Firstly, $200\ \mu\text{L}$ of lysis buffer was added to a Microcon-10kDa centrifugal filter (Merck Millipore cat. MRCPRT010), followed by $50\ \mu\text{g}$ of protein sample, which were mixed gently and then centrifuged for 30 minutes at $17,000\text{ xg}$, discarding the filtrate afterward. Following this, $200\ \mu\text{L}$ of lysis buffer was added to the relatively dry filter. Proteins were reduced to break disulfide bonds by adding $40\ \mu\text{L}$ of $50\ \text{mM}$ TCEP to samples and incubating at 37°C and $600\ \text{rpm}$ for 30 min. Subsequently, $100\ \mu\text{L}$ of $0.05\ \text{M}$ IAA in lysis buffer was added to each filter and was left in the dark with gentle shaking at room temperature for 20 minutes, and then centrifuged at $17,000\text{ xg}$ for 25 min, and the filtrate was discarded. Each filter was washed thrice with $200\ \mu\text{L}$ of $0.1\ \text{ABC}$ and then centrifuged for 30 minutes at $14,000\text{ xg}$ for each successive wash. The filter was transferred to a new holding Eppendorf. Next, $50\ \mu\text{L}$ of $50\ \text{mM}$ ABC was added to the filter followed by trypsin in a 1:100 ratio (trypsin to protein). The sample was incubated overnight in a chamber at 37°C for trypsin digestion. Peptide digest was collected by centrifuging the filter at $17,000\text{ xg}$ for 15 minutes. Subsequently, $50\ \mu\text{L}$ of $0.5\ \text{M}$ NaCl was added to the filter, followed by another centrifugation at $17,000\text{ xg}$ for 20 minutes. For peptide purification, Harvard Apparatus MicroSpin C-18 SPE columns were activated by the addition of $200\ \mu\text{L}$ of 0.1% FA in ACN and centrifuged at $300\ \text{xg}$ for 2 minutes, and was repeated twice. Columns were equilibrated by adding $200\ \mu\text{L}$ of 0.1% FA in water and centrifuging at $300\ \text{xg}$ for 2 minutes, which was again repeated. Another $200\ \mu\text{L}$ of 0.1% formic acid in water was added and allowed to stand for 15 minutes before a centrifugation at 300g for 2 minutes. The sample was loaded onto the column, followed by centrifugation at $300\ \text{xg}$ for 2 minutes. After washing the samples three times with $200\ \mu\text{L}$ of

0.1% FA in water, peptides were eluted stepwise using 200 μ L of 0.1% FA in 50% ACN and then 200 μ L of 0.1% FA in 80% ACN, and a final 200 μ L of 0.1% FA in 100% ACN, each successive step was followed by centrifugation at 300 \times g for 2 minutes. Lastly, the eluate was transferred to a 1.5 mL LowBinding microtube (Axygen) and evaporated to dryness in a vacuum centrifuge. Samples were then stored at -80°C until LC-MS measurement.

4.3.7. LC-MS/MS sample measurement

Ultimate 3000 RSLCnano (ThermoFisher Scientific, MA, USA) coupled to Orbitrap Exploris 480 mass spectrometer (ThermoFisher Scientific, MA, USA) were used for LC-MS/MS m. Initially, the dried peptides were reconstituted in 30 μ L of loading buffer (0.08% TFA in 2.5% ACN). Afterward, the peptide concentration was determined using NanoDrop 2000 (Thermo Scientific, MA, USA) with absorbance set to 220 nm.

Samples were injected into the LC system and analyzed in 3 technical replicates. First, peptides were trapped and concentrated on an Acclaim™ PepMap™ 100C18 trap column (ThermoFisher Scientific, MA, USA) of 300 μ m i.d x 5 mm and packed with 5 μ m, 100 Å particles. Subsequently, 75 μ m i.d x 15 cm Acclaim™ PepMap™ RSLC 100 reverse phase C18 analytical column packed with 2 μ m particles having 100 Å pore size was used to separate peptides at a flow rate of 300 nL/min. The analytical column was first equilibrated for 10 min at 2.5% solvent B (0.1% (v/v) TFA in acetonitrile) and 97.5% of solvent A (0.1% (v/v) TFA in water). Following, to separate the peptides, solvent B was increased in a 90-min gradient between 2.5% and 40%. A post-gradient cycle of 99% B followed a post-run equilibration at 2.5% B. The analytical column was coupled to the Nanospray Flex Ion Source. Recording of all MS measurements and spectra was done in resolution positive ion mode with a high voltage of 2500 V. For DDA mode, the ion transfer tube was heated to a temperature of 250 °C. MS1 data was acquired by setting the Orbitrap resolution to 120000, an MS scan range of 350–1200 m/z, an AGC (automatic gain control) level to custom mode, and a maximum injection time mode to auto. Precursor ions with charges between +2 and +6 and intensity values over 5.0e3 were selected for subsequent higher-energy C-trap

dissociation (HCD) fragmentation and MS2 scanning. Precursors were isolated with an m/z window of 2 and fragmented by HCD set at 30%. The product/fragment ions were channeled to the Orbitrap for the MS2 acquisition at a scan resolution of 120000 and a maximum injection time mode set to 60 ms. Repeated sampling was avoided by setting the dynamic exclusion mode to custom, exclusion after n times to 1, exclusion duration to 20, and mass tolerance to 10 parts per million (ppm). For data-independent acquisition (DIA) mode, the precursor mass range was 350-1100, isolation window (m/z) of 13, window overlap (m/z) of 1, number of scan events set 62, collision energy of 30%, and in a fixed mode, orbitrap resolution of 30000, and the scan range mode was set to auto. In addition, the AGC target was set to custom mode, the maximum injection time mode was set to 25 ms, and microscan was set to 1.

4.3.8. MS data processing and analysis

A study-specific spectral library was generated from DDA data using MSFragger¹⁹⁴ (v3.8) in the FragPipe (v20.0) proteomics pipeline. Both reviewed (20361 entries) and unreviewed (58691 entries) UniProt human proteome database (79220 total entries), downloaded on May 17th, 2022, was used. Most algorithms were used in the default setting with few adjustments. Precursor mass tolerance and fragment mass tolerance were set to 8 and 10 ppm, respectively. For proteomics data, the enzymatic protein digestion button was checked and set to trypsin, while for immunopeptidomics, the nonspecific protein digestion was switched on. The generated library consisted of 4107 protein isoforms, 4107 protein groups, and 44227 precursors in 34541 elution groups. DIA-MS RAW files were processed with Data-Independent Acquisition by Neural Networks (DIA-NN) software (v1.8)¹¹³. DIA-NN was operated in both library and library-free modes. The report output matrix was used for further analysis in an R statistical environment (v4.1.0). Protein and peptide intensity values were logarithmically transformed and normalized by the median for pHLA-I data and VSN method for proteomics data. Statistical tests were performed using the DEqMS R package.

4.4. Results and discussion

4.4.1. Experimental design and study workflow

Sample preparation and MS data acquisition were conducted as described in the preceding sections. Normoxia was defined as 21% oxygen, while 0.1%, 1%, and 5% were defined as different levels of hypoxia. Cells were grown in hypoxia or normoxia for 48 hours or 168 hours, and then both proteomics and immunopeptidomics analysis was performed (**Figure 4-1a, b, c**)

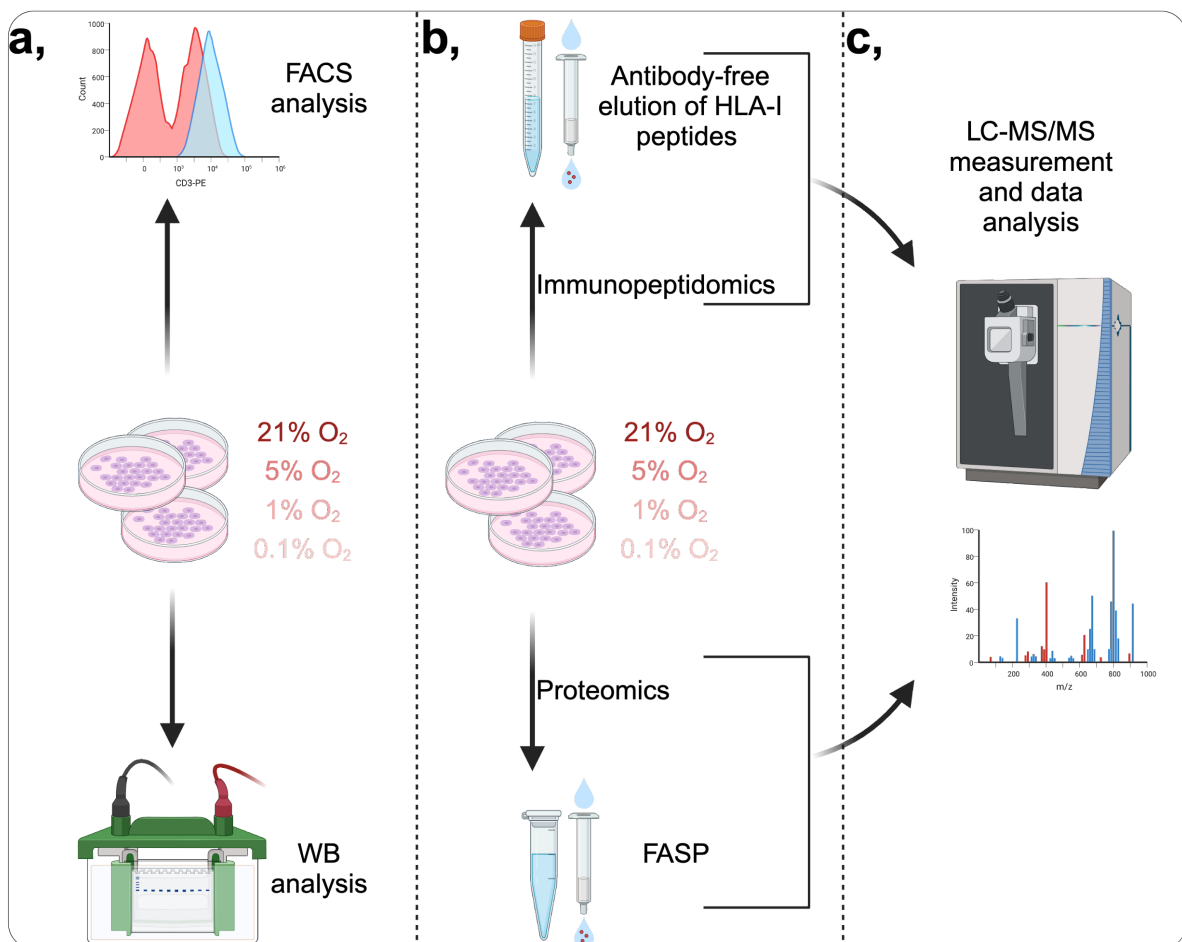


Figure 4-1. Schematic workflow for proteomics and immunopeptidomics analysis. Panel a, cultivation of U-87 MG cell lines under normoxia or hypoxia, Western blotting analysis, and fluorescence-activated cell sorting (FACS) analysis. Panel b, immunopeptidomics, and proteomics sample preparation. Panel c, LC-MS/MS measurement, and data analysis. Created with [BioRender.com](https://www.biorender.com).

4.4.2. HIF1 α is markedly expressed at 0.1% hypoxia

HIF1 α stabilizes under low oxygen concentrations and is a critical marker for detecting cellular response to hypoxia⁷¹. To determine the responsiveness of U-87 cell lines to varying hypoxic levels, we performed WB to check HIF1 α expression. We observed a marked increase in HIF1 α levels in 0.1% hypoxia, as shown by the relatively bigger and dense band, indicating the cells' adaptation to the hypoxic conditions (**Figure 4-2a**). Besides, relatively small increases were noticed with 1% and 5% hypoxia conditions. Furthermore, WB analysis of the cells' response to extended hypoxia (168 hours) showed that HIF1 α protein levels increased, similar to what was seen after 48 hours of exposure (**Figure 4-2b**). This finding confirms the established understanding of HIF1 α as a marker for cellular adaptation to hypoxia^{71,233,243,244}. Increased induction of HIF1 α in 0.1% hypoxia indicates a strong cellular response to severe hypoxic conditions. In a study that characterized the response of several patient-derived GB cell lines to hypoxia, both HIF-1 α and HIF-2 α proteins were found induced at 1% O₂ concentration²⁴⁵.

Next, we performed a flow cytometry analysis to determine the expression of surface HLA-I molecules. We observed that the cell lines expressed surface HLA-I molecules under both normoxic and hypoxic conditions. However, there was no remarkable difference in the expression levels between normoxia and any of the hypoxic levels (**Figure 4-2c, d, and e**). The 5% hypoxic condition reported a relatively higher count of median fluorescence intensity compared to either 0.1% or 1% hypoxia level, suggesting that the expression of HLA-I molecules may have been slightly reduced in a more hypoxic environment. Previous flow cytometry analysis demonstrated that hypoxia downregulates HLA-I molecules²⁴⁰. Notably, the confirmation of HIF1 α induction and expression of HLA-I molecules under hypoxia signaled that these cells were fit for subsequent experimental procedures to decipher the interplay between hypoxia and antigen presentation.

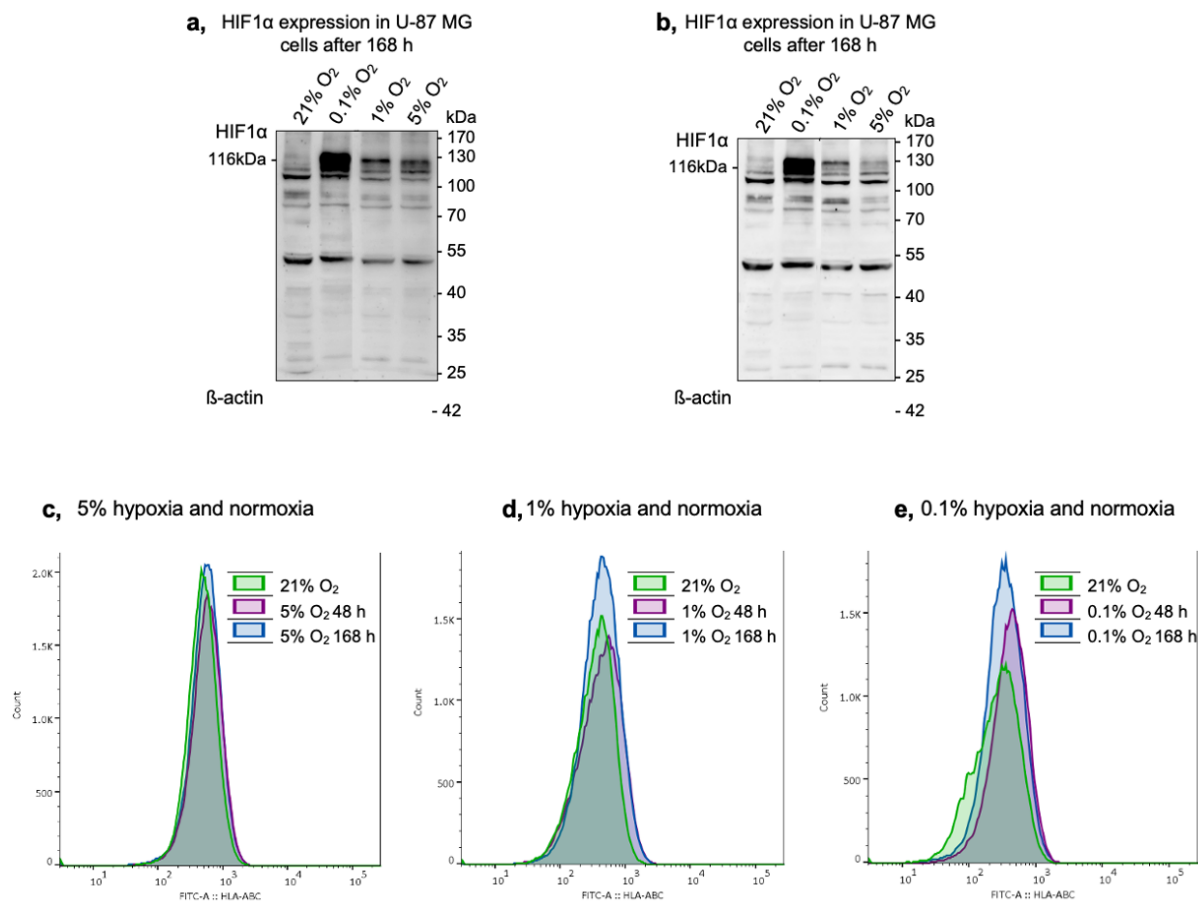


Figure 4-2. Analysis of U-87 MG GBM cells' responsiveness to hypoxia and determination of surface expression of HLA-I molecules. **a** and **b**, WB analysis of HIF1 α protein expression under normoxia or hypoxia for 48 hours and 168 hours, respectively. **c-e**, FACS analysis of surface expression of HLA-I molecules under normoxia or hypoxia (5%, 1%, and 0.1%, respectively) for 48 hours and 168 hours.

4.4.3. Hypoxia-induced proteome changes in U-87 GBM cell lines

Proteomics data were processed using DIA-NN, as mentioned earlier. The resulting protein-level data matrix was analyzed in an R statistical environment. Over 3000 and 25000 proteins and peptides were respectively confidently identified and quantified, with a high (>70%) fraction of samples having complete data (**Supplementary Figure 4-1a, b**). This number of IDs shows a relatively deep coverage of the proteome. Next, we did control checks to assess the quantitative quality of the data. The data was of good quality, and as illustrated in **Supplementary Figure 4-1c**, there were no instances of sample loading

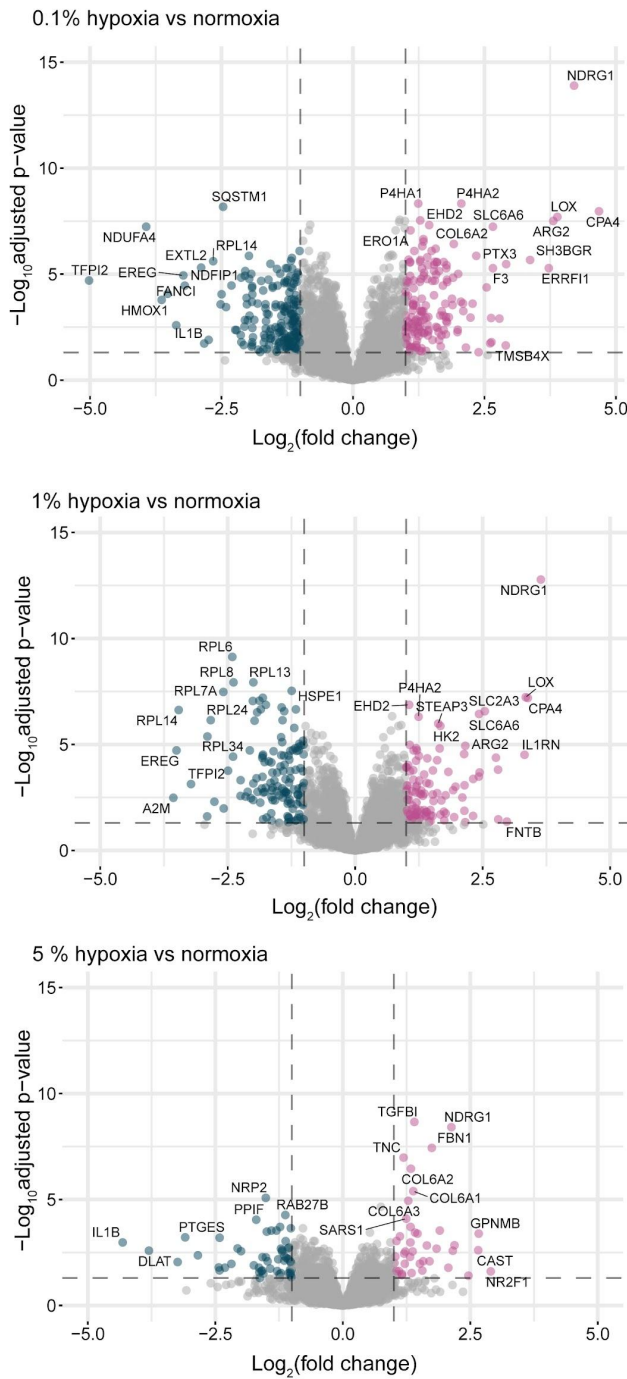
problems. Quantitative values were logarithmically transformed and normalized using the variance stabilizing normalization method (VSN).

To detect differential abundance between groups, differential expression analysis of quantitative mass spectrometry data (DEqMS)²⁴⁶ was used to perform statistical testing. All three hypoxic levels were compared to normoxia, and the resulting data was visualized with volcano plots (**Figure 4-3a**). A total of 361 proteins were differentially expressed between 0.1% hypoxia and normoxia (top panel). For 1% hypoxia versus normoxia, 250 proteins had significant differential abundance (middle panel). A comparison between 5% hypoxia and normoxia had 101 proteins with significant differential abundance (bottom panel). All differentially expressed proteins from the three comparison groups were overlapped, as illustrated in **Figure 4-3b**. About 23 proteins were common in all the comparison groups, and an overlap between 0.1% hypoxia vs. normoxia and 1% hypoxia vs. normoxia reported the highest number (133) of shared proteins with differential abundance. Next, we assessed the magnitude of differential expression by plotting a heatmap of log₂ fold changes for the top most differentially expressed proteins common in all three comparison groups.

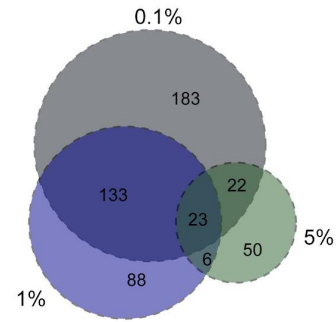
The 0.1% hypoxia vs. normoxia had the highest fold changes, followed by 1% hypoxia vs. normoxia and then 5% hypoxia vs. normoxia (**Figure 4-3c**). These results show that different hypoxic levels have varied changes in the protein expression profile of U-87 GBM cells. Moreover, the highest number of differentially expressed proteins (361) was observed under the most severe hypoxic condition (0.1% O₂). This suggests that these cells under extreme hypoxic conditions undergo significant molecular changes, perhaps as a response to cope with such low oxygen availability. Furthermore, it seems that 5% O₂ is the least severe hypoxia hypoxic environment, possibly suggesting that U-87 cells might find 5% hypoxia relatively tolerable and less of a stress factor compared to 0.1% or 1% hypoxic conditions. The observed higher fold change at 0.1% hypoxia indicates more dramatic changes to the proteome in response to severe hypoxia. The finding that 0.1% O₂ resulted in the highest number of differentially expressed proteins corroborates the WB analysis and indicates a

heightened cellular response under extreme hypoxic stress. A label-free shotgun proteomics analysis identified 2928 proteins and found 1034 proteins unique to the hypoxia condition. Further, 17 proteins were differentially expressed in U-87 GBM cells cultured under hypoxic conditions (1% O₂) compared to normoxic conditions (21% O₂)²⁴⁷. The differentially expressed proteins were involved in various cellular processes, including metabolism, cell cycle regulation, and hypoxia response.

a, Comparison of protein abundances



b, Overlap of differentially expressed proteins (DEP)



c, Most DEP common in all groups

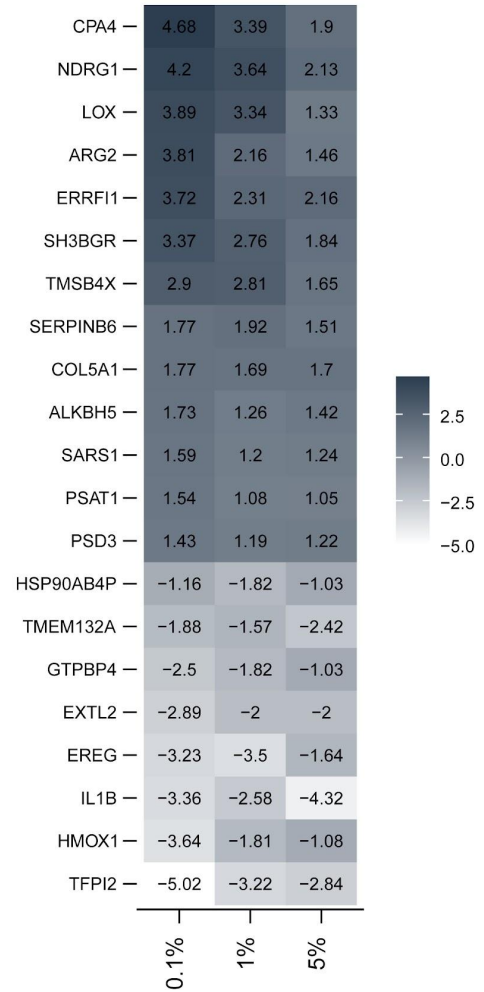


Figure 4-3. Hypoxia-dependent differential regulation of U-87 MG cells' proteome. *a*, Volcano plots of differentially regulated proteins. Top, middle, and bottom panels represent 0.1% hypoxia vs. normoxia, 1% hypoxia vs. normoxia, and 5% hypoxia vs. normoxia, respectively. The X-axis represents the log₂ fold change in protein expression, whereas the Y-axis indicates -log₁₀ of adjusted p-values. The muted

raspberry color highlights significantly upregulated proteins under hypoxia, while dark teal displays downregulated ones. Proteins with relatively much lower adjusted p-values or higher fold change are labeled. b, Venn Diagram showing the overlap of differentially regulated proteins among comparison groups. Each circle in the diagram represents a set of proteins, either upregulated or downregulated. The intersecting areas indicate sets of proteins with shared regulation across the experimental groups. c, A heatmap displaying the most differentially regulated proteins that are shared by all comparison groups. Rows represent the individual proteins, and the columns represent the different comparison groups. The color intensity indicates the fold change of expression of each protein under hypoxic (0.1%, 1% and 5% oxygen) compared to normoxia, with a gradient of color (from white to dark grey) indicating a range from low to high expression levels.

4.4.4. Metabolic reprogramming is a key adaptation to extreme hypoxic stress

To understand the top regulated pathways, we subsetted upregulated proteins at 0.1% hypoxia and performed gene set enrichment analysis using the KEGG pathway database. Metabolic-related pathways—including glycolysis/gluconeogenesis, carbon metabolism, biosynthesis of amino acids, protein digestion, and absorption—and the HIF-1 signaling pathway were the most enriched (**Figure 4-4a**). To understand whether the set of differentially regulated proteins common to all three hypoxic levels is involved in the fundamental cellular process, we used gene ontology (GO) enrichment analysis and discovered the top ten enriched processes (**Figure 4-4b**). Some of these biological processes include regulation of cytokine production, vasculature development, and tube development. These findings are consistent with previous studies showing hypoxia-induced metabolic switch from oxidative phosphorylation to glycolysis. A TMT 6-plex proteome analysis of hypoxic glioblastoma cells by Kora *et al.*²⁴⁸ identified a number of proteins that were differentially expressed under hypoxic conditions, including the glycolytic pathway.

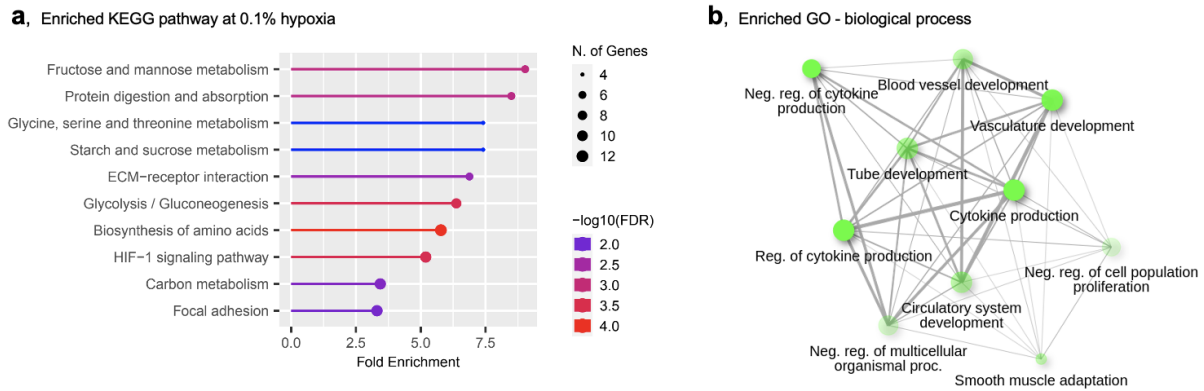


Figure 4-4. Enrichment analysis of differentially regulated proteins. *a*, Lollipop plot of top enriched KEGG pathways at 0.1% hypoxia. The horizontal coordinate represents the magnitude enrichment. Adjusted *p*-values are color-coded from blue (low) to red (red). The size of the lollipop circle corresponds to the number of genes for the respective pathway. Network of enriched GO (biological process) for the top significantly regulated proteins common in three different hypoxic levels. Dark green nodes represent more significantly enriched gene sets, and bigger nodes represent larger ones. Thicker edges represent more overlapped genes.

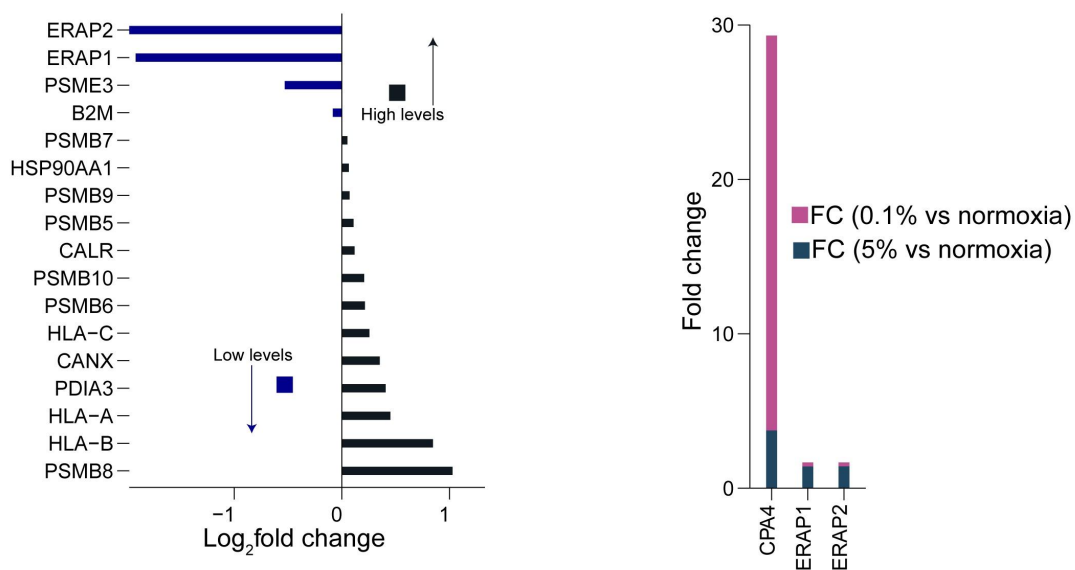
4.4.5. Hypoxia-dependent dysregulation of APM components

The antigen processing machinery (APM) comprises different components, including enzymes (ERAP), TAP, HLA, etc., which are necessary for effective antigen processing and presentation by the HLA-I molecules^{249,250}. Dysregulation of any of these components can significantly alter HLA-I antigen repertoire, thereby limiting recognition of tumor cells by cytotoxic CD8+ T cells^{240,251}. Using our proteomics data, we investigated whether 0.1% hypoxia altered the expression pattern of APM components. As illustrated in the barplot (**Figure 4-5a**) showing the log₂ fold change (<-1 or > 1), most of the APM components were not significantly different except for the enzymes (ERAP 1 and 2), which were downregulated. This finding is similar to Mohan *et al.*²⁴¹; however, for their study, they had a contrary observation with the enzymes, which were not significant. ERAPs are involved in trimming of peptides for HLA-I presentation; therefore, dysregulation of their expression profiles potentially impacts the kind and number of peptides presented by HLA-I. We noticed

a surprising observation with carboxypeptidase A4 (CPA4), which was increased many folds (**Figure 4-5b**). CPA4 is a member of the A family of carboxypeptidases, which are exopeptidases that cleave C-terminal amino acids from peptides and proteins²⁵². The oncogenic effects of CPA4 on tumor progression have been reported with some findings showing its potential utility as a poor prognostic biomarker^{253–255}. Moreover, Moon *et al.*²⁵⁶ demonstrated a HIF-1 α -dependent elevation of CPA4 in hypoxic human adipose-derived stem cells, which is consistent with our finding.

While there is no previous study on the involvement of CPA4 in trimming peptides for HLA-I presentation, we speculated that its marked elevation under hypoxia might, to some degree, lead to the cleaving of C-terminal amino acids from proteins prior to proteasomal degradation. To further investigate this hypothesis, we studied the peptide sampling from source proteins and assessed the sampling pattern at the C-terminus and N-terminus of proteins that were dysregulated the most. To achieve this, we used our immunopeptidomics data and mapped peptides to their respective source proteins using COBALT alignment tool²⁵⁷. We observed a distinct distribution of peptides in the C- and N-terminus for hypoxia compared to normoxia. For instance, under hypoxia, filamin-A protein (FLNA) had fewer peptides at the C-terminus with zero peptide sampling at position ~2000, whereas under normoxia, peptides are sampled from this position (**Figure 4-5c**). This finding suggests that marked elevation of CPA4 under hypoxia potentially shortens the proteins from the C-terminus before or during proteasomal degradation. Therefore, we are proposing that CPA4 has a potential role in shaping the HLA-I-bound peptide repertoire under a hypoxic environment.

a, Antigen processing and presentation machinery **b**, Comparison of ERAP and CPA4 genes



c, Peptide sampling in filamin-A protein

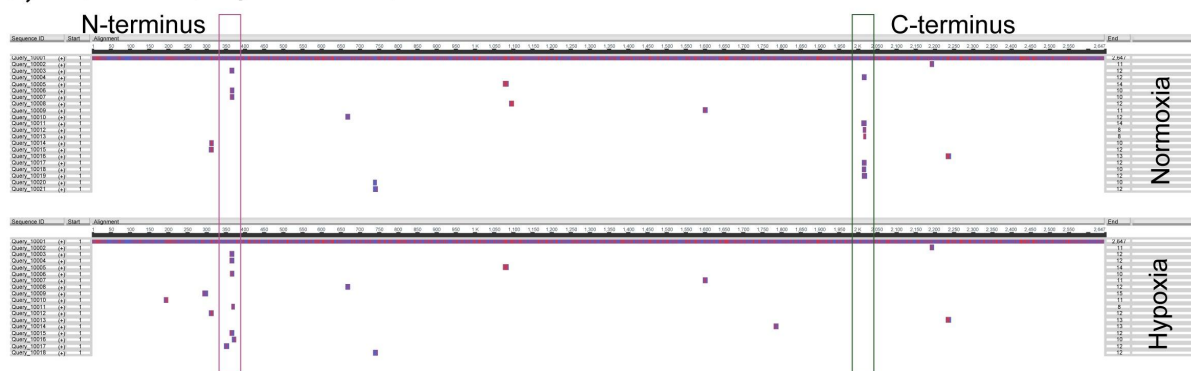


Figure 4-5. Hypoxia-dependent dysregulation of APM components. **a**, A barplot displaying log₂ fold change between 0.1% hypoxia vs. normoxia for APM proteins. Blue and black bars represent downregulated and upregulated proteins, respectively. **b**, A barplot for a fold change comparison of CPA4 and ERAPs enzymes. The muted raspberry color indicates a fold change for the 0.1% hypoxia vs. normoxia, while the dark teal color shows a fold change for the 5% hypoxia vs. normoxia. **c**, COBALT alignment visualization for filamin-A protein. The top panel displays peptide alignment from normoxia data whereas the bottom panel is for 0.1% hypoxia. Aligned amino acid sequences are shown in different colors, from red to blue, and appear like dots on the alignment map.

4.4.6. Hypoxia alters HLA-I peptide repertoire

To further explore the impact of hypoxia on HLA-I peptide repertoire, peptide characteristics including identification counts, and length distribution were evaluated. Initially, we explored proteins that presented the most by HLA-I molecules. The 0.1% hypoxia demonstrated the highest peptide sampling from the source proteins. Top sampled proteins include G3P, ENOA, VIME, and ALDOA, among others (**Figure 4-6a**). Next, we compared peptide counts from the three different hypoxic levels to normoxia. The normoxic group had the highest number of peptides followed by 5% hypoxia, then 1% hypoxia, and 0.1% hypoxia had the least peptide counts (**Figure 4-6b**). Peptides bound to HLA-I are between 8 to 15 amino acids long with a predominant 9 mer. We compared the peptide length distribution patterns for normoxia, 0.1%, 1%, and 5% hypoxia for U-87 cells cultured for 48 hours or 168 hours. As was expected, 9 mer had the highest count—confirming the classical profile of HLA-I peptides. We noticed that the 48-hour experiment had more count compared to 168 hours (**Figure 4-6c**).

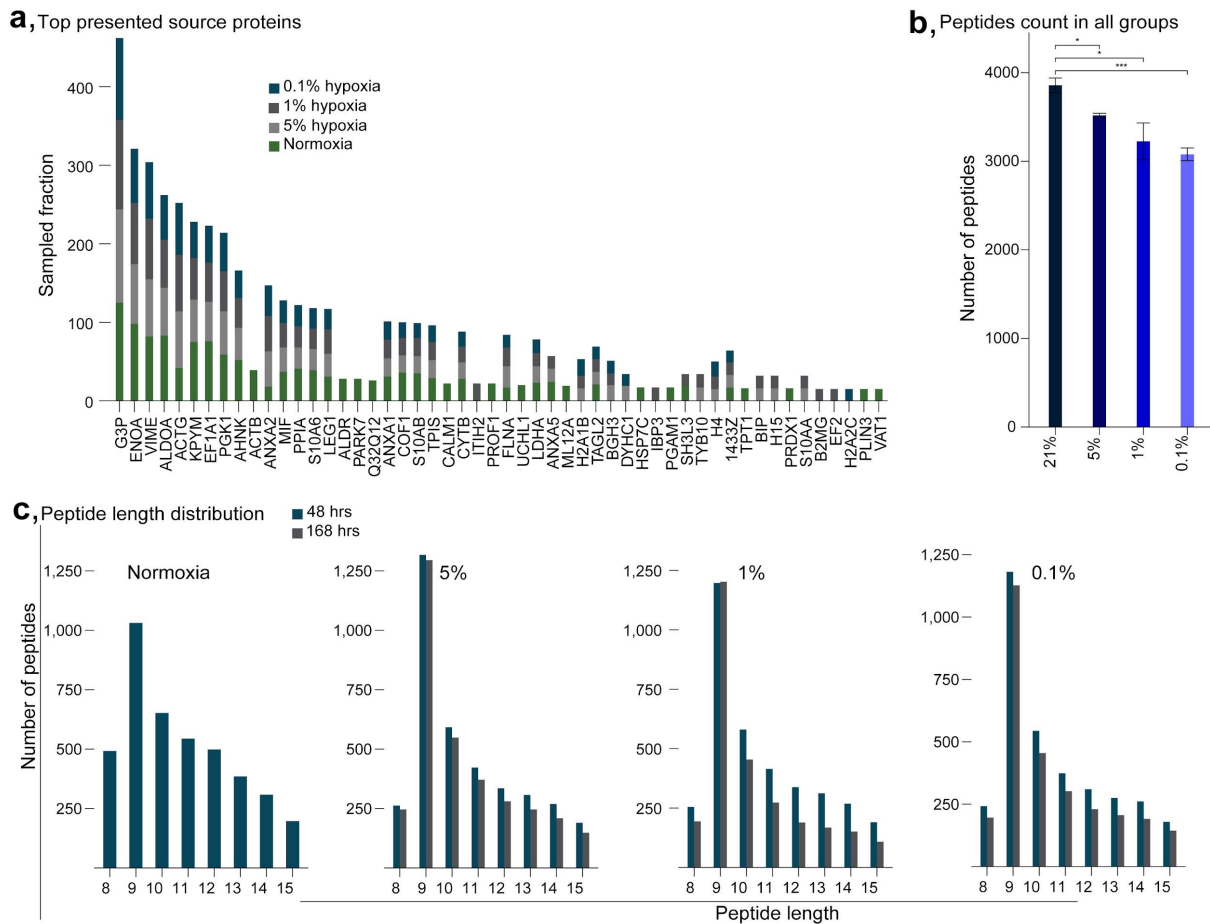


Figure 4-6. Hypoxia alters HLA-I peptide repertoire. *a*, A barplot of most presented source proteins for different hypoxic levels and normoxia. The horizontal axis represents the individual proteins and the vertical axis displays a cumulative number of peptides for the corresponding protein. *b*, A barplot showing the number of identified peptides from each experimental group. *c*, Barplots displaying peptide length distribution for all experimental groups and different cell culturing durations.

4.5. Conclusion

In conclusion, this study provides insights into the effects of hypoxia on antigen processing and presentation. The proteomic analysis revealed significant alterations in the U-87 GBM cell proteome under extreme hypoxic stress (0.1% O₂), with 361 differentially expressed proteins compared to normoxia. Metabolic reprogramming emerged as a central adaptation mechanism, with pathways related to glycolysis/gluconeogenesis, carbon metabolism, amino

acid biosynthesis, and HIF-1 signaling being highly enriched. Hypoxic stress resulted in the dysregulation of components of the APM, particularly the downregulation of ERAP enzymes. Marked elevation of CPA4 was observed under hypoxia. This finding suggests a potential role of CPA4 in shaping the HLA-I-bound peptide repertoire under hypoxia by cleaving C-terminal amino acids from proteins before or during proteasomal degradation. Overall, our results suggest that hypoxia has a complex and multifaceted impact on antigen presentation in GBM. Future studies are needed to further elucidate the underlying mechanisms and to identify potential therapeutic targets to overcome hypoxia-mediated immune suppression in GBM.

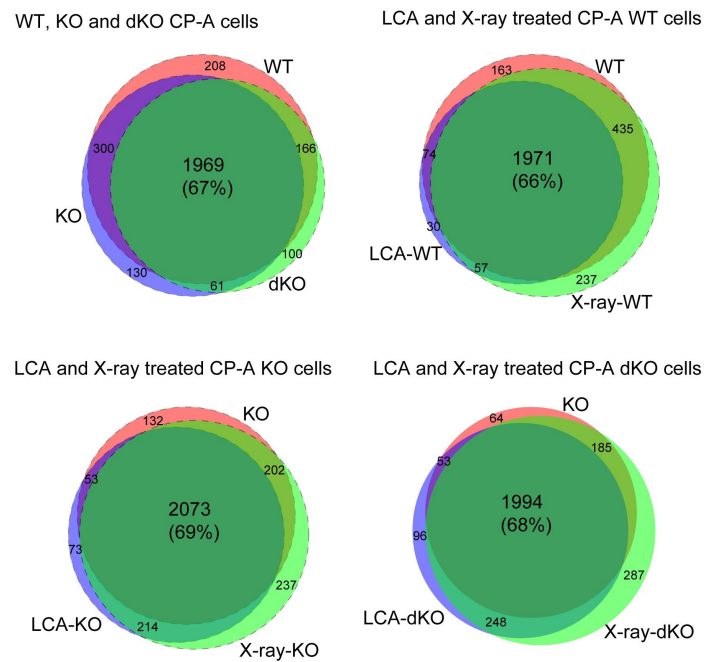
CHAPTER 5: Conclusions and future directions

In conclusion, the three studies presented in this thesis have made significant contributions to the field of proteomics and immuno-oncology. The first study developed a novel microscale proteomics method, microPOTS, which can identify and quantify differentially expressed proteins from limited numbers of cells. This method has the potential to be used for early detection of cancer and other diseases, as well as for monitoring the response to therapy. The second study established a simple and precise DIA-MS workflow for proteome analysis of FFPE tissues. This easy-to-implement workflow can be used to detect and quantify immune- and GBM-relevant proteins from FFPE patient materials.

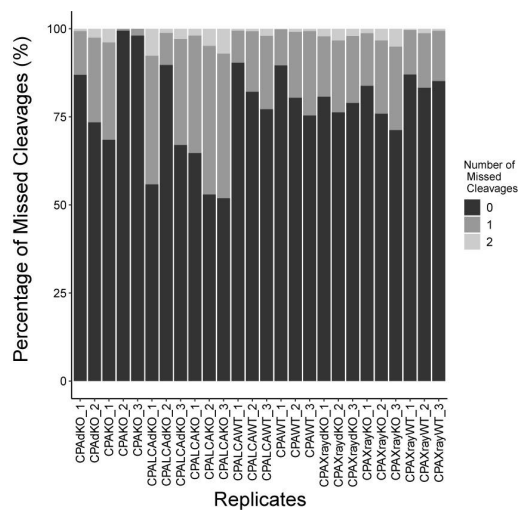
The third study investigated the interplay between hypoxia and antigen presentation in GBM using an integrated approach—combining MS-based proteomics and immunopeptidomics strategies. The study found that hypoxia induces significant changes in the GBM proteome, downregulates critical enzymes in the APM, and reduces the HLA-I-associated antigen peptide repertoire. These findings provide new insights into the mechanisms of immune escape in GBM, and suggest that targeting the APM components to enhance their expression under hypoxia may be a promising therapeutic strategy. Future directions for this research include further development and optimization of these methodologies for applications in clinical settings. Further investigation of the interplay between hypoxia and antigen presentation in GBM, including confirming and improving these new findings and exploring the possibilities of reversing hypoxia-induced downregulation of APM components as a new immunotherapeutic strategy.

Appendices

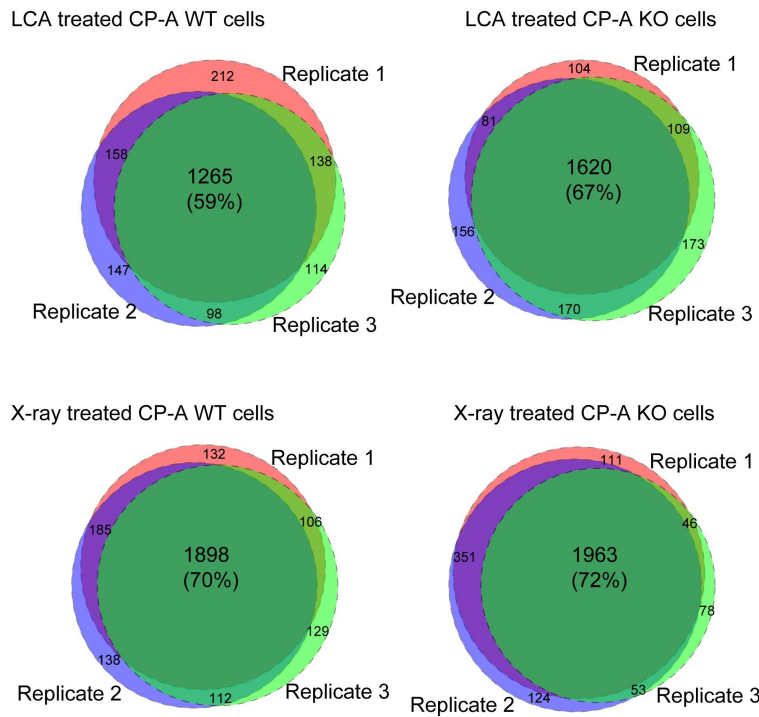
Appendix I: Supplementary material for chapter 2



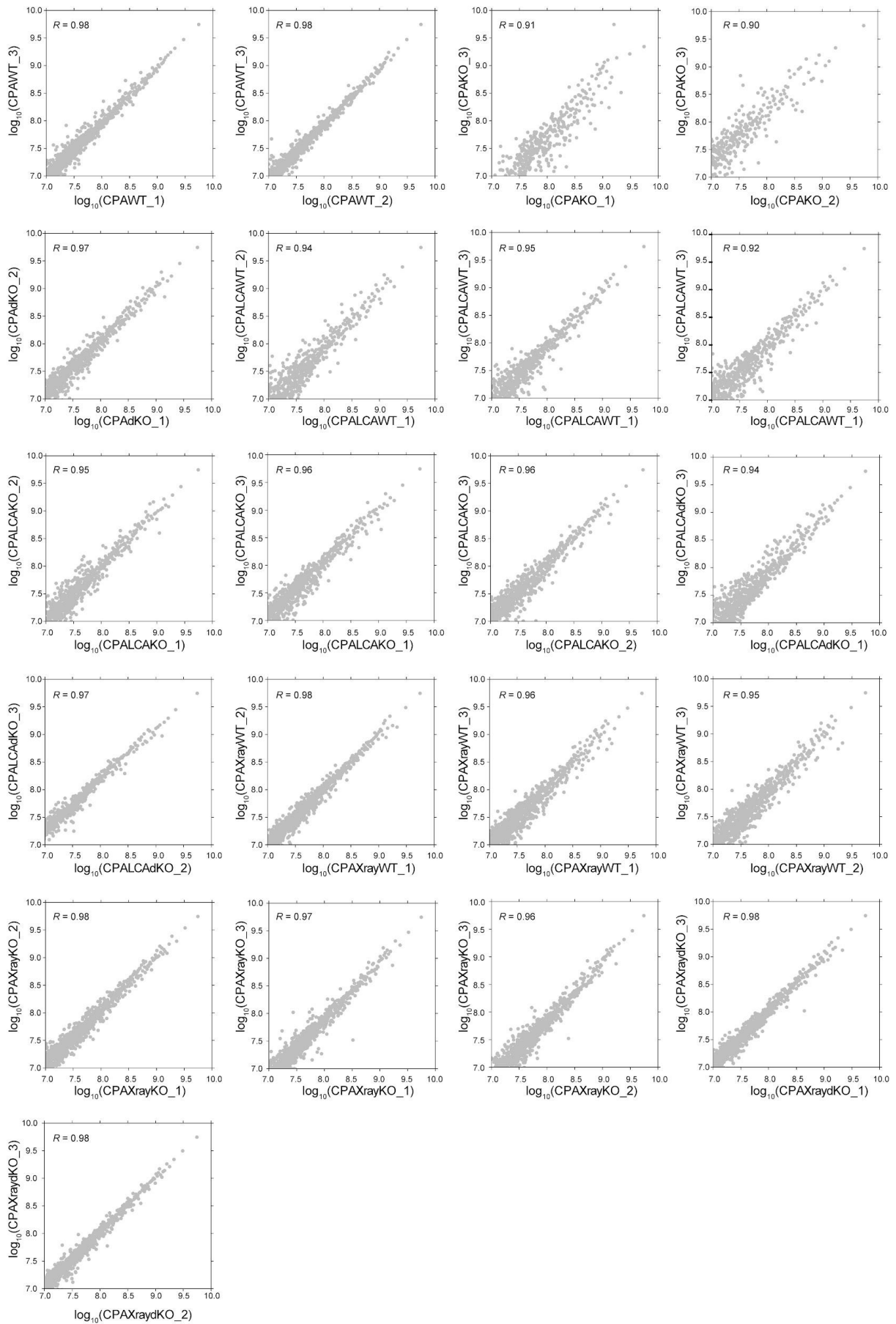
Supplementary Figure 2-1. Venn diagram illustrating the shared number of identified proteins among different sample types, as well as proteins that are unique to each sample type.



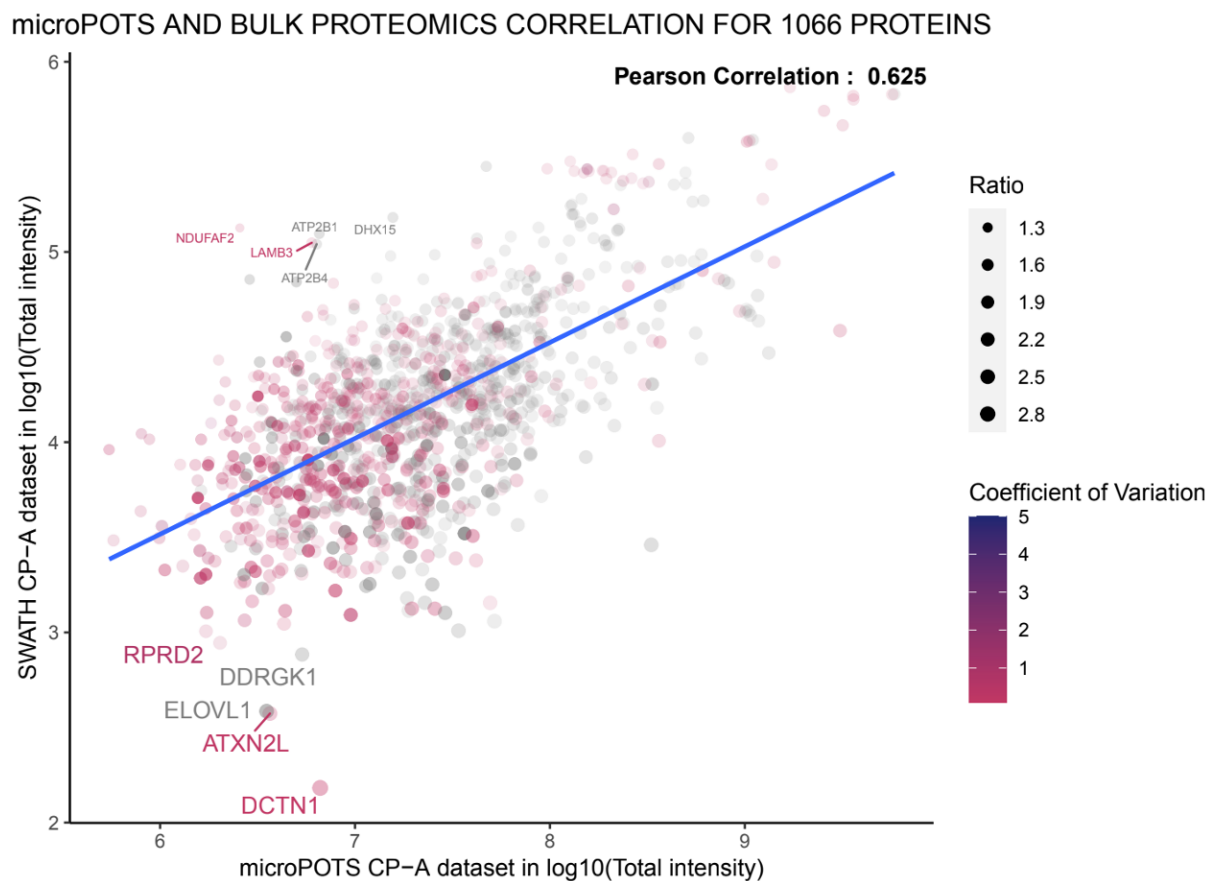
Supplementary Figure 2-2. The number of missed cleavages is shown for all replicates and is expressed in percentage.



Supplementary Figure 2-3. Qualitative assessment of the reproducibility of the microPOTS system. Area proportional Venn diagrams illustrate the overlap of the number of identified proteins among the replicates for each sample type.



Supplementary Figure 2-4. Quantitative assessment of the reproducibility of the microPOTS system. The scatter plots illustrate the pairwise comparison of the logarithmically (base 10) transformed LFQ.



Supplementary Figure 2-5. Scatter plot with associated Pearson's correlation coefficient ($R = 0.625$) between microPOTS and bulk proteomics for all overlapping 1066 proteins that we identified. Geometric mean normalization was used to normalize the protein intensities and a coefficient of variation (CV) was computed for each dataset. The vertical and horizontal coordinates respectively represent the log10 transformed protein intensities for bulk and microPOTS analyses of CP-A cells. Each dot shows an individual protein and the size of the dots indicates the ratio of protein intensities for microPOTS versus bulk proteomics analyses. The color ranges from violet (low CV in microPOTS) to dark blue (high CV in microPOTS). Color transparency is controlled by the CV values for bulk analysis, with more

intense color illustrating high CV values while low CV values are more transparent. Proteins with potential outlier intensity values are labeled on the plot.

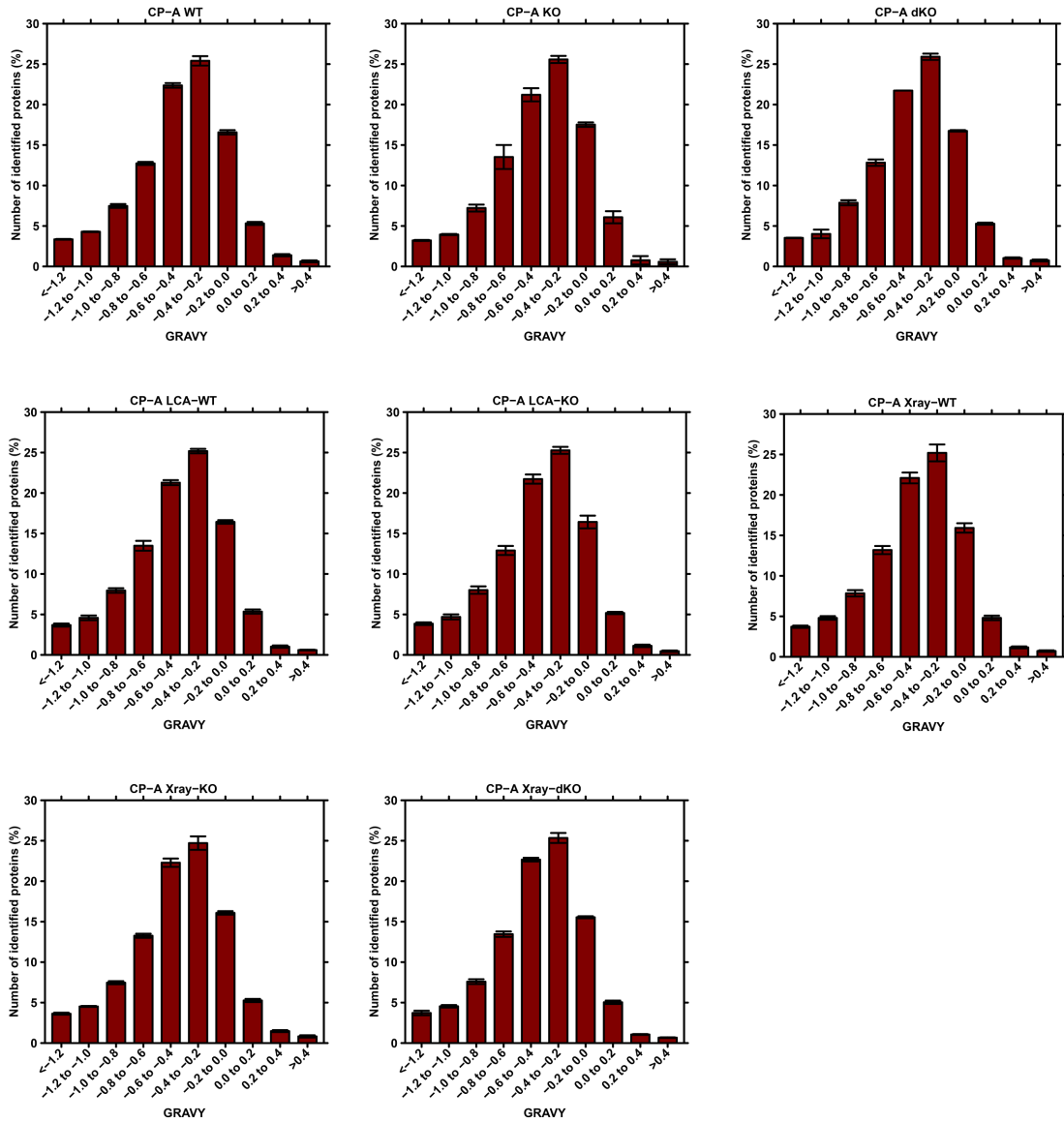
Notes 2-1. Sample preparation and LC-MS analysis for bulk proteomics dataset.

Samples were prepared according to filter-aided sample preparation protocol (FASP)⁸⁰. Evaporated digests were dissolved in 50 µl of loading buffer composed of 5% acetonitrile (ACN), and 0.05% trifluoroacetic acid (TFA) in liquid chromatography-mass spectrometry (LC-MS) grade water. Approximately, 1 µg of peptide sample was loaded into LC fluidics. Peptide trapping, peptide separation, and data-dependent mass spectrometry were performed as described by Way *et al.*²⁵⁸ and Faktor *et al.*²⁵⁹. Briefly, peptide concentration and separation were performed on Eksigent Ekspert nanoLC 400 system (AB-SCIEX) nano-LC system online coupled to TripleTOF 5600+ (AB-SCIEX) mass spectrometer. Peptide samples were concentrated and desalted on a cartridge trap column (300 µm i.d. x5 mm) packed with C18 PepMap100 sorbent with 5 µm particle size (Thermo Fisher Scientific). Analytical separation was carried out on a 25 cm fused-silica emitter column with 75 µm inner diameter (New Objective), packed in-house with ProntoSIL C18 AQ 3 µm beads (Bischoff Analysentechnik GmbH). LC mobile phases were composed of 0.1% formic acid (FA) in LC-MS grade water (mobile phase A) and 0.1% FA in LCMS grade ACN (mobile phase B). A gradient starting at 5% B linearly increasing up to 40% B in 120 min with a flow rate of 300 nL/min was employed to separate peptides prior to their introduction into mass spectrometer (MS). MS was operated in data-dependent mode, and the top 20 intense precursor ions were fragmented with an exclusion time set to 12 s. The minimum precursor ion intensity was set to 50 cps and 100 ms accumulation time per precursor.

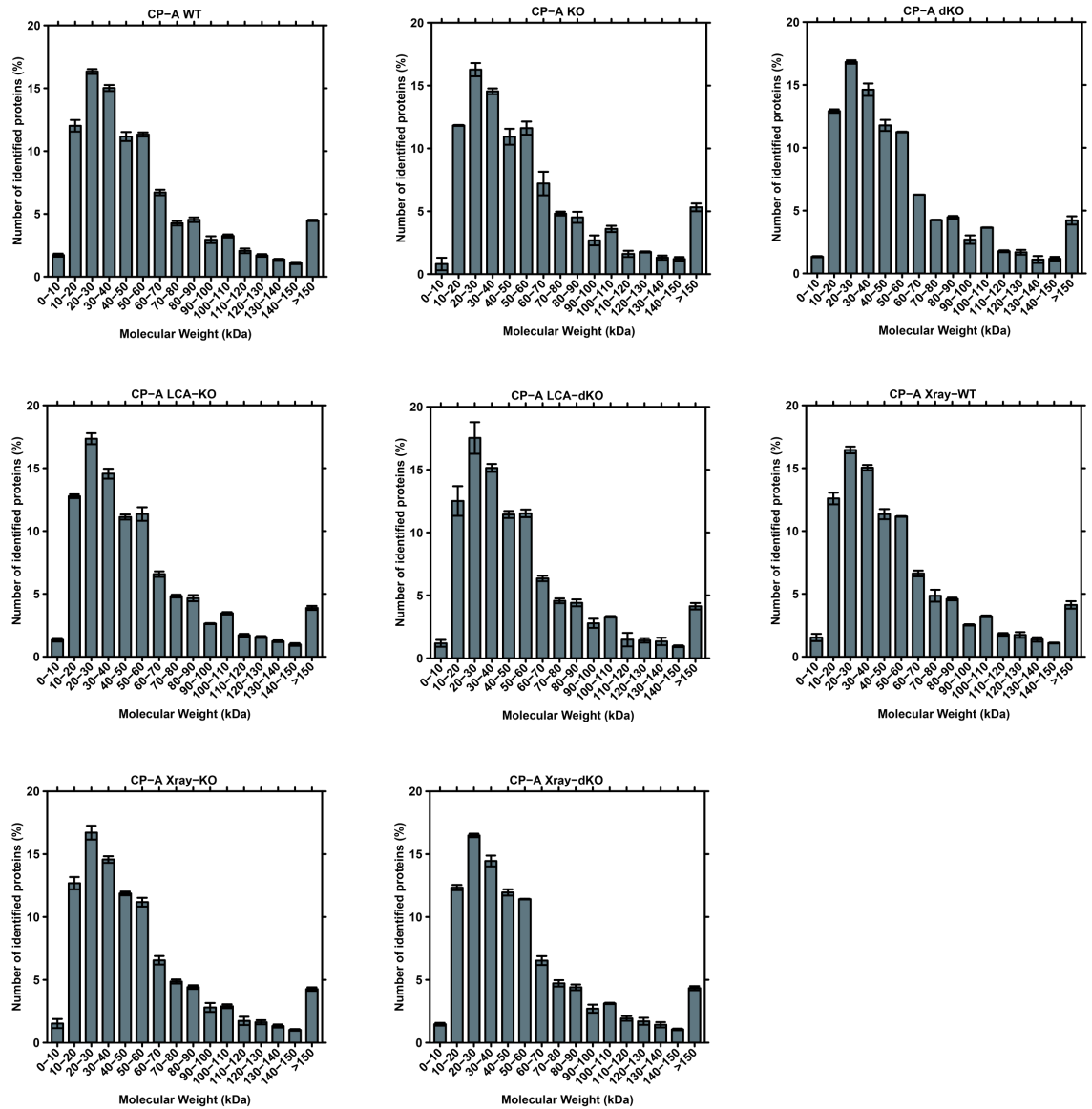
Notes 2-2. Bulk data analysis and comparison to microPOTS data.

The Brno dataset was analyzed using the bioinformatics methods implemented in FragPipe (v14.0). Since the two datasets were generated by different mass spectrometers, we decided to re-analyze the microPOTS data using the same analysis pipeline to avert any subtle variation that could have possibly been introduced while using different bioinformatics

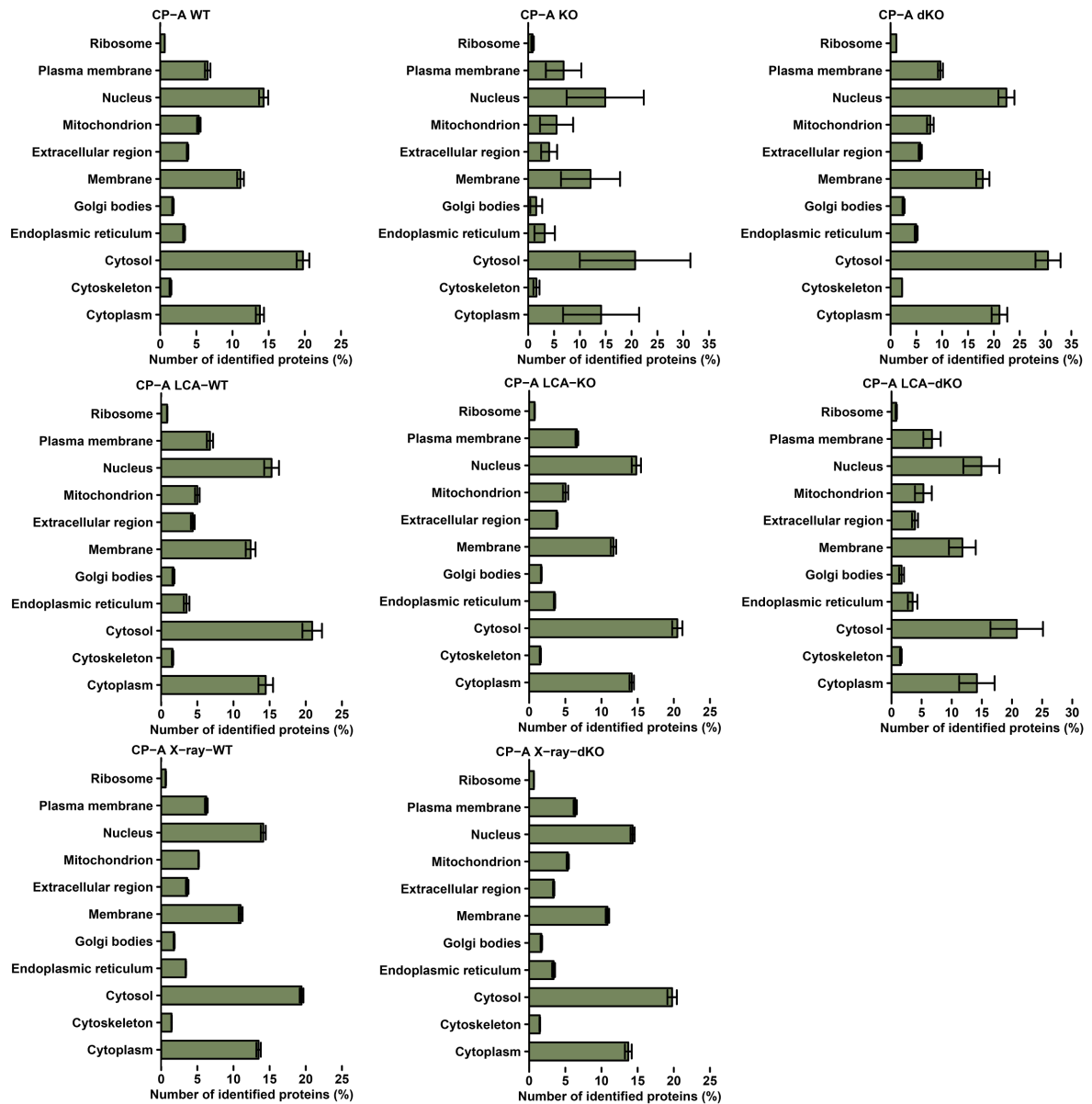
workflows for each dataset. Briefly, MSConvert with peakPicking filter enabled was used to convert the .wiff and .raw files to mzML file format. Subsequently, the files were loaded in the FragPipe software, and peptide identification was performed using the MSFragger software (v3.1.1)¹⁹⁴ to search against the reviewed UniProt human proteome database (2020 release with a total of 20379 reviewed entries). All search parameters were used in their default setting. For the Brno dataset (AB-SCIEX), the MS1 quantification was performed with lonQuant activated and *m/z* window (ppm) and RT window (minutes) set to 30 and 8 respectively. For the microPOTS dataset, the MS1 quantification was performed with lonQuant activated and *m/z* window (ppm) and RT window (minutes) set to 10 and 5 respectively. In either case, the Protein quant was set to MaxLFQ. The resulting .tsv file was imported to the R statistical environment (v4.0.3) for correlation analysis and subsequent visualization with the ggplot2 package.



Supplementary Figure 2-6. Assessment of physicochemical characteristics. Distribution of GRAVY scores of the identified proteins in each sample type.



Supplementary Figure 2-7. Assessment of physicochemical characteristics. Distribution of molecular weight (kDa) of the identified proteins in each sample type.



Supplementary Figure 2-8. Subcellular localization. Distribution of the identified proteins in various subcellular locations for each sample type.

| CP-A Cells | WT | KO | dKO | LCA-WT | LCA-KO | LCA-dKO | Xray-WT | Xray-KO | Xray-dKO |
|---------------|-------|-------|-------|--------|--------|---------|---------|---------|----------|
| Median CV (%) | 14.08 | 64.61 | 17.41 | 26.78 | 22.15 | 41.65 | 18.67 | 26.49 | 14.97 |

Supplementary Table 2-1. Median coefficient of variation (CV) for quantile normalized protein LFQ values for each sample type.

| No. | Protein name | UniProt ID | Ratio [‡] | Adjusted p.values | Direction [§] |
|-----|-------------------|------------|--------------------|-------------------|------------------------|
| 1 | ARF4 | P18085 | -2.87 | 0.0301 | ↓ |
| 2 | ASNS [†] | P08243 | 2.97 | 0.000801 | ↑ |
| 3 | CHCHD2 | Q9Y6H1 | -1.76 | 0.0091 | ↓ |
| 4 | CSRP1 | P21291 | 3.1 | 2.52E-09 | ↑ |
| 5 | CTSD | P07339 | -1.41 | 0.043 | ↓ |
| 6 | DBI | P07108 | 2.89 | 0.0115 | ↑ |
| 7 | DDX47 | Q9H0S4-2 | -2.02 | 0.047 | ↓ |
| 8 | DSP | P15924 | -1.8 | 0.0119 | ↓ |
| 9 | EPHA2 | P29317 | 1.34 | 0.0115 | ↑ |
| 10 | ESD | P10768 | 3.23 | 2.63E-05 | ↑ |
| 11 | FAM120A | Q9NZB2 | 1.54 | 0.00151 | ↑ |
| 12 | FKBP3 | Q00688 | 1.71 | 0.0393 | ↑ |
| 13 | G3BP2 | Q9UN86-2 | -2.39 | 0.04 | ↓ |
| 14 | GSR | P00390-5 | 1.35 | 0.0234 | ↑ |
| 15 | GTF2I | P78347-2 | -1.17 | 0.0364 | ↓ |
| 16 | HIST1H1E | P10412 | 1.13 | 0.00111 | ↑ |
| 17 | HNRNP3 | P31942-2 | -2.56 | 9.95E-07 | ↓ |
| 18 | IDH1 | O75874 | 0.913 | 0.00537 | ↑ |
| 19 | IL18 | Q14116 | 4.08 | 0.00105 | ↑ |
| 20 | LAMB3 | Q13751 | 1.37 | 0.0174 | ↑ |
| 21 | LAMC2 | Q13753 | 2 | 0.0459 | ↑ |
| 22 | LARP1 | Q6PKG0 | -1.45 | 0.00684 | ↓ |
| 23 | LRPPRC | P42704 | -0.992 | 0.000234 | ↓ |
| 24 | MCM6 | Q14566 | -1.44 | 0.0485 | ↓ |
| 25 | OCIAD2 | Q56VL3 | -2.4 | 0.00972 | ↓ |
| 26 | RALY | Q9UKM9-2 | 2.32 | 7.72E-11 | ↑ |
| 27 | SAR1A | Q9NR31 | 2.65 | 0.00895 | ↑ |
| 28 | SEC61B | P60468 | 1.35 | 0.00833 | ↑ |
| 29 | SLC25A1 | P53007 | 1.99 | 0.00802 | ↑ |
| 30 | TMX1 | Q9H3N1 | 1.27 | 0.0319 | ↑ |
| 31 | TRMT112 | Q9UI30 | -6.9 | 0.000226 | ↓ |
| 32 | TXNRD1 | Q16881-5 | 2.05 | 0.00937 | ↑ |
| 33 | UBE2M | P61081 | 2.33 | 0.0115 | ↑ |
| 34 | UPP1 | Q16831 | 1.27 | 0.0247 | ↑ |

Supplementary Table 2-2. List of differentially expressed proteins between CP-A dKO cells with or without LCA treatment. ‡ indicates the ratio between the specific value of expression intensity of the protein in LCA-treated CP-A dKO cells and untreated CP-A dKO cells. §

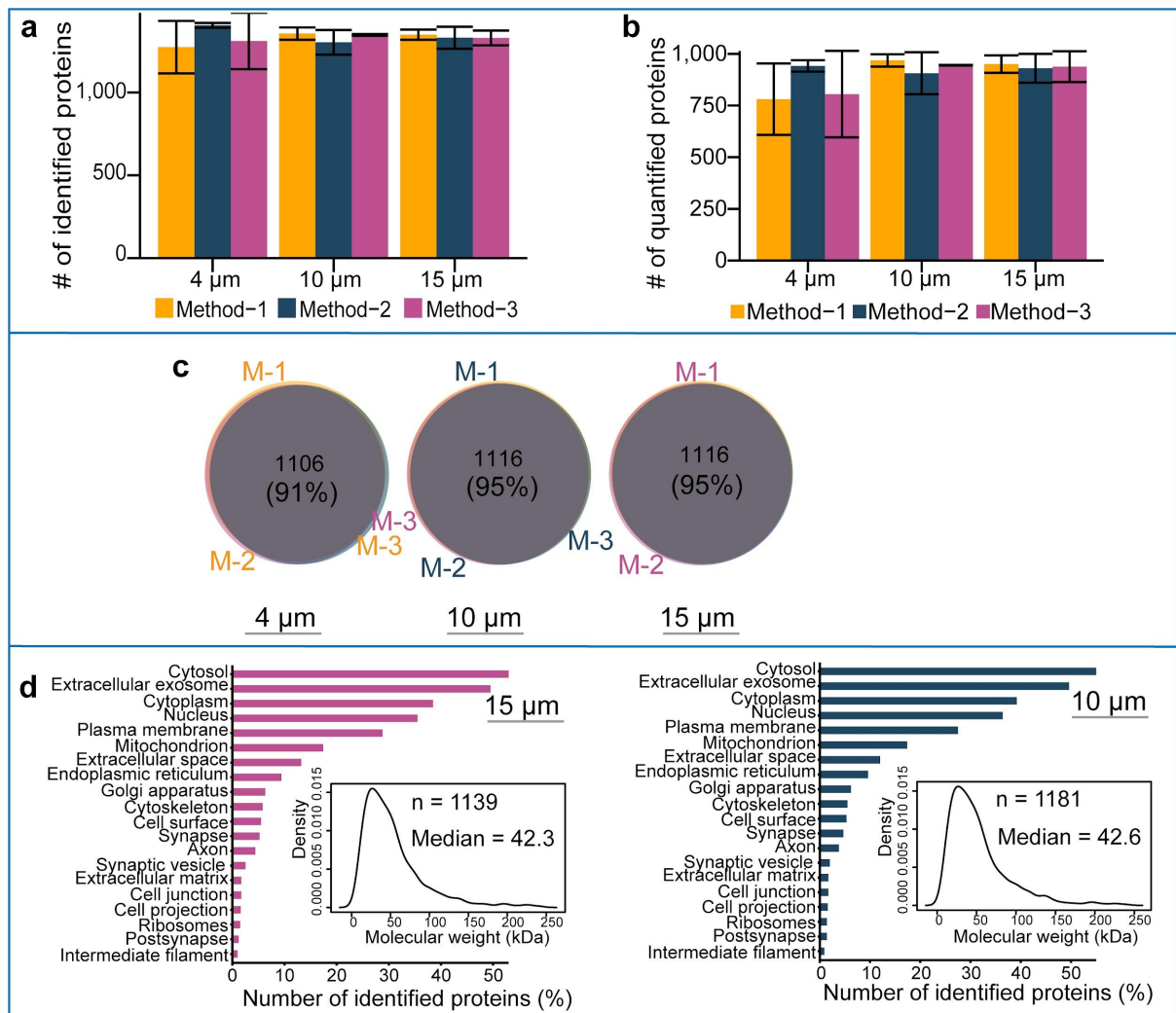
indicates the direction of protein expression, whether upregulation or downregulation. † indicates proteins that were to be consistently differentially expressed with bulk analyses in other studies.

| No. | Protein name | UniProt ID | Ratio [†] | Adjusted p.values | Direction [§] |
|-----|-------------------|------------|--------------------|-------------------|------------------------|
| 1 | AGR2 [†] | O95994 | 1.51 | 0.0192 | ↑ |
| 2 | CALU | O43852 | 2.94 | 0.0102 | ↑ |
| 3 | COPS6 | Q7L5N1 | -1.65 | 0.0281 | ↓ |
| 4 | COX4I1 | P13073 | 2.07 | 0.00846 | ↑ |
| 5 | CSRP1 | P21291 | 3.22 | 8.56E-11 | ↑ |
| 6 | CYC1 | P08574 | 1.59 | 0.0219 | ↑ |
| 7 | DDX47 | Q9H0S4-2 | -2.83 | 0.000393 | ↓ |
| 8 | DHCR24 | Q15392-2 | 1.78 | 0.0211 | ↑ |
| 9 | DTYMK | P23919-2 | 1.88 | 0.0439 | ↑ |
| 10 | EIF1 | P41567 | 1.95 | 0.0429 | ↑ |
| 11 | ESD | P10768 | 2.41 | 0.00313 | ↑ |
| 12 | FAM120A | Q9NZB2 | 1.28 | 0.00878 | ↑ |
| 13 | FARSA | Q9Y285-2 | -2.41 | 0.0189 | ↓ |
| 14 | FKBP3 | Q00688 | 1.95 | 0.00626 | ↑ |
| 15 | GLUL | P15104 | 2.69 | 0.000427 | ↑ |
| 16 | HIST1H1A | Q02539 | -1.22 | 0.0167 | ↓ |
| 17 | HIST1H2AJ | Q99878 | -1.26 | 0.00258 | ↓ |
| 18 | HNRNPH3 | P31942-2 | -1.88 | 0.000655 | ↓ |
| 19 | IFI16 | Q16666-3 | 1.56 | 0.000387 | ↑ |
| 20 | RALY | Q9UKM9-2 | 2.43 | 8.66E-13 | ↑ |
| 21 | IPO9 | Q96P70 | -2.66 | 0.00507 | ↓ |
| 22 | LDLR | P01130-3 | -2.12 | 0.00873 | ↓ |
| 23 | LIMA1 | Q9UHB6-2 | 1.24 | 0.0143 | ↑ |
| 24 | MRTO4 | Q9UKD2 | 1.81 | 0.0482 | ↑ |
| 25 | MTDH | Q86UE4 | 1.91 | 8.91E-05 | ↑ |
| 26 | NDUFS3 | O75489 | -1.38 | 0.048 | ↓ |
| 27 | NONO | Q15233-2 | 2.4 | 0.00373 | ↑ |
| 28 | OCIAD1 | Q9NX40 | 1.75 | 0.0167 | ↑ |
| 29 | PLEC.1 | Q15149-4 | 0.931 | 0.000371 | ↑ |
| 30 | PRRC2C | Q9Y520-4 | 1.57 | 0.000229 | ↑ |
| 31 | PSMC3 | P17980 | -3.1 | 0.00409 | ↓ |
| 32 | PSMD14 | O00487 | -2.39 | 0.0424 | ↓ |
| 33 | PSMD7 | P51665 | 1.25 | 0.0482 | ↑ |
| 34 | RAB21 | Q9UL25 | -2.32 | 0.0187 | ↓ |

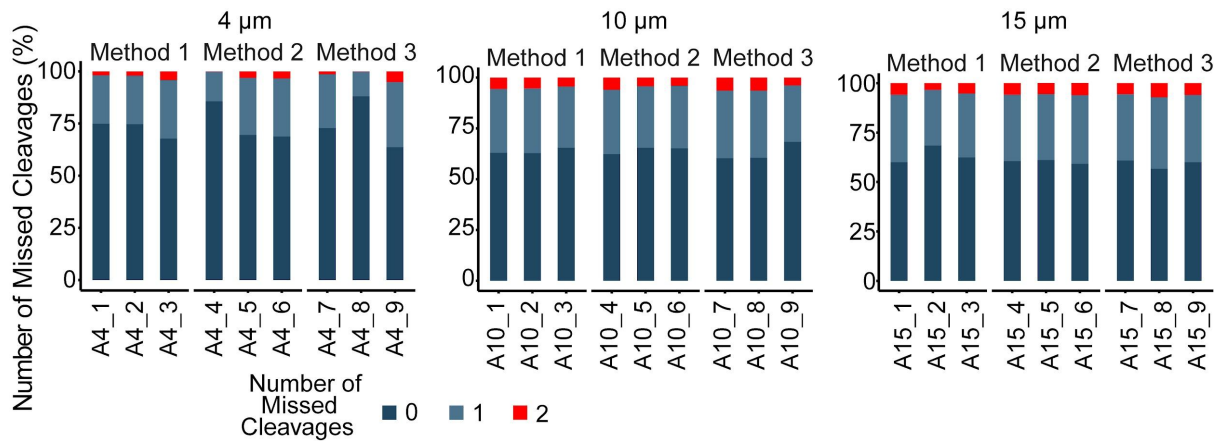
| | | | | | |
|----|---------------------|----------|-------|----------|---|
| 35 | S100A9 [†] | P06702 | 3.47 | 5.13E-07 | ↑ |
| 36 | S100P | P25815 | -4.97 | 2.63E-09 | ↓ |
| 37 | SAR1A | Q9NR31 | 2.55 | 0.00781 | ↑ |
| 38 | SEC13 | P55735-2 | 1.14 | 0.00888 | ↑ |
| 39 | SEC61B | P60468 | 1.24 | 0.0116 | ↑ |
| 40 | SH3BGRL3 | Q9H299 | 3.31 | 0.00526 | ↑ |
| 41 | ZNF207 | O43670-2 | 1.73 | 0.00965 | ↑ |
| 42 | TCEA1 | P23193 | -2.12 | 0.0211 | ↓ |
| 43 | THRAP3 | Q9Y2W1 | 1.32 | 0.0408 | ↑ |
| 44 | TMED10 | P49755 | 2.72 | 0.0333 | ↑ |
| 45 | TMX1 | Q9H3N1 | 1.14 | 0.0453 | ↑ |
| 46 | UBE2M | P61081 | 2.91 | 0.00025 | ↑ |

Supplementary Table 2-3. List of differentially expressed proteins between CP-A dKO cells with or without X-ray treatment. ‡ indicates the ratio between the specific value of expression intensity of the protein in X-ray-treated CP-A dKO cells and untreated CP-A dKO cells. § indicates the direction of protein expression, whether upregulation or downregulation. † indicates proteins that were to be consistently differentially expressed with bulk analyses in other studies.

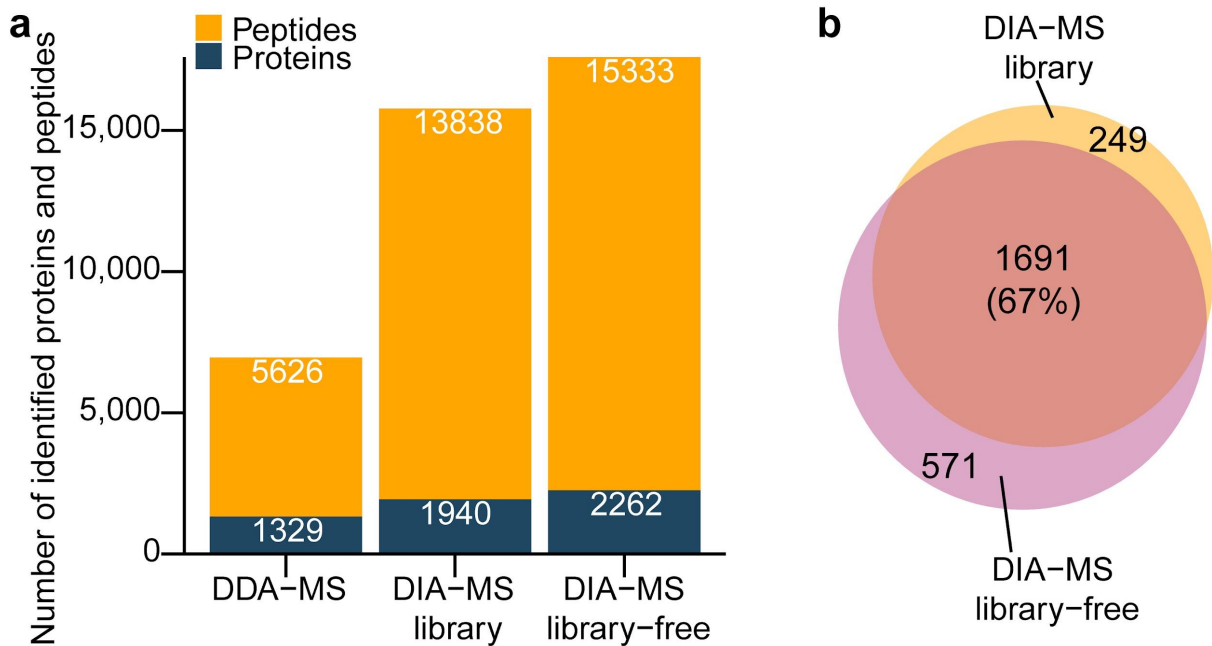
Appendix II: Supplementary material for chapter 3



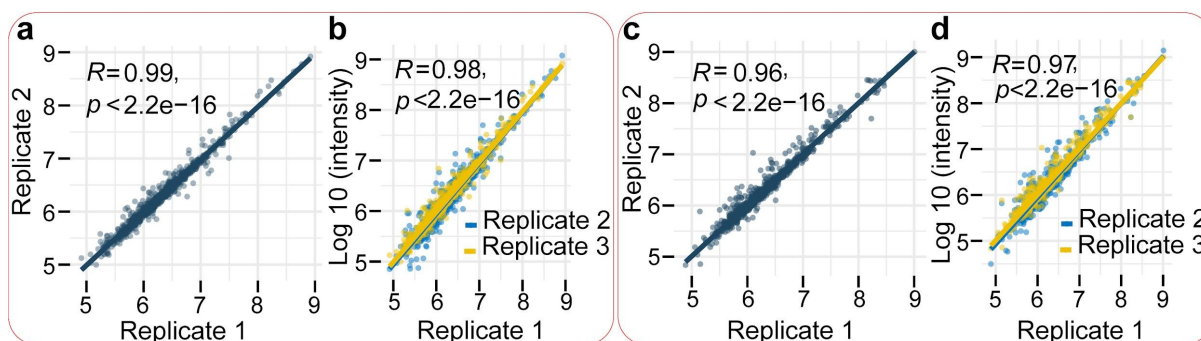
Supplementary Figure 3-1. Method evaluation. **a**, A bar plot representing the number of identified proteins in each tissue section that was analyzed (cello and mulberry correspond to 4 μm , 10 μm , and 15 μm sizes, respectively). **b**, Number of quantified proteins. **c**, Area-proportional Venn diagrams illustrating the number and proportion of overlapping identified proteins for the three methods (M-1, M-2, and M-3) for each tissue thickness. **d**, Bar plots showing the proportion of a subset of the gene ontology (GO) terms for the identified proteins (cello and mulberry correspond to 10 μm and 15 μm , respectively). Inset, density plots showing the molecular weight (Mw) distribution of the identified proteins in 10 μm and 15 μm thick GBM tissue sections. The letter n marks the number of proteins.



Supplementary Figure 3-2. Evaluation of digestion efficiency. Bar plots showing the proportion of the missed cleavages sites for Method 1, Method 2, and Method 3 for each tissue thickness (4 μm, 10 μm, or 15 μm). Dark blue, light blue, and red colors represent 0, 1, and 2 missed cleavages, respectively.

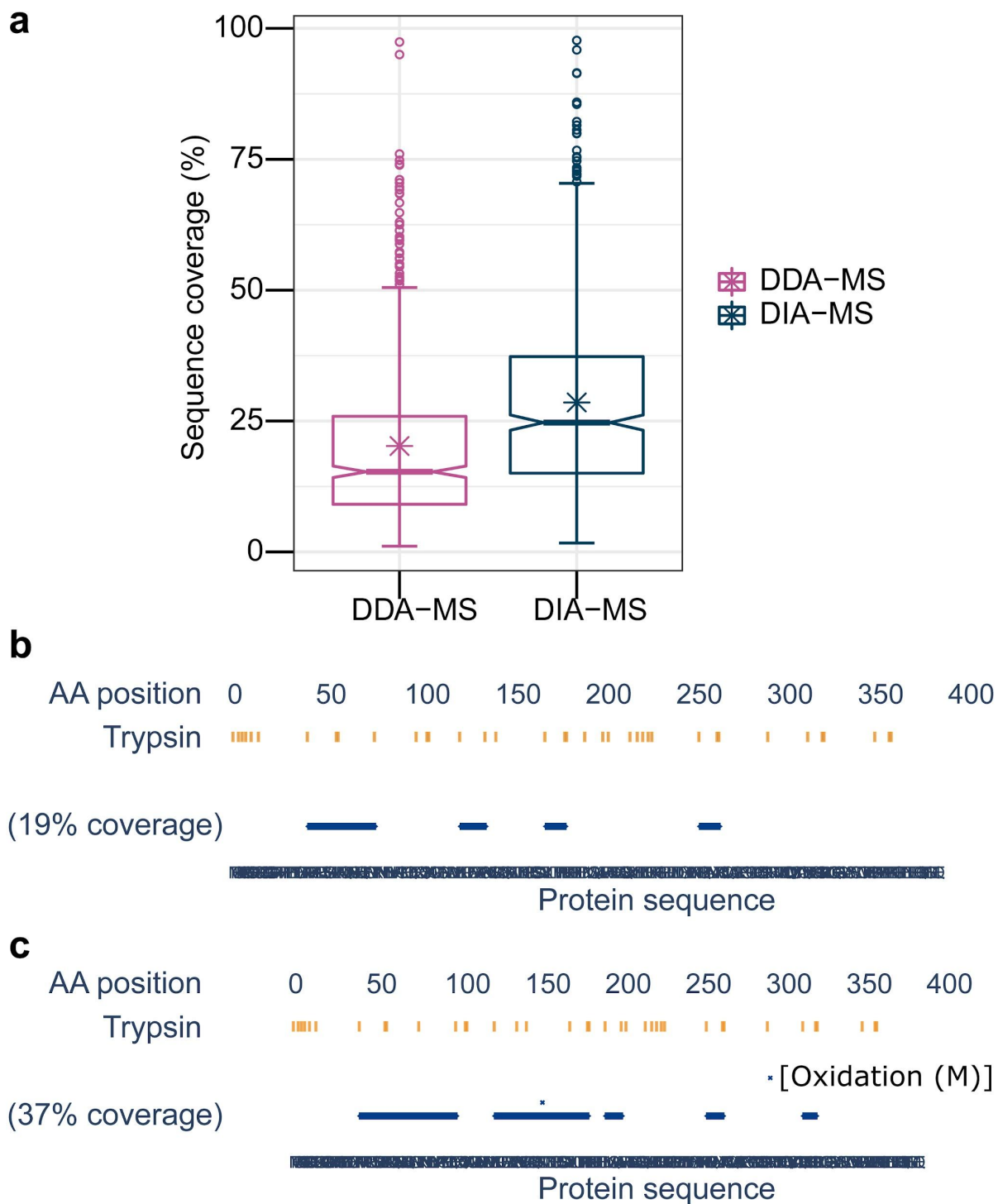


Supplementary Figure 3-3. Comparison of protein identifications using different approaches. **a**, Bar plots showing the number of identified peptides and proteins with DDA-MS, DIA-MS in a spectral library mode and DIA-MS in library-free mode. **b**, Area-proportional Venn diagram showing the number of proteins shared and distinct between spectral library and library-free modes.



Supplementary Figure 3-4. Quantitative comparison of DIA-NN library and library-free approaches. **a**, Pearson's correlation coefficient analysis between two technical replicates for data processed using a library approach. **b**, Two technical replicates were combined and correlated with a third technical replicate for data processed with spectral library approach. **c** and **d**, same as **a** and **b**, but with data processed using a library-free approach.

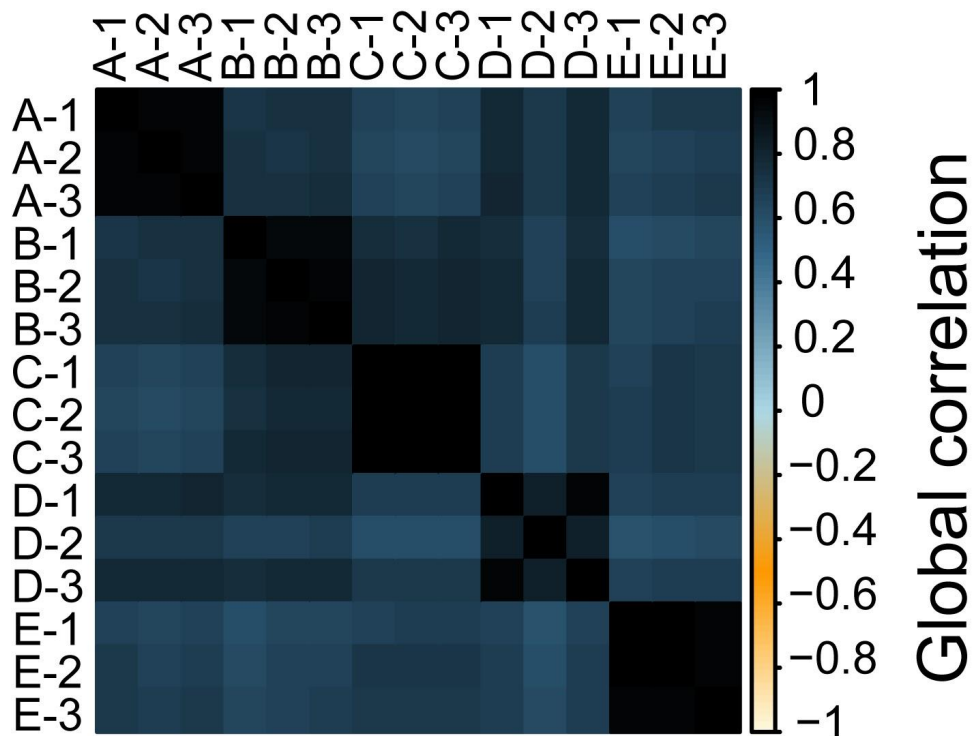
Pearson's correlation coefficient analysis between technical replicates for data processed with either library or library-free revealed a relatively higher R value for the library than library-free approach (**Supplementary Figure 3-4a, b, c, and d**). As illustrated in **Supplementary Figure 3-3**, the library-free quantification provided more hits than library approach. However, the quantification precision based on the transition picked up by library-free approach seems to be less precise compared to the quantification based on the transition picked up by library approach. Therefore, we decided to proceed with the data processed with the spectral library. It is important to note that a thorough comparison of the quantitative precision of two approaches might be necessary prior to deciding the approach to use for some data.



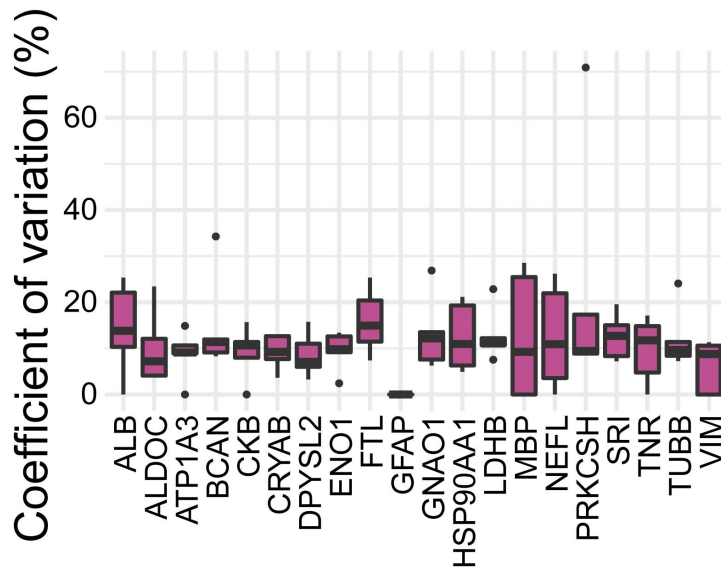
Supplementary Figure 3-5. Comparison of sequence coverage by DDA-MS and DIA-MS. **a**, Notched boxplot comparison of sequence coverage between DDA-MS and DIA-MS. The centerlines show the medians, box limits indicate the 25th and 75th percentiles and the mean sequence coverage is marked by the asterisks. **b**, Sequence plot for Interleukin enhancer-binding factor 2 (ILF2) - Q12905 for DDA-MS analysis showing the percent

sequence coverage and amino acid position within the ILF2 protein sequence. **c**, Same as **b** but for DIA-MS analysis.

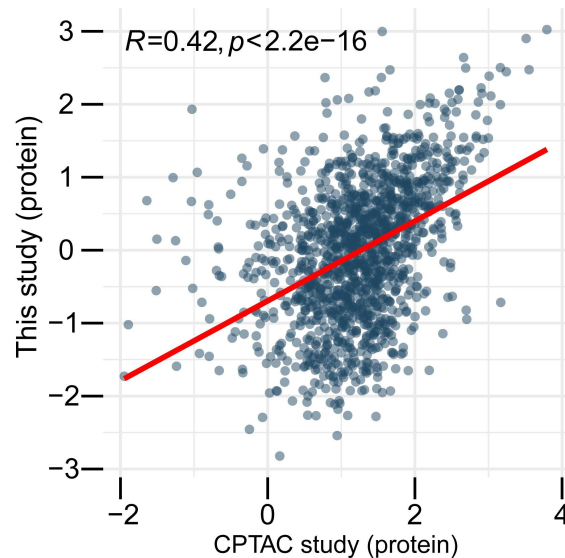
Sequence coverage is the fraction of amino acids covered in the sequence of the identified protein. **Supplementary Figure 3-5a** shows the distribution of the sequence coverage for all the identified proteins using either DDA-MS or DIA-MS. We also compared the sequence coverage between DDA-MS and DIA-MS using a single protein. For this comparison, we randomly selected one of the immune-relevant proteins (ILF2) and determined its sequence coverage in either DDA-MS or DIA-MS data. DIA-MS reported a higher sequence coverage (37%) for ILF2 compared to DDA-MS, which reported 19% (**Supplementary Figure 3-5b,c**).



Supplementary Figure 3-6. Global correlation matrix of protein relative expression levels between samples. The protein intensity level in each sample is compared to intensity levels in all other samples, and Pearson's correlation coefficient is calculated, where an R value of 1 and black color indicate a strong positive correlation.



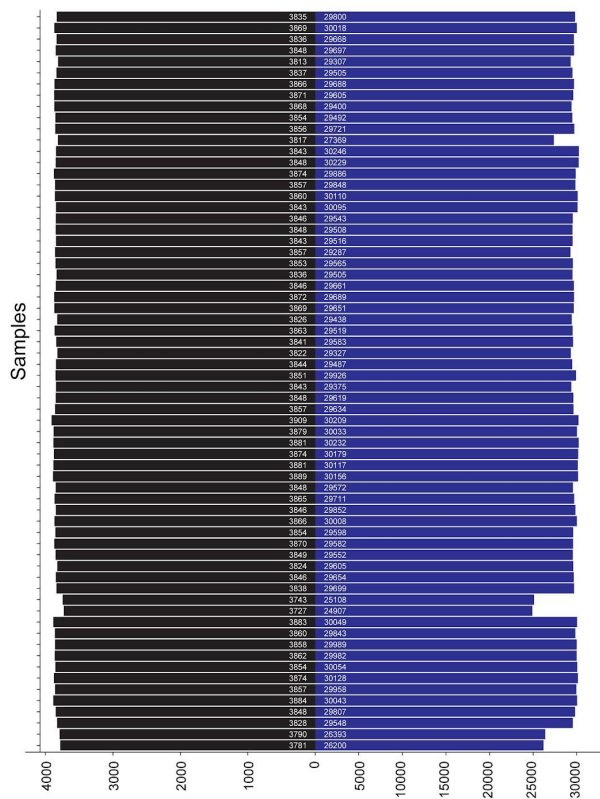
Supplementary Figure 3-7. Boxplot showing the interpatient CV distribution for the top 20 proteins. Each dot represents a CV that is calculated based on the intensity values of the individual protein in all tissue sections for the patient in case.



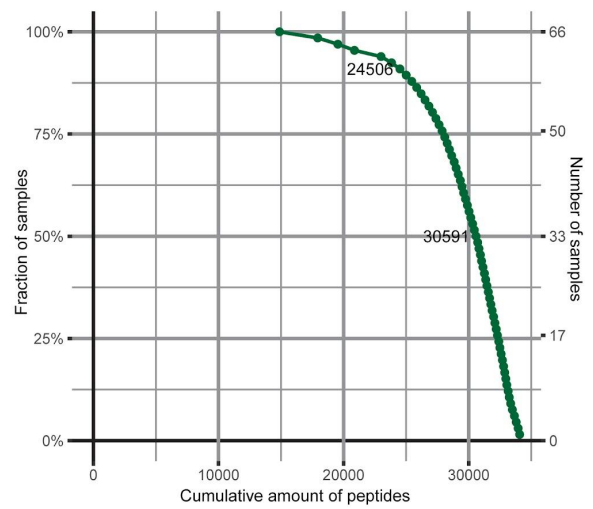
Supplementary Figure 3-8. Scatter plot showing Pearson's correlation coefficient ($R = 0.42$) between DIA-MS (this study) and DDA-MS (CPTAC - Clinical Proteomic Tumor Analysis Consortium) proteomics data. Protein values were transformed (\log_{10}) and subsequently Z-scored to standardize the data. Each dot represents the expression intensity value of an individual protein, and the red line indicates the regression line.

Appendix IV: Supplementary material for chapter 4

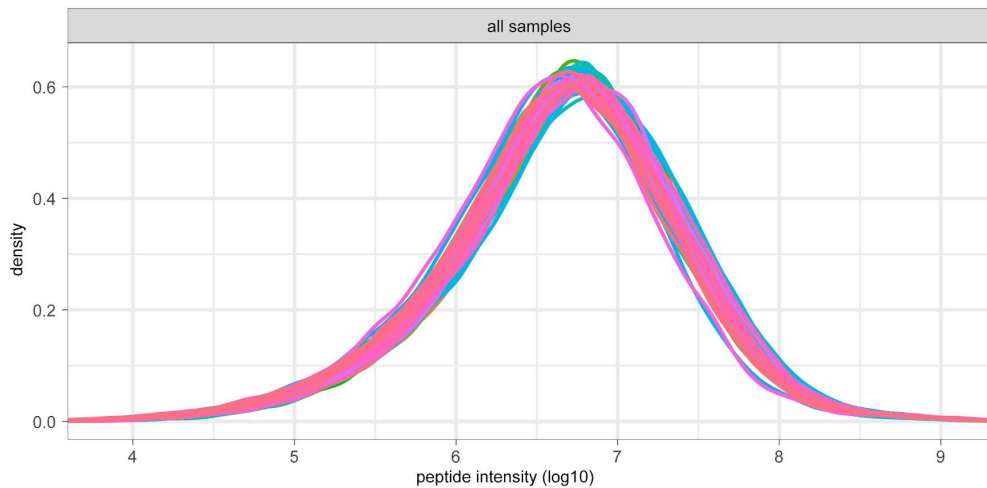
a, Number of proteins/peptides identified and quantified



b, Completeness of data in the entire dataset



c, Abundance distributions



Supplementary Figure 4-1. Overview of proteomics data. **a**, Mirrored bar plots representing the number of proteins (in black) and peptides (in blue) identified and quantified with horizontal axis displaying the counts while the vertical axis showing bars for respective

samples. **b**, Scatter plot illustrating data completeness in the entire dataset. **c**, density plot showing the abundance distribution of all measured samples.

References

1. Cigudosa, J. C. *et al.* Cytogenetic analysis of 363 consecutively ascertained diffuse large B-cell lymphomas. *Genes. Chromosomes Cancer* **25**, 123–133 (1999).
2. Hanselmann, R. G. & Welter, C. Origin of Cancer: Cell work is the Key to Understanding Cancer Initiation and Progression. *Front. Cell Dev. Biol.* **10**, (2022).
3. Hassanpour, S. H. & Dehghani, M. Review of cancer from perspective of molecular. *J. Cancer Res. Pract.* **4**, 127–129 (2017).
4. Kashyap, A. K. & Dubey, S. K. CHAPTER 5 - Molecular mechanisms in cancer development. in *Understanding Cancer* (eds. Jain, B. & Pandey, S.) 79–90 (Academic Press, 2022). doi:10.1016/B978-0-323-99883-3.00016-0.
5. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
6. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
7. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* **12**, 31–46 (2022).
8. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
9. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA. Cancer J. Clin.* **72**, 7–33 (2022).
10. Ferlay, J. *et al.* Cancer statistics for the year 2020: An overview. *Int. J. Cancer* **149**, 778–789 (2021).
11. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
12. Brennan, C. W. *et al.* The Somatic Genomic Landscape of Glioblastoma. *Cell* **155**, 462–477 (2013).
13. Wu, G. *et al.* The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nat. Genet.* **46**, 444–450 (2014).
14. Sawada, G. *et al.* Genomic Landscape of Esophageal Squamous Cell Carcinoma in a

- Japanese Population. *Gastroenterology* **150**, 1171–1182 (2016).
15. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
 16. Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
 17. Zhang, B. *et al.* Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382–387 (2014).
 18. Song, Y.-C. *et al.* Proteogenomic characterization and integrative analysis of glioblastoma multiforme. *Oncotarget* **8**, 97304–97312 (2017).
 19. Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 (2016).
 20. Yanovich-Arad, G. *et al.* Proteogenomics of glioblastoma associates molecular patterns with survival. *Cell Rep.* **34**, (2021).
 21. Wang, L.-B. *et al.* Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell* **39**, 509-528.e20 (2021).
 22. Bauernfeind, A. L. & Babbitt, C. C. The predictive nature of transcript expression levels on protein expression in adult human brain. *BMC Genomics* **18**, 322 (2017).
 23. Kwon, Y. W. *et al.* Application of Proteomics in Cancer: Recent Trends and Approaches for Biomarkers Discovery. *Front. Med.* **8**, (2021).
 24. Dolecek, T. A., Propp, J. M., Stroup, N. E. & Kruchko, C. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2005–2009. *Neuro-Oncol.* **14**, v1–v49 (2012).
 25. Miller, K. D. *et al.* Brain and other central nervous system tumor statistics, 2021. *CA. Cancer J. Clin.* **71**, 381–406 (2021).
 26. Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol. (Berl.)* **131**, 803–820 (2016).
 27. Louis, D. N. *et al.* The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-Oncol.* **23**, 1231–1251 (2021).

28. Gritsch, S., Batchelor, T. T. & Gonzalez Castro, L. N. Diagnostic, therapeutic, and prognostic implications of the 2021 World Health Organization classification of tumors of the central nervous system. *Cancer* **128**, 47–58 (2022).
29. Liu, X.-Y. *et al.* Frequent ATRX mutations and loss of expression in adult diffuse astrocytic tumors carrying IDH1/IDH2 and TP53 mutations. *Acta Neuropathol. (Berl.)* **124**, 615–625 (2012).
30. Ebrahimi, A. *et al.* ATRX immunostaining predicts IDH and H3F3A status in gliomas. *Acta Neuropathol. Commun.* **4**, 60 (2016).
31. Yang, J. *et al.* Cyclin-Dependent Kinase Inhibitor 2A/B Homozygous Deletion Prediction and Survival Analysis. *Brain Sci.* **13**, 548 (2023).
32. Boots-Sprenger, S. H. E. *et al.* Significance of complete 1p/19q co-deletion, IDH1 mutation and MGMT promoter methylation in gliomas: use with caution. *Mod. Pathol.* **26**, 922–929 (2013).
33. Wang, T.-J. *et al.* Comparisons of Tumor Suppressor p53, p21, and p16 Gene Therapy Effects on Glioblastoma Tumorigenicity in Situ. *Biochem. Biophys. Res. Commun.* **287**, 173–180 (2001).
34. Cairncross, G. *et al.* Phase III Trial of Chemoradiotherapy for Anaplastic Oligodendroglioma: Long-Term Results of RTOG 9402. *J. Clin. Oncol.* **31**, 337–343 (2013).
35. Rebocho, A. P. & Marais, R. ARAF acts as a scaffold to stabilize BRAF:CRAF heterodimers. *Oncogene* **32**, 3207–3212 (2013).
36. Brat, D. J. *et al.* cIMPACT-NOW update 3: recommended diagnostic criteria for “Diffuse astrocytic glioma, IDH-wildtype, with molecular features of glioblastoma, WHO grade IV”. *Acta Neuropathol. (Berl.)* **136**, 805–810 (2018).
37. Tesileanu, C. M. S. *et al.* Survival of diffuse astrocytic glioma, IDH1/2 wildtype, with molecular features of glioblastoma, WHO grade IV: a confirmation of the cIMPACT-NOW criteria. *Neuro-Oncol.* **22**, 515–523 (2020).
38. Weller, M. *et al.* MGMT promoter methylation in malignant gliomas: ready for

- personalized medicine? *Nat. Rev. Neurol.* **6**, 39–51 (2010).
39. Grochans, S. *et al.* Epidemiology of Glioblastoma Multiforme—Literature Review. *Cancers* **14**, 2412 (2022).
 40. Grech, N. *et al.* Rising Incidence of Glioblastoma Multiforme in a Well-Defined Population. *Cureus* **12**, (2020).
 41. Tamimi, A. F. & Juweid, M. Epidemiology and Outcome of Glioblastoma. in *Glioblastoma* (ed. De Vleeschouwer, S.) (Codon Publications, 2017).
 42. McKinnon, C., Nandhabalan, M., Murray, S. A. & Plaha, P. Glioblastoma: clinical presentation, diagnosis, and management. *BMJ* **374**, n1560 (2021).
 43. De Biase, G., Garcia, D. P., Bohnen, A. & Quiñones-Hinojosa, A. Perioperative Management of Patients with Glioblastoma. *Neurosurg. Clin. N. Am.* **32**, 1–8 (2021).
 44. Brown, T. J. *et al.* Association of the Extent of Resection With Survival in Glioblastoma: A Systematic Review and Meta-analysis. *JAMA Oncol.* **2**, 1460–1469 (2016).
 45. Roa, W. *et al.* Abbreviated Course of Radiation Therapy in Older Patients With Glioblastoma Multiforme: A Prospective Randomized Clinical Trial. *J. Clin. Oncol.* **22**, 1583–1588 (2004).
 46. Stupp, R. *et al.* Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005).
 47. Robert, C. A decade of immune-checkpoint inhibitors in cancer therapy. *Nat. Commun.* **11**, 3801 (2020).
 48. Leach, D. R., Krummel, M. F. & Allison, J. P. Enhancement of Antitumor Immunity by CTLA-4 Blockade. *Science* **271**, 1734–1736 (1996).
 49. Rizvi, N. A. *et al.* Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
 50. Sener, U., Ruff, M. W. & Campian, J. L. Immunotherapy in Glioblastoma: Current Approaches and Future Perspectives. *Int. J. Mol. Sci.* **23**, 7046 (2022).
 51. Bausart, M., Pr at, V. & Malfanti, A. Immunotherapy for glioblastoma: the promise of combination strategies. *J. Exp. Clin. Cancer Res.* **41**, 35 (2022).

52. Pinheiro, S. L. R. *et al.* Immunotherapy in glioblastoma treatment: Current state and future prospects. *World J. Clin. Oncol.* **14**, 138–159 (2023).
53. Zanders, E. D., Svensson, F. & Bailey, D. S. Therapy for glioblastoma: is it working? *Drug Discov. Today* **24**, 1193–1201 (2019).
54. Yang, K. *et al.* Glioma targeted therapy: insight into future of molecular approaches. *Mol. Cancer* **21**, 39 (2022).
55. Jain, K. K. A Critical Overview of Targeted Therapies for Glioblastoma. *Front. Oncol.* **8**, (2018).
56. Kaka, N. *et al.* Progression-Free but No Overall Survival Benefit for Adult Patients with Bevacizumab Therapy for the Treatment of Newly Diagnosed Glioblastoma: A Systematic Review and Meta-Analysis. *Cancers* **11**, 1723 (2019).
57. Gramatzki, D. *et al.* Bevacizumab may improve quality of life, but not overall survival in glioblastoma: an epidemiological study. *Ann. Oncol.* **29**, 1431–1436 (2018).
58. Taylor, T. E., Furnari, F. B. & Cavenee, W. K. Targeting EGFR for Treatment of Glioblastoma: Molecular Basis to Overcome Resistance. *Curr. Cancer Drug Targets* **12**, 197–209.
59. Westphal, M., Maire, C. L. & Lamszus, K. EGFR as a Target for Glioblastoma Treatment: An Unfulfilled Promise. *CNS Drugs* **31**, 723–735 (2017).
60. Liu, X. *et al.* The third-generation EGFR inhibitor AZD9291 overcomes primary resistance by continuously blocking ERK signaling in glioblastoma. *J. Exp. Clin. Cancer Res.* **38**, 219 (2019).
61. Faisal, S. M. *et al.* The complex interactions between the cellular and non-cellular components of the brain tumor microenvironmental landscape and their therapeutic implications. *Front. Oncol.* **12**, 1005069 (2022).
62. Thol, K., Pawlik, P. & McGranahan, N. Therapy sculpts the complex interplay between cancer and the immune system during tumour evolution. *Genome Med.* **14**, 137 (2022).
63. Prestwich, R. J. *et al.* The Immune System — is it Relevant to Cancer Development, Progression and Treatment? *Clin. Oncol.* **20**, 101–112 (2008).

64. Schreiber, R. D., Old, L. J. & Smyth, M. J. Cancer Immunoediting: Integrating Immunity's Roles in Cancer Suppression and Promotion. *Science* **331**, 1565–1570 (2011).
65. Rabinovich, G. A., Gabrilovich, D. & Sotomayor, E. M. IMMUNOSUPPRESSIVE STRATEGIES THAT ARE MEDIATED BY TUMOR CELLS. *Annu. Rev. Immunol.* **25**, 267–296 (2007).
66. Pishesha, N., Harmand, T. J. & Ploegh, H. L. A guide to antigen processing and presentation. *Nat. Rev. Immunol.* **22**, 751–764 (2022).
67. Kotsias, F., Cebrian, I. & Alloatti, A. Chapter Two - Antigen processing and presentation. in *International Review of Cell and Molecular Biology* (eds. Lhuillier, C. & Galluzzi, L.) vol. 348 69–121 (Academic Press, 2019).
68. Trolle, T. *et al.* The Length Distribution of Class I–Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele–Specific Binding Preference. *J. Immunol.* **196**, 1480–1487 (2016).
69. Kumar, B. *et al.* Tumor collection/processing under physioxia uncovers highly relevant signaling networks and drug sensitivity. *Sci. Adv.* **8**, eabh3375.
70. Chi, J.-T. *et al.* Gene Expression Programs in Response to Hypoxia: Cell Type Specificity and Prognostic Significance in Human Cancers. *PLOS Med.* **3**, e47 (2006).
71. Lee, P., Chandel, N. S. & Simon, M. C. Cellular adaptation to hypoxia through hypoxia inducible factors and beyond. *Nat. Rev. Mol. Cell Biol.* **21**, 268–283 (2020).
72. Bai, R. *et al.* The hypoxia-driven crosstalk between tumor and tumor-associated macrophages: mechanisms and clinical treatment strategies. *Mol. Cancer* **21**, 177 (2022).
73. Facciabene, A. *et al.* Tumour hypoxia promotes tolerance and angiogenesis via CCL28 and Treg cells. *Nature* **475**, 226–230 (2011).
74. Wasinger, V. C. *et al.* Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* **16**, 1090–1094 (1995).
75. Xie, S., Moya, C., Bilgin, B., Jayaraman, A. & Walton, S. P. Emerging Affinity-Based Techniques in Proteomics. *Expert Rev. Proteomics* **6**, 573–583 (2009).

76. Baker, E. S. *et al.* Mass spectrometry for translational proteomics: progress and clinical implications. *Genome Med.* **4**, 63 (2012).
77. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C. & Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **113**, 2343–2394 (2013).
78. Botelho, D. *et al.* Top-Down and Bottom-Up Proteomics of SDS-Containing Solutions Following Mass-Based Separation. *J. Proteome Res.* **9**, 2863–2870 (2010).
79. Yu, Y.-Q., Gilar, M., Lee, P. J., Bouvier, E. S. P. & Gebler, J. C. Enzyme-Friendly, Mass Spectrometry-Compatible Surfactant for In-Solution Enzymatic Digestion of Proteins. *Anal. Chem.* **75**, 6023–6028 (2003).
80. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
81. Ctorteczka, C. & Mechtler, K. The rise of single-cell proteomics. *Anal. Sci. Adv.* **2**, 84–94 (2021).
82. Zhu, Y. *et al.* Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.* **9**, 882 (2018).
83. Liang, Y. *et al.* Fully Automated Sample Processing and Analysis Workflow for Low-Input Proteome Profiling. *Anal. Chem.* **93**, 1658–1666 (2021).
84. Wang, H. *et al.* Development and Evaluation of a Micro- and Nanoscale Proteomic Sample Preparation Method. *J. Proteome Res.* **4**, 2397–2403 (2005).
85. Wang, N., Xu, M., Wang, P. & Li, L. Development of Mass Spectrometry-Based Shotgun Method for Proteome Analysis of 500 to 5000 Cancer Cells. *Anal. Chem.* **82**, 2262–2271 (2010).
86. Li, S. *et al.* An Integrated Platform for Isolation, Processing, and Mass Spectrometry-based Proteomic Profiling of Rare Cells in Whole Blood. *Mol. Cell. Proteomics MCP* **14**, 1672–1683 (2015).
87. Zhang, Z., Dubiak, K. M., Huber, P. W. & Dovichi, N. J. Miniaturized Filter-Aided Sample Preparation (MICRO-FASP) Method for High Throughput, Ultrasensitive Proteomics Sample Preparation Reveals Proteome Asymmetry in *Xenopus laevis* Embryos. *Anal.*

- Chem.* **92**, 5554–5560 (2020).
88. Martin, K. *et al.* Facile One-Pot Nanoproteomics for Label-Free Proteome Profiling of 50–1000 Mammalian Cells. *J. Proteome Res.* (2021)
doi:10.1021/acs.jproteome.1c00403.
89. Tsai, C.-F. *et al.* Surfactant-assisted one-pot sample preparation for label-free single-cell proteomics. *Commun. Biol.* **4**, 1–12 (2021).
90. Ethier, M., Hou, W., Duewel, H. S. & Figeys, D. The Proteomic Reactor: A Microfluidic Device for Processing Minute Amounts of Protein Prior to Mass Spectrometry Analysis. *J. Proteome Res.* **5**, 2754–2759 (2006).
91. Tian, R., Alvarez-Saavedra, M., Cheng, H.-Y. M. & Figeys, D. Uncovering the Proteome Response of the Master Circadian Clock to Light Using an AutoProteome System*. *Mol. Cell. Proteomics* **10**, M110.007252 (2011).
92. Chen, W. *et al.* Simple and Integrated Spintip-Based Technology Applied for Deep Proteome Profiling. *Anal. Chem.* **88**, 4864–4871 (2016).
93. Wen, B. & Zhu, M. Applications of mass spectrometry in drug metabolism: 50 years of progress. *Drug Metab. Rev.* **47**, 71–87 (2015).
94. Wang, J. *et al.* Portable mass spectrometry system: instrumentation, applications, and path to 'omics analysis. *PROTEOMICS* **22**, 2200112 (2022).
95. Neagu, A.-N. *et al.* Applications of Tandem Mass Spectrometry (MS/MS) in Protein Analysis for Biomedical Research. *Molecules* **27**, 2411 (2022).
96. Wetie, A. G. N., Woods, A. G. & Darie, C. C. Mass Spectrometric Analysis of Post-translational Modifications (PTMs) and Protein–Protein Interactions (PPIs). in *Advancements of Mass Spectrometry in Biomedical Research* (eds. Woods, A. G. & Darie, C. C.) 205–235 (Springer International Publishing, 2014).
doi:10.1007/978-3-319-06068-2_9.
97. Yakubu, R. R., Nieves, E. & Weiss, L. M. The Methods Employed in Mass Spectrometric Analysis of Posttranslational Modifications (PTMs) and Protein–Protein Interactions (PPIs). in *Advancements of Mass Spectrometry in Biomedical Research* (eds. Woods, A.

- G. & Darie, C. C.) 169–198 (Springer International Publishing, 2019).
doi:10.1007/978-3-030-15950-4_10.
98. Zecha, J. *et al.* Linking post-translational modifications and protein turnover by site-resolved protein turnover profiling. *Nat. Commun.* **13**, 165 (2022).
99. El-Aneed, A., Cohen, A. & Banoub, J. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Appl. Spectrosc. Rev.* **44**, 210–230 (2009).
100. Rankin-Turner, S., Sears, P. & Heaney, L. M. Applications of ambient ionization mass spectrometry in 2022: An annual review. *Anal. Sci. Adv.* **4**, 133–153 (2023).
101. Hofstadler, S. A., Bakhtiar, R. & Smith, R. D. Electrospray Ionization Mass Spectroscopy: Part I. Instrumentation and Spectral Interpretation. *J. Chem. Educ.* **73**, A82 (1996).
102. Haag, A. M. Mass Analyzers and Mass Spectrometers. in *Modern Proteomics – Sample Preparation, Analysis and Practical Applications* (eds. Mirzaei, H. & Carrasco, M.) 157–169 (Springer International Publishing, 2016).
doi:10.1007/978-3-319-41448-5_7.
103. Mann, M., Hendrickson, R. C. & Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **70**, 437–473 (2001).
104. Domon, B. & Aebersold, R. Mass Spectrometry and Protein Analysis. *Science* **312**, 212–217 (2006).
105. Glish, G. L. & Burinsky, D. J. Hybrid mass spectrometers for tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **19**, 161–172 (2008).
106. Ortea, I., Rodríguez-Ariza, A., Chicano-Gálvez, E., Arenas Vacas, M. S. & Jurado Gámez, B. Discovery of potential protein biomarkers of lung adenocarcinoma in bronchoalveolar lavage fluid by SWATH MS data-independent acquisition and targeted data extraction. *J. Proteomics* **138**, 106–114 (2016).
107. Li, K. W., Gonzalez-Lozano, M. A., Koopmans, F. & Smit, A. B. Recent Developments in Data Independent Acquisition (DIA) Mass Spectrometry: Application of Quantitative

- Analysis of the Brain Proteome. *Front. Mol. Neurosci.* **13**, 248 (2020).
108. Guan, S., Taylor, P. P., Han, Z., Moran, M. F. & Ma, B. Data Dependent–Independent Acquisition (DDIA) Proteomics. *J. Proteome Res.* **19**, 3230–3237 (2020).
109. Davies, V. *et al.* Rapid Development of Improved Data-Dependent Acquisition Strategies. *Anal. Chem.* **93**, 5676–5683 (2021).
110. Krasny, L. & H. Huang, P. Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. *Mol. Omics* **17**, 29–42 (2021).
111. Kim, Y. J. *et al.* Data-Independent Acquisition Mass Spectrometry To Quantify Protein Levels in FFPE Tumor Biopsies for Molecular Diagnostics. *J. Proteome Res.* **18**, 426–435 (2019).
112. Ting, Y. S. *et al.* PECAN: library-free peptide detection for data-independent acquisition tandem mass spectrometry data. *Nat. Methods* **14**, 903–908 (2017).
113. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2020).
114. Zhi, W., Wang, M. & She, J.-X. Selected reaction monitoring (SRM) mass spectrometry without isotope labeling can be used for rapid protein quantification. *Rapid Commun. Mass Spectrom. RCM* **25**, 1583–1588 (2011).
115. Ankney, J. A., Muneer, A. & Chen, X. Relative and Absolute Quantitation in Mass Spectrometry–Based Proteomics. *Annu. Rev. Anal. Chem.* **11**, 49–77 (2018).
116. Nosti, A. J., Barrio, L. C., Calderón Celis, F., Soldado, A. & Encinar, J. R. Absolute quantification of proteins using element mass spectrometry and generic standards. *J. Proteomics* **256**, 104499 (2022).
117. Lindemann, C. *et al.* Strategies in relative and absolute quantitative mass spectrometry based proteomics. *Biol. Chem.* **398**, 687–699 (2017).
118. Anand, S., Samuel, M., Ang, C.-S., Keerthikumar, S. & Mathivanan, S. Label-Based and Label-Free Strategies for Protein Quantitation. in *Proteome Bioinformatics* (eds. Keerthikumar, S. & Mathivanan, S.) 31–43 (Springer, 2017).

doi:10.1007/978-1-4939-6740-7_4.

119. Mayne, J. *et al.* Bottom-Up Proteomics (2013–2015): Keeping up in the Era of Systems Biology. *Anal. Chem.* **88**, 95–121 (2016).
120. Fagerberg, L. *et al.* Large-Scale Protein Profiling in Human Cell Lines Using Antibody-Based Proteomics. *J. Proteome Res.* **10**, 4066–4075 (2011).
121. Branca, R. M. M. *et al.* HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat. Methods* **11**, 59–62 (2014).
122. Wang, G. *et al.* Balanced-PCR amplification allows unbiased identification of genomic copy changes in minute cell and tissue samples. *Nucleic Acids Res.* **32**, e76 (2004).
123. Parapatics, K. *et al.* Proteomic analysis of low quantities of cellular material in the range obtainable from scarce patient samples. *J. Integr. OMICS* **5**, 30-43–43 (2015).
124. Clair, G. *et al.* Spatially-Resolved Proteomics: Rapid Quantitative Analysis of Laser Capture Microdissected Alveolar Tissue Samples. *Sci. Rep.* **6**, (2016).
125. Coscia, F. *et al.* A streamlined mass spectrometry–based proteomics workflow for large-scale FFPE tissue analysis. *J. Pathol.* **251**, 100–112 (2020).
126. Martin, J. G., Rejtar, T. & Martin, S. A. Integrated Microscale Analysis System for Targeted Liquid Chromatography Mass Spectrometry Proteomics on Limited Amounts of Enriched Cell Populations. *Anal. Chem.* **85**, 10680–10685 (2013).
127. Yamaguchi, H. *et al.* Rapid and efficient proteolysis for proteomic analysis by protease-immobilized microreactor. *Electrophoresis* **30**, 3257–3264 (2009).
128. Myers, S. A. *et al.* Streamlined Protocol for Deep Proteomic Profiling of FAC-sorted Cells and Its Application to Freshly Isolated Murine Immune Cells. *Mol. Cell. Proteomics MCP* **18**, 995–1009 (2019).
129. Gao, W. *et al.* An integrated strategy for high-sensitive and multi-level glycoproteome analysis from low micrograms of protein samples. *J. Chromatogr. A* **1600**, 46–54 (2019).
130. Dou, M. *et al.* Nanowell-mediated two-dimensional liquid chromatography enables deep proteome profiling of <1000 mammalian cells †Electronic supplementary information (ESI) available. See DOI: 10.1039/c8sc02680g. *Chem. Sci.* **9**, 6944–6951

- (2018).
131. Zhu, Y. *et al.* Single-cell proteomics reveals changes in expression during hair-cell development. *eLife* **8**, e50777 (2019).
 132. Shao, X. & Zhang, X. Design of five-layer gold nanoparticles self-assembled in a liquid open tubular column for ultrasensitive nano-LC-MS/MS proteomic analysis of 80 living cells. *PROTEOMICS* **17**, 1600463 (2017).
 133. Chen, Q., Yan, G., Gao, M. & Zhang, X. Ultrasensitive Proteome Profiling for 100 Living Cells by Direct Cell Injection, Online Digestion and Nano-LC-MS/MS Analysis. *Anal. Chem.* **87**, 6674–6680 (2015).
 134. Zhu, Y. *et al.* Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS. *Angew. Chem. Int. Ed.* **57**, 12370–12374 (2018).
 135. Shao, X. *et al.* Integrated Proteome Analysis Device for Fast Single-Cell Protein Profiling. *Anal. Chem.* **90**, 14003–14010 (2018).
 136. Zhou, M. *et al.* Sensitive Top-Down Proteomics Analysis of a Low Number of Mammalian Cells Using a Nanodroplet Sample Processing Platform. *Anal. Chem.* **92**, 7087–7095 (2020).
 137. Xu, K. *et al.* Benchtop-compatible sample processing workflow for proteome profiling of < 100 mammalian cells. *Anal. Bioanal. Chem.* **411**, 4587–4596 (2019).
 138. Ross-Innes, C. S. *et al.* Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat. Genet.* **47**, 1038–1046 (2015).
 139. Peters, Y. *et al.* Barrett oesophagus. *Nat. Rev. Dis. Primer* **5**, 35 (2019).
 140. Hofmann, A. F. The Continuing Importance of Bile Acids in Liver and Intestinal Disease. *Arch. Intern. Med.* **159**, 2647–2658 (1999).
 141. Fitzgerald, R. C. *et al.* Cytosponge-trefoil factor 3 versus usual care to identify Barrett's oesophagus in a primary care setting: a multicentre, pragmatic, randomised controlled trial. *The Lancet* **396**, 333–344 (2020).

142. Ross-Innes, C. S. *et al.* Risk stratification of Barrett's oesophagus using a non-endoscopic sampling method coupled with a biomarker panel: a cohort study. *Lancet Gastroenterol. Hepatol.* **2**, 23–31 (2017).
143. Williams, S. M. *et al.* Automated Coupling of Nanodroplet Sample Preparation with Liquid Chromatography–Mass Spectrometry for High-Throughput Single-Cell Proteomics. *Anal. Chem.* **92**, 10588–10596 (2020).
144. DEP: Differential Enrichment analysis of Proteomics data version 1.10.0 from Bioconductor. <https://rdrr.io/bioc/DEP/>.
145. Zhang, X. *et al.* Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nat. Protoc.* **13**, 530–550 (2018).
146. test_diff: Differential enrichment test in DEP: Differential Enrichment analysis of Proteomics data. https://rdrr.io/bioc/DEP/man/test_diff.html.
147. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
148. P'ng, C. *et al.* BPG: Seamless, automated and interactive visualization of scientific data. *BMC Bioinformatics* **20**, 42 (2019).
149. Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**, 488 (2008).
150. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).
151. UniProt. <https://www.uniprot.org/>.
152. Protein GRAVY. https://www.bioinformatics.org/sms2/protein_gravy.html.
153. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 161 (2018).
154. Dou, M. *et al.* High-Throughput Single Cell Proteomics Enabled by Multiplex Isobaric Labelling in a Nanodroplet Sample Preparation Platform. *Anal. Chem.* **91**, 13119–13127

- (2019).
155. Weng, S. S. H. *et al.* Sensitive Determination of Proteolytic Proteoforms in Limited Microscale Proteome Samples*. *Mol. Cell. Proteomics* **18**, 2335–2347 (2019).
 156. Kasuga, K., Kato, Y., Nagase, K. & Igarashi, K. Microproteomics with microfluidic-based cell sorting: Application to 1000 and 100 immune cells. *Proteomics* **17**, (2017).
 157. Sun, J. *et al.* Pathway analysis and transcriptomics improve protein identification by shotgun proteomics from samples comprising small number of cells - a benchmarking study. *BMC Genomics* **15**, S1 (2014).
 158. Tanca, A. *et al.* Critical comparison of sample preparation strategies for shotgun proteomic analysis of formalin-fixed, paraffin-embedded samples: insights from liver tissue. *Clin. Proteomics* **11**, 28 (2014).
 159. Föll, M. C. *et al.* Reproducible proteomics sample preparation for single FFPE tissue slices using acid-labile surfactant and direct trypsinization. *Clin. Proteomics* **15**, 11 (2018).
 160. Broeckx, V. *et al.* Comparison of multiple protein extraction buffers for GeLC-MS/MS proteomic analysis of liver and colon formalin-fixed, paraffin-embedded tissues. *Mol. Biosyst.* **12**, 553–565 (2016).
 161. Tabb, D. L. *et al.* Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography–Tandem Mass Spectrometry. *J. Proteome Res.* **9**, 761–776 (2010).
 162. Delmotte, N. *et al.* Repeatability of peptide identifications in shotgun proteome analysis employing off-line two-dimensional chromatographic separations and ion-trap MS. *J. Sep. Sci.* **32**, 1156–1164 (2009).
 163. Resing, K. A. *et al.* Improving Reproducibility and Sensitivity in Identifying Human Proteins by Shotgun Proteomics. *Anal. Chem.* **76**, 3556–3568 (2004).
 164. Washburn, M. P., Ulaszek, R. R. & Yates, J. R. Reproducibility of Quantitative Proteomic Analyses of Complex Biological Mixtures by Multidimensional Protein

- Identification Technology. *Anal. Chem.* **75**, 5054–5061 (2003).
165. Magdeldin, S. & Yamamoto, T. Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues. *PROTEOMICS* **12**, 1045–1058 (2012).
166. Maes, E. *et al.* Analysis of the formalin-fixed paraffin-embedded tissue proteome: pitfalls, challenges, and future perspectives. *Amino Acids* **45**, 205–218 (2013).
167. Prongvitaya, S. *et al.* High expression of CCDC25 in cholangiocarcinoma tissue samples. *Oncol. Lett.* **14**, 2566–2572 (2017).
168. Brychtova, V., Vojtesek, B. & Hrstka, R. Anterior gradient 2: a novel player in tumor cell biology. *Cancer Lett.* **304**, 1–7 (2011).
169. Suwanmanee, G. *et al.* Upregulation of AGR2vH facilitates cholangiocarcinoma cell survival under endoplasmic reticulum stress via the activation of the unfolded protein response pathway. *Int. J. Mol. Med.* **45**, 669–677 (2020).
170. van Oostrum, M. *et al.* Surfaceome dynamics reveal proteostasis-independent reorganization of neuronal surface proteins during development and synaptic plasticity. *Nat. Commun.* **11**, 4990 (2020).
171. Hou, G. *et al.* Biomarker Discovery and Verification of Esophageal Squamous Cell Carcinoma Using Integration of SWATH/MRM. *J. Proteome Res.* **14**, 3793–3803 (2015).
172. Liu, Y. *et al.* Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. *PROTEOMICS* **13**, 1247–1256 (2013).
173. Ludwig, C. *et al.* Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **14**, e8126 (2018).
174. Zhang, F., Ge, W., Ruan, G., Cai, X. & Guo, T. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *PROTEOMICS* **20**, 1900276 (2020).
175. Collins, B. C. *et al.* Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat. Commun.* **8**, 291 (2017).
176. Bouchal, P. *et al.* Breast Cancer Classification Based on Proteotypes Obtained by SWATH Mass Spectrometry. *Cell Rep.* **28**, 832-843.e7 (2019).

177. Steffen, P. *et al.* Molecular Features of Lymph Node Metastasis in T1/2 Colorectal Cancer from Formalin-Fixed Paraffin-Embedded Archival Specimens. *J. Proteome Res.* **20**, 1304–1312 (2021).
178. Gao, Y. *et al.* Quantitative proteomics by SWATH-MS reveals sophisticated metabolic reprogramming in hepatocellular carcinoma tissues. *Sci. Rep.* **7**, 45913 (2017).
179. Marchione, D. M. *et al.* HYPERsol: High-Quality Data from Archival FFPE Tissue for Clinical Proteomics. *J. Proteome Res.* **19**, 973–983 (2020).
180. Pirog, A. *et al.* Comparison of different digestion methods for proteomic analysis of isolated cells and FFPE tissue samples. *Talanta* **233**, 122568 (2021).
181. Newton, Y. *et al.* Large scale, robust, and accurate whole transcriptome profiling from clinical formalin-fixed paraffin-embedded samples. *Sci. Rep.* **10**, 17597 (2020).
182. Zhu, Y. *et al.* High-throughput proteomic analysis of FFPE tissue samples facilitates tumor stratification. *Mol. Oncol.* **13**, 2305–2328 (2019).
183. Shi, S. R., Key, M. E. & Kalra, K. L. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. *J. Histochem. Cytochem.* **39**, 741–748 (1991).
184. Yamashita, S. Heat-induced antigen retrieval: Mechanisms and application to histochemistry. *Prog. Histochem. Cytochem.* **41**, 141–200 (2007).
185. Wiśniewski, J. R., Ostasiewicz, P. & Mann, M. High Recovery FASP Applied to the Proteomic Analysis of Microdissected Formalin Fixed Paraffin Embedded Cancer Tissues Retrieves Known Colon Cancer Markers. *J. Proteome Res.* **10**, 3040–3049 (2011).
186. Nazarian, J., Santi, M., Hathout, Y. & MacDonald, T. J. Protein profiling of formalin fixed paraffin embedded tissue: Identification of potential biomarkers for pediatric brainstem glioma. *PROTEOMICS – Clin. Appl.* **2**, 915–924 (2008).
187. Luebker, S. A. & Koepsell, S. A. Optimization of Urea Based Protein Extraction from Formalin-Fixed Paraffin-Embedded Tissue for Shotgun Proteomics. *Int. J. Proteomics*

- 2016**, e4324987 (2016).
188. Alkhas, A. *et al.* Standardization of a Sample Preparation and Analytical Workflow for Proteomics of Archival Endometrial Cancer Tissue. *J. Proteome Res.* **10**, 5264–5271 (2011).
189. Ostasiewicz, P., Zielinska, D. F., Mann, M. & Wiśniewski, J. R. Proteome, Phosphoproteome, and N-Glycoproteome Are Quantitatively Preserved in Formalin-Fixed Paraffin-Embedded Tissue and Analyzable by High-Resolution Mass Spectrometry. *J. Proteome Res.* **9**, 3688–3700 (2010).
190. Hughes, C. S. *et al.* Quantitative Profiling of Single Formalin Fixed Tumour Sections: proteomics for translational research. *Sci. Rep.* **6**, 34949 (2016).
191. Bergmann, N. *et al.* The Intratumoral Heterogeneity Reflects the Intertumoral Subtypes of Glioblastoma Multiforme: A Regional Immunohistochemistry Analysis. *Front. Oncol.* **10**, (2020).
192. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
193. Weke, K. *et al.* MicroPOTS Analysis of Barrett's Esophageal Cell Line Models Identifies Proteomic Changes after Physiologic and Radiation Stress. *J. Proteome Res.* **20**, 2195–2205 (2021).
194. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**, 513–520 (2017).
195. Dapic, I., Uwugiaren, N., Jansen, P. J. & Corthals, G. L. Fast and Simple Protocols for Mass Spectrometry-Based Proteomics of Small Fresh Frozen Uterine Tissue Sections. *Anal. Chem.* **89**, 10769–10775 (2017).
196. Tang, J., He, D., Yang, P., He, J. & Zhang, Y. Genome-wide expression profiling of glioblastoma using a large combined cohort. *Sci. Rep.* **8**, 15104 (2018).
197. Liu, M. *et al.* The Identification of Key Genes and Pathways in Glioma by

- Bioinformatics Analysis. *J. Immunol. Res.* **2017**, e1278081 (2017).
198. Wang, W.-Q., Jensen, O. N., Møller, I. M., Hebelstrup, K. H. & Rogowska-Wrzesinska, A. Evaluation of sample preparation methods for mass spectrometry-based proteomic analysis of barley leaves. *Plant Methods* **14**, 72 (2018).
199. Hu, A., Noble, W. S. & Wolf-Yadlin, A. Technical advances in proteomics: new developments in data-independent acquisition. *F1000Research* **5**, F1000 Faculty Rev-419 (2016).
200. Gillet, L. C. *et al.* Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis *. *Mol. Cell. Proteomics* **11**, (2012).
201. Wei, P. *et al.* Serum GFAP autoantibody as an ELISA-detectable glioma marker. *Tumor Biol.* **34**, 2283–2292 (2013).
202. Zhuang, Z. *et al.* Proteomic identification of glutamine synthetase as a differential marker for oligodendrogliomas and astrocytomas. *J. Neurosurg.* **115**, 789–795 (2011).
203. Herpers, M. J. H. M. & Budka, H. Glial fibrillary acidic protein (GFAP) in oligodendroglial tumors: Gliofibrillary oligodendroglioma and transitional oligoastrocytoma as subtypes of oligodendroglioma. *Acta Neuropathol. (Berl.)* **64**, 265–272 (1984).
204. Thirant, C. *et al.* Proteomic analysis of oligodendrogliomas expressing a mutant isocitrate dehydrogenase-1. *PROTEOMICS* **11**, 4139–4154 (2011).
205. Anagnostopoulos, A. K. *et al.* Proteomics Studies of Childhood Pilocytic Astrocytoma. *J. Proteome Res.* **10**, 2555–2565 (2011).
206. Lemée, J.-M. *et al.* Integration of transcriptome and proteome profiles in glioblastoma: looking for the missing link. *BMC Mol. Biol.* **19**, 13 (2018).
207. Bathke, J., Konzer, A., Remes, B., McIntosh, M. & Klug, G. Comparative analyses of the variation of the transcriptome and proteome of *Rhodobacter sphaeroides* throughout growth. *BMC Genomics* **20**, 358 (2019).
208. Giusti, L., Angeloni, C. & Lucacchini, A. Update on proteomic studies of

- formalin-fixed paraffin-embedded tissues. *Expert Rev. Proteomics* **16**, 513–520 (2019).
209. Dapic, I. *et al.* Proteome analysis of tissues by mass spectrometry. *Mass Spectrom. Rev.* **38**, 403–441 (2019).
210. Gustafsson, O. J. R., Arentz, G. & Hoffmann, P. Proteomic developments in the analysis of formalin-fixed tissue. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1854**, 559–580 (2015).
211. Kawashima, Y. *et al.* Optimization of Data-Independent Acquisition Mass Spectrometry for Deep and Highly Sensitive Proteomic Analysis. *Int. J. Mol. Sci.* **20**, 5932 (2019).
212. Oh, S. *et al.* Integrated pharmaco-proteogenomics defines two subgroups in isocitrate dehydrogenase wild-type glioblastoma with prognostic and therapeutic opportunities. *Nat. Commun.* **11**, 3288 (2020).
213. La Rocca, G. *et al.* Glioblastoma CUSA Fluid Protein Profiling: A Comparative Investigation of the Core and Peripheral Tumor Zones. *Cancers* **13**, 30 (2021).
214. Ye, S. *et al.* Quantitative proteomics analysis of glioblastoma cell lines after lncRNA HULC silencing. *Sci. Rep.* **11**, 12587 (2021).
215. Lam, K. H. B. *et al.* Topographic mapping of the glioblastoma proteome reveals a triple-axis model of intra-tumoral heterogeneity. *Nat. Commun.* **13**, 116 (2022).
216. Herpers, M. J. H. M., Ramaekers, F. C. S., Aldeweireldt, J., Moesker, O. & Slooff, J. Co-expression of glial fibrillary acidic protein- and vimentin-type intermediate filaments in human astrocytomas. *Acta Neuropathol. (Berl.)* **70**, 333–339 (1986).
217. Zhao, J. *et al.* High Expression of Vimentin is Associated With Progression and a Poor Outcome in Glioblastoma. *Appl. Immunohistochem. Mol. Morphol.* **26**, 337–344 (2018).
218. Tang, H. *et al.* High brain acid soluble protein 1(BASP1) is a poor prognostic factor for cervical cancer and promotes tumor growth. *Cancer Cell Int.* **17**, 97 (2017).
219. Zhou, Q. *et al.* Quantitative proteomics identifies brain acid soluble protein 1 (BASP1) as a prognostic biomarker candidate in pancreatic cancer tissue. *eBioMedicine* **43**,

- 282–294 (2019).
220. Akiyama, S. K., Olden, K. & Yamada, K. M. Fibronectin and integrins in invasion and metastasis. *Cancer Metastasis Rev.* **14**, 173–189 (1995).
221. Ohnishi, T. *et al.* Role of fibronectin-stimulated tumor cell migration in glioma invasion in vivo: clinical significance of fibronectin and fibronectin receptor expressed in human glioma tissues. *Clin. Exp. Metastasis* **16**, 729–741 (1998).
222. Alban, T. J. *et al.* Glioblastoma Myeloid-Derived Suppressor Cell Subsets Express Differential Macrophage Migration Inhibitory Factor Receptor Profiles That Can Be Targeted to Reduce Immune Suppression. *Front. Immunol.* **11**, 1191 (2020).
223. Ha, W. *et al.* Ibudilast sensitizes glioblastoma to temozolomide by targeting Macrophage Migration Inhibitory Factor (MIF). *Sci. Rep.* **9**, 2905 (2019).
224. Yalamarty, S. S. K. *et al.* Mechanisms of Resistance and Current Treatment Options for Glioblastoma Multiforme (GBM). *Cancers* **15**, 2116 (2023).
225. Nakagawa, T. *et al.* Identification of glioblastoma-specific antigens expressed in patient-derived tumor cells as candidate targets for chimeric antigen receptor T cell therapy. *Neuro-Oncol. Adv.* **5**, vdac177 (2022).
226. Miyaguchi, K. *et al.* Activated T cell therapy targeting glioblastoma cancer stem cells. *Sci. Rep.* **13**, 196 (2023).
227. Brown, N. F., Carter, T. J., Ottaviani, D. & Mulholland, P. Harnessing the immune system in glioblastoma. *Br. J. Cancer* **119**, 1171–1181 (2018).
228. Rajani, K. R. *et al.* Harnessing Radiation Biology to Augment Immunotherapy for Glioblastoma. *Front. Oncol.* **8**, 656 (2018).
229. Colwell, N. *et al.* Hypoxia in the glioblastoma microenvironment: shaping the phenotype of cancer stem-like cells. *Neuro-Oncol.* **19**, 887–896 (2017).
230. Nicolas, S., Abdellatif, S., Haddad, M. A., Fakhoury, I. & El-Sibai, M. Hypoxia and EGF Stimulation Regulate VEGF Expression in Human Glioblastoma Multiforme (GBM) Cells by Differential Regulation of the PI3K/Rho-GTPase and MAPK Pathways. *Cells* **8**, 1397 (2019).

231. Wang, Y. *et al.* Hypoxia and macrophages promote glioblastoma invasion by the CCL4-CCR5 axis. *Oncol. Rep.* **36**, 3522–3528 (2016).
232. Zhang, Y. Y., Ruan, L. X. & Zhang, S. Rapid progression of glioblastoma multiforme: A case report. *Oncol. Lett.* **12**, 4803–4806 (2016).
233. Park, J. H. & Lee, H. K. Current Understanding of Hypoxia in Glioblastoma Multiforme and Its Response to Immunotherapy. *Cancers* **14**, 1176 (2022).
234. Krzywinska, E. & Stockmann, C. Hypoxia, Metabolism and Immune Cell Function. *Biomedicines* **6**, (2018).
235. Vito, A., El-Sayes, N. & Mossman, K. Hypoxia-Driven Immune Escape in the Tumor Microenvironment. *Cells* **9**, 992 (2020).
236. Brown, S. D. & Holt, R. A. Neoantigen characteristics in the context of the complete predicted MHC class I self-immunopeptidome. *OncolImmunology* **8**, 1556080 (2019).
237. Bedran, G. *et al.* The Immunopeptidome from a Genomic Perspective: Establishing the Noncanonical Landscape of MHC Class I–Associated Peptides. *Cancer Immunol. Res.* OF1–OF16 (2023) doi:10.1158/2326-6066.CIR-22-0621.
238. Stopfer, L. E., D’Souza, A. D. & White, F. M. 1,2,3, MHC: a review of mass-spectrometry-based immunopeptidomics methods for relative and absolute quantification of pMHCs. *Immuno-Oncol. Technol.* **11**, (2021).
239. Kajiwara, T. *et al.* Hypoxia augments MHC class I antigen presentation via facilitation of ERO1- α -mediated oxidative folding in murine tumor cells. *Eur. J. Immunol.* **46**, 2842–2851 (2016).
240. Sethumadhavan, S. *et al.* Hypoxia and hypoxia-inducible factor (HIF) downregulate antigen-presenting MHC class I molecules limiting tumor cell recognition by T cells. *PLOS ONE* **12**, e0187314 (2017).
241. Mohan, N. *et al.* Effects of hypoxia on antigen presentation and T cell-based immune recognition of HPV16-transformed cells. *Front. Immunol.* **13**, (2022).
242. Shraibman, B. *et al.* Identification of Tumor Antigens Among the HLA Peptidomes of Glioblastoma Tumors and Plasma *^[S]. *Mol. Cell. Proteomics* **18**, 1255–1268 (2019).

243. Joseph, J. V. *et al.* Hypoxia enhances migration and invasion in glioblastoma by promoting a mesenchymal shift mediated by the HIF1 α –ZEB1 axis. *Cancer Lett.* **359**, 107–116 (2015).
244. Abou Khouzam, R. *et al.* The Effect of Hypoxia and Hypoxia-Associated Pathways in the Regulation of Antitumor Response: Friends or Foes? *Front. Immunol.* **13**, (2022).
245. Chédeville, A. L., Lourdusamy, A., Monteiro, A. R., Hill, R. & Madureira, P. A. Investigating Glioblastoma Response to Hypoxia. *Biomedicines* **8**, 310 (2020).
246. Zhu, Y. *et al.* DEqMS: A Method for Accurate Variance Estimation in Differential Protein Expression Analysis*. *Mol. Cell. Proteomics* **19**, 1047–1057 (2020).
247. Indira Chandran, V. *et al.* Global extracellular vesicle proteomic signature defines U87-MG glioma cell hypoxic status with potential implications for non-invasive diagnostics. *J. Neurooncol.* **144**, 477–488 (2019).
248. Kore, R. A. *et al.* Hypoxia-derived exosomes induce putative altered pathways in biosynthesis and ion regulatory channels in glioblastoma cells. *Biochem. Biophys. Rep.* **14**, 104–113 (2018).
249. Leone, P. *et al.* MHC Class I Antigen Processing and Presenting Machinery: Organization, Function, and Defects in Tumor Cells. *JNCI J. Natl. Cancer Inst.* **105**, 1172–1187 (2013).
250. Hoof, I., Baarle, D. van, Hildebrand, W. H. & Keşmir, C. Proteome Sampling by the HLA Class I Antigen Processing Pathway. *PLOS Comput. Biol.* **8**, e1002517 (2012).
251. Seliger, B., Kloor, M. & Ferrone, S. HLA class II antigen-processing pathway in tumors: Molecular defects and clinical relevance. *OncolImmunology* **6**, e1171447 (2017).
252. Tanco, S. *et al.* Characterization of the Substrate Specificity of Human Carboxypeptidase A4 and Implications for a Role in Extracellular Peptide Processing*. *J. Biol. Chem.* **285**, 18385–18396 (2010).
253. Wei, C. *et al.* Comprehensive Analysis of CPA4 as a Poor Prognostic Biomarker Correlated with Immune Cells Infiltration in Bladder Cancer. *Biology* **10**, 1143 (2021).
254. Handa, T. *et al.* Carboxypeptidase A4 accumulation is associated with an aggressive

- phenotype and poor prognosis in triple-negative breast cancer. *Int. J. Oncol.* **54**, 833–844 (2019).
255. Fu, Y. *et al.* Downregulation of CPA4 inhibits non small–cell lung cancer growth by suppressing the AKT/c-MYC pathway. *Mol. Carcinog.* **58**, 2026–2039 (2019).
256. Moon, Y. *et al.* HIF-1 α -Dependent Induction of Carboxypeptidase A4 and Carboxypeptidase E in Hypoxic Human Adipose-Derived Stem Cells. *Mol. Cells* **43**, 945–952 (2020).
257. Papadopoulos, J. S. & Agarwala, R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**, 1073–1079 (2007).
258. Way, L. *et al.* Rearrangement of mitochondrial pyruvate dehydrogenase subunit dihydrolipoamide dehydrogenase protein–protein interactions by the MDM2 ligand nutlin-3. *Proteomics* **16**, 2327–2344 (2016).
259. Faktor, J. *et al.* The effects of p53 gene inactivation on mutant proteome expression in a human melanoma cell model. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1864**, 129722 (2020).

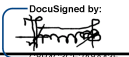
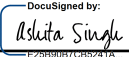
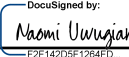
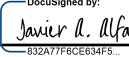
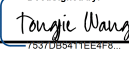
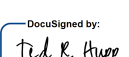
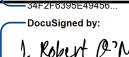
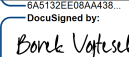
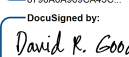
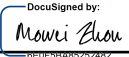

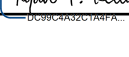
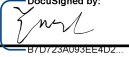
November 13th, 2023.

Dear Council members,

RE: Authorship Contribution Statement

We, the undersigned contributing authors, hereby declare our active involvement in the creation of the manuscript "**Weke *et al.*, MicroPOTS Analysis of Barrett's Esophageal Cell Line Models Identifies Proteomic Changes after Physiologic and Radiation Stress. *J. Proteome Res.* 20, 2195–2205 (2021).**" which is being included as **Chapter 2** in Kenneth Weke's Ph.D. thesis. We affirm that his contribution to this work was notably extensive, as he was primarily responsible for conducting the research, performing data analysis and interpretation, and drafting the manuscript.

Below is a summary of the authors' contributions:

| Name | Contribution | Signature |
|-------------------|---|---|
| Kenneth Weke | Data processing, analysis, visualization & interpretation, writing – original draft, writing – review & editing |  |
| Ashita Singh | Performing the experiment, writing – review & editing |  |
| Naomi Uwugiaren | Performing the experiment, writing – review & editing |  |
| Javier A. Alfaro | Writing – review & editing |  |
| Tongjie Wang | Writing – review & editing |  |
| Ted R. Hupp | Conceptualization, design of methodology, supervision, writing – review & editing |  |
| J. Robert O'Neill | Writing – review & editing |  |
| Borek Vojtesek | Writing – review & editing |  |
| David R. Goodlett | Writing – review & editing |  |
| Sarah M. Williams | Writing – review & editing | |
| Mowei Zhou | Writing – review & editing |  |
| Ryan T. Kelly | Conceptualization, design of methodology, writing – review & editing |  |
| Ying Zhu | Conceptualization, design of methodology, writing – review & editing |  |
| Irena Dapic | Conceptualization, design of methodology, supervision, project administration, writing – review & editing |  |

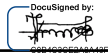



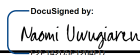
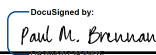
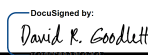

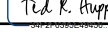
November 13th, 2023.

Dear Council members,

RE: Authorship Contribution Statement

We, the undersigned contributing authors, hereby declare our active involvement in the creation of the manuscript "**Weke *et al.*, DIA-MS proteome analysis of formalin-fixed paraffin-embedded glioblastoma tissues. *Anal. Chim. Acta* 1204, 339695 (2022).**" which is being included as **Chapter 3** in Kenneth Weke's Ph.D. thesis. We affirm that his contribution to this work was notably extensive, as he was primarily responsible for conducting the research, performing data analysis and interpretation, and drafting the manuscript.

Below is a summary of the authors' contributions:

| Name | Contribution | Signature |
|-------------------|--|---|
| Kenneth Weke | Performing the experiment, data generation, data processing, data analysis, visualization & interpretation, writing – original draft, writing – review & editing |  |
| Sachin Kote | Data generation, data processing, data analysis, visualization & interpretation, writing – review & editing |  |
| Jakub Faktor | Data generation & processing, writing – review & editing |  |
| Sofian Al Shboul | Data analysis & interpretation, writing – review & editing |  |
| Naomi Uwugiaren | Performing the experiment, writing – review & editing |  |
| Paul M. Brennan | Data processing, writing – review & editing |  |
| David R. Goodlett | Writing – review & editing |  |
| Ted R. Hupp | Conceptualization, design of methodology, supervision, writing – review & editing |  |
| Irena Dapic | Conceptualization, design of methodology, supervision, project administration, writing – review & editing |  |


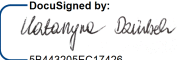


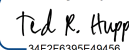
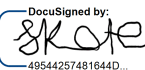
November 13th, 2023.

Dear Council members,

RE: Authorship Contribution Statement

We, the undersigned contributing authors, hereby declare our active involvement in the creation of the manuscript "**Weke et al., The interplay of hypoxia and antigen presentation in glioblastoma tumor: Insights from proteomics and immunopeptidomics. (Unpublished manuscript).**" which is being included as **Chapter 4** in Kenneth Weke's Ph.D. thesis. We affirm that his contribution to this work was notably extensive, as he was primarily responsible for conducting the research, performing data analysis and interpretation, and drafting the manuscript.

Below is a summary of the authors' contributions:

| Name | Contribution | Signature |
|-------------------|--|---|
| Kenneth Weke | Conceptualization, design of methodology, performing the experiment, data generation, data processing, data analysis, visualization & interpretation, writing – original draft, writing – review & editing |  DocuSigned by: Kenneth Weke 68B423CF2A8A13F... |
| Katarzyna Dziubek | Performing the experiment, data processing, data analysis, visualization & interpretation, writing – review & editing |  DocuSigned by: Katarzyna Dziubek 5B443205EC17426... |
| Ines Papak | Cytometry data generation, writing – review & editing |  DocuSigned by: Ines Papak 5E1C50A08CB8A4FA... |
| Martyna Muszczek | Writing – review & editing |  DocuSigned by: Muszczek C340B016333940F... |
| Ted R. Hupp | Conceptualization, design of methodology, supervision, writing – review & editing |  DocuSigned by: Ted R. Hupp 34F2F6395E49456... |
| Sachin Kote | Conceptualization, design of methodology, supervision, project administration, writing – review & editing |  DocuSigned by: SKote 49544257481644D... |