STANDING IN THE WAY OF PROFILING DEPTH: HOW THE SELECTION OF CHROMATOGRAPHY AND MASS SPECTROMETRY DATA ACQUISITION LIMIT OUR VIEW OF THE PROTEOME

by

Daniel Graham Delafield

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: 3/24/2023

The dissertation is approved by the following members of the Final Oral Committee:
Lingjun Li, Professor, Department of Chemistry and Division of Pharmaceutical Sciences
Joshua J. Coon, Professor, Department of Chemistry
Lloyd M. Smith, Professor, Department of Chemistry
William A. Ricke, Professor, Department of Urology

© Copyright by Daniel. G. Delafield

All Rights Reserved

Acknowledgements

The completion and successful defense of a doctoral dissertation is a monumental achievement. Standing at the end of my time in graduate school, it is only now fully apparent the extent to which individuals in my life have both directly and inadvertently contributed to my success. I do not know enough words that could suitably thank you for all you have given me – for the years of nurturing, the moments of guidance, the numerous opportunities to correct and teach me, or for your unwavering encouragement. Regardless of the thanks written here or repaid in kind, please know that the conclusion of this degree and the man I have become are not a testament of my abilities, but a reflection the role you have played in my life.

Though I may wish to believe that my enthusiasm for science and education are innate characteristics, my early educators deserve unequivocal praise for inspiring my pursuit and molding me into an adaptable, curious, and ambitious student. To Mrs. Jill Sullins, my eighth-grade science teacher: thank you for displaying consistent excitement for science education and empowering my curiosity. I still remember the day you introduced us to chemical formulas and taught us to balance ionic equations. Whether due my impressionable disposition or my sincere affinity for the subject, I credit you with inspiring my interest in chemistry and giving me the confidence to pursue greater heights. To Mr. Mike Lowery, my high school band director: thank you for being a role model that demonstrated diligence, persistence, and excellence. The trials, tribulations, pinnacles, and successes your education provided me taught the value of hard work and admonished me to always rise to the occasion. It is rare to have a daily reminder of personal responsibility and to be inspired towards collective achievement – your guidance has been invaluable. To Miss Torie York, my two-time high school social studies teacher: thank you for teaching me to have character and to have confidence in my understanding. I vividly recall and

take to heart the opportunities you presented me to consider others in the way I act, to be mindful, and to be generous. You displayed the unique ability to connect with any student and to inspire participation, regardless of reluctance – qualities not found in the average educator. I am so thankful for the opportunity to be mentored by someone of outstanding quality. Finally, to Mrs. Cheri Blackwood, my first formal chemistry educator and three-time science teacher: thank you for inspiring me to pursue science. Your unquestionable chemistry comprehension and genuine excitement for teaching both captured my attention and cemented my love of learning. When I moved on to college and found myself lost while searching for the right course of study, I was reminded of the immense joy your class brought to my life and it gave me the confidence to select chemistry as my major. Had I liked the subject less or not been shown that I can learn and succeed, my life today would look very different, and this doctoral degree would not be within reach. Thank you for the life path you helped me create.

The completion of two graduate degrees would not be possible without the guidance and mentorship of my advisors. To Dr. Si Wu, thank you for taking the chance on an undergraduate student who just needed to complete his honors research. I had no idea what proteomics was and had very little exposure to separation science and mass spectrometry but the exposure to these topics in your lab could not have come at a better time. Until joining your group, I felt that maybe science wasn't a career I wished to pursue. But learning how my education and skills were applicable to real-world problems and being taught the importance of analytical curiosity changed everything. Your guidance and mentorship not only provided me skills and understanding, it secured my appreciation for this field of study and laid the groundwork for my future career. To Dr. Lingjun Li, my Ph.D. advisor: thank you for allowing me the freedom to grow. I came into your lab ready to work and having decided on the direction I wish to take. It would be natural for

any advisor to correct my course or adjust my interests to fit a set of criteria, but from the beginning you were encouraging of any direction I wish to take and provided me the resources to do so. When I look back at my time at the University of Wisconsin, I will remember how pivotal these years have been and I cannot thank you enough for encouraging me to ask questions and start new projects without the expectation or promise of return. I am so appreciative for the chance to be shaped into a mature, skilled scientist.

I also wish to thank my friends for providing me support and a means of social outlet during the arduous years in Madison. Nhu Vu, Ashley Phetsanthad, Austin Carr, Dylan Tabang, Chris Sauer, and many others, thank you for providing me fun, laughter, and helpful criticism throughout our time together. I am so thankful to have become close to such talented scientists and I look forward to our future interactions as part of the same research community.

Finally, the largest acknowledgements are dedicated to my family. It goes without stating that your unswerving patience and unconditional love have been paramount to my success. To Dewey and Patricia Coffman, my grandparents: thank you for nurturing my imagination throughout my life. You have always wished for me to grow, explore, learn, and achieve and I am so thankful you have always been there to offer support. Not everyone has the privilege of knowing and growing up with a loving grandparent; I have the honor of knowing two. To my sister Aubrey: thank you for pioneering the world ahead of me. Your social and adventurous nature caused you to seek activities and gain experiences that brought you joy and long-lasting memories. While I wish to claim I am just as adventurous, I know that without your example I would never have pursued the interests, craved the opportunities, or made the friends that have so deeply enriched my life. Thank you for constantly demonstrating how to pursue happiness. To my brother Cooper: thank you for providing an excellent example – in some ways a foil – against which I may compare

myself. I did not receive any formal lessons into how to be a good older brother, but I can see that is an example you never needed. You have grown to be an intelligent, kind, understanding man with an agreeable temperament and good heart. I am so proud of you and I am very thankful to be a part of your life. To my father, Fred Delafield: thank you for never wanting anything less than the best for me. You have cared for me, taught me, and guided me during my formative years and through it all you never sought return or placed expectations before me. Your deepest desire has always been for me to find my own dream and to create a life that is only mine rather than to take after you or to pursue a dream you have for me. Without you, I would not be the man I am today and I hope I provide to my future family everything that you have given me. And finally, to my mother, Mary Delafield: thank you for always loving me. Though I may have been a difficult child at times, and possibly an equally difficult adult, I am thankful that you have always been my biggest supporter and source of encouragement. Throughout my life you have sought to lift up my achievements to spur me on and you have provided me comfort in times of uncertainty. I still maintain that my aptitude for speaking, engagement, and participation come from the invaluable example you set. Thank you for every teachable moment that has brought me into adulthood and prepared me for my life ahead.

Table of Contents

List of Tables

List of Figures

List of Supplemental Figures

List of Abbreviations

ACN	Acetonitrile
ACTN1	Actin Alpha 1
AGC	Automatic Gain Control
AKT	Protein Kinase B
ANLN	Annilin
APO	Apolipoprotein
APP	Amyloid Precursor Protein
AURKB	Aurora Kinase B
BCA	Bicinchonic Acid
BCAM25	Basal Cell Adhesion Molecule
BCAT1	Branched Chain Amino Acid Transferase
BCaP	BPH-1-derived Cancer Progression
BPH	Benign Prostate Hypertrophy
BUB1	Mitotic Checkpoint Serine/Threonine Kinase
CALD126	Caldesmon
CAV1	Caveolin-1
CCS	Collisional Cross Section
CDC20	Cell Division Cycle 20
CEP78	Centrosomal protein
CID	Collisional Induced Dissociation
CIMS	Cyclic Ion Mobility Mass Spectrometry
CIU	Collisional Induced Unfolding

CLDN1	Caludin-1
COL17A1	Collagen 17
CRPC	Castration-Resistant Prostate Cancer
CSF	Cerebrospinal Fluid
CV	Coefficient of Variance
DDA	Data-Dependent Acquisition
DDX	DEAD-box Helicase
DHT	Hormone Dihydrotestosterone
DI-SPA	Direct Infusion Shotgun Proteomics Analysis
DIA	Data-Independent Acquisition
DIA-MS	Data-Independent Acquisition Mass Spectrometry
DIA-NN	Data-Independent Acquisition Neural Network
DIMS	Differential Ion Mobility Mass Spectrometry
DPS	DNA-binding protein from starved cells
DRE	Digital Rectal Examination
DTIMS	Drift Tube Ion Mobility Mass Spectrometry
DTT	Dithiothreitol
ECD	Electron Capture Dissociation
ECM	Extracellular Matrix
EGFR	epidermal growth factor receptor
ELISA	Enzyme-Linked Immunosorbent Assay
EMT	Epithelial-Mesenchymal Transition
ERLIC	Electrostatic Repulsion-Hydrophilic Interaction Chromatography

ESI	Electrospray Ionization
ETD	Electron Transfer Dissociation
FA	Formic Acid
FAH	fumarylacetoacetate hydrolase
FAIMS	High Field Asymmetric Waveform Mass Spectrometry
FDR	False Discovery Rate
FSTL118	Follistatin-Related Protein 1
FUT8	Fucosyltransferase 8
FWHM	Full Width Half Max
G6PD	Glucose-6-Phosphate Dehydrogenase
GCLC	Glutamate-Cysteine Ligase Catalytic Subunit
GCLM	Glutamate-Cysteine Ligase Modifier Subunit
GO	Gene Ontology
GOT	Glutamic Oxaloacetate Transaminase
GPX	Glutathione Peroxidase
GRAVY	Grand Average of Hydropathy
GSH	Glutathione
GSN	Gelsolin
GSS	Glutathione Synthetase
GSTM4	Glutathione S-Transferase Mu
HCC	Hepatocellular Carcinoma
HCD	Higher Energy Collisional Induced Dissociation
HGSC	High-Grade Serious Carcinoma

HIF	hypoxia induced factor
HIF1a	Hypoxia Inducible Factor 1 Subunit Alpha
HILIC	Hydrophilic Interaction Chromatography
HLA	Human Leukocyte Antigen
HYOU1	Hypoxia Upregulated 1
IAA	Iodoacetamide
IFI35	Interferon Induced Protein
IFIT	Interferon Induced Proteins with Tetratricopeptide Repeats
IFN	Interferon
IGFBP	Insulin-like Growth Factor Binding Protein
IM-IMS	Ion Mobility Mass Spectrometry
INA	Internexin
INCENP	Inner Centromere Protein
ISG	Interferon-Stimulated Gene
ISRE	Interferon-Stimulated Response Element
iTRAQ	Isotopic Tags for Relative and Absolute Quantitation
IVL	Involucrin
JAK	Janus kinase
KIF	Kinesin Superfamily Protein
KLK4	Kallikrein Related Peptidase 4
KNL1	Kinetochore Null Protein
KNSTRN	Kinetochore Localized Astrin
KRT1	Keratin

LC-MS	Liquid Chromatography Mass Spectrometry
LFQ	Label-Free Quantitation
LNCaP	Lymph Node Carcinoma of the Prostate
LOY	Loss of Y
MALDI	Matrix-Assisted Laser Desorption/Ionization
MAPT	Microtubule Associated Protein Tau
MBR	Match Between Runs
MCI	Mild Cognitive Impairment
mdSUGAR	Millidalton Isobaric Multiplex Labeling Reagents for Carbonyl-Containing Compound
MMP1	Matrix Metalloproteinase
MPRIP	Myosin Phosphotase Rho
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
mTOR	Mammalian Target of Rapimycin
MYH	Myosin Heavy Chain 14
MYL	Myosin Light Polypeptide
MYO	Mysoin
NDC80	Kinetochore Protein NDC80 Homolog
NF-KB	Nuclear Factor Kappa B
NUF2	Kinetochore Protein Nuf2
OAS	Oligoadenylate synthase

	Protein Aggregate Magnesium-Ammonium Phospholinoleate-Palmitoleate
P-MAPA	Anhydride
PANC1	Pancreatic Cancer-1
PASEF	Parallel Accumulation Serial Fragmentation
PBS	Phosphorylated Buffered Saline
PDK1	3-Phosphoinositide-Dependent Kinase 1
PFKM	ATP-Dependent 6-Phosphofructokinase, Muscle
PGC	Porous Graphitic Carbon
PGM1	Phosphoglucomutase 1
PI3K	Phosphoinositide 3-kinase
PKM2	Pyruvate Kinase 2
PKN1	Protein Kinase N1
PPL	Periplakin
PPP2R5D	Protein Phosphatase 2 Regulatory Subunit B'delta
PRC1	Protein Regulator of Cytokinesis
PRM	Parallel Reaction Monitoring
PSA	Prostate Specific Antigen
PSAT1	Phosphoserine Aminotransferase 1
PSC	Pluripotent Stem Cell
PSM	Peptide Spectral Match
PSMB	Proteasome 20S Subunit Beta
PSPH	Phosphoserine Phosphatase
PTM	Post-Translational Modification

PYCR	Pyrroline-5-carboxylate reductaseÂ
Q-TOF	Quadrupole-Time of Flight
RF1	Release Factor 1
ROC	Receiver Operating Characteristic
ROS	Reactive Oxygen Species
RPLC	Reversed-Phase Liquid Chromatography
RPLC-MS	Reversed-Phase Liquid Chromatography-Mass Spectrometry
RPTOR	RPTOR regulatory associated protein of MTOR complex 1
rUGM	Urogenital mesenchymal
SCE	Stepped Higher Energy Collisionally Induced Dissociation
SDS	Sodium Dodecyl Sulfate
SID	Surface Induced Dissociation
SILAC	Stable Isotope Labeling by Amino Acids in Cell Culter
SLC	Solute Carrier
SMS	Spermine Synthase
SPC	Kinetochore Protein SPC
SPC24	Kinetochore Protein Spc24
SPDL1	Spindly-Like Protein 1
SPEG	Striated Muscle Preferentially Expressed Protein Kinase
SRM	Select Reaction Monitoring
STAT	Signal Transducer and Activator of Transcription
STOM	Stomatin
SUGAR	Isobaric Multiplex Labeling Reagents for Carbonyl-Containing Compound

SWATH-MS	Sequential Window Acquisition of All Theoretical Ions
TCA	Tricarboxylic Acid
TFA	Trifluoroacetic Acid
TFEB	Transcription Factor EB
TIMS	Trapped Ion Mobility Mass Spectrometry
TMT	Tandem Mass Tag
TOSIL	Tandem ^{18}O Stable Isotope Labeling
TPM	tropomyosin
TPP1	Tripeptidyl-Peptidase 1
TRAMP	Transgenic Adenocarcinoma of the Mouse Prostate
TWIMS	Travelin Wave Ion Mobility Mass Spectrometry
UBE2	Ubiquitin Conjugating Enzyme E2
XIC	Extracted Ion Chromatogram

Abstract

This thesis outlines research focusing on the development and application of chromatography separation and mass spectrometry data acquisition methods aimed at improving profiling depth in proteomic analyses. This work is briefly outlined and explained in **Chapter 1**, highlighting how reliance on common analytical methodologies inherently limits our view of the proteome. **Chapter 2** details the initial reports of utilizing porous graphitic carbon (PGC) chromatography as an alternative to reversed-phase liquid chromatography (RPLC). This report describes demonstrable improvements in peptide, glycopeptide, and protein identifications while highlighting the various analytical advantages seen over traditional separations. In this preliminary work, we observed elevated column temperatures imparted potentially detrimental effects of various glycopeptides, a phenomenon explored in **Chapter 3**. This following report provides heuristic guidance in PGC-based glycoproteomic analyses and details how column temperature must be optimized to suit the given analytical need. **Chapter 4** provides a topical overview of quantitative approaches in glycan and glycopeptide analyses, providing meaningful consideration and comparison of methodologies most appropriate for use in future glycoproteomic analyses. In **Chapter 5**, we return to PGC separations with a specific focus on further detailing the benefits found in proteomic analyses. This report highlights substantial improvements in the number of peptide and protein identifications compared to what was seen previously and highlights the breadth of information lost during routine RPLC analyses. **Chapter 6** departs from chromatographic separation and turns attention towards quantitative mass spectrometry (MS) methods useful for biological discovery. The knowledge and information garnered from this report informed the experimental design used in **Chapter 7**, which describes the employment of data-independent acquisition (DIA)-MS to study a novel, progressive prostate cancer cell model. Here

we quantified 6,614 proteins across 9 biological samples, finding 1,242 to be significantly dysregulated in malignant cancer phenotypes. Thes proteins demonstrate potential for disease diagnosis, phenotypic stratification, and therapeutic targeting. In **Chapter 8** we expand on the utility of DIA-MS, detailing the ability to reuse and repurpose prior proteomic measurements for enhanced biomolecular identification. We apply this workflow to the analysis of neurological disorder patient cohorts, revealing 1,642 dysregulated proteins that speak to the biomolecular organization related to Alzheimer's Disease. **Chapter 9** describes new and emerging ion mobility (IM)-based analytical modalities and discuss their capacity for biomolecular interrogation and structural analysis. Given instrumentation of this kind is not utilized in the works preceding, this report explains potential advantages and necessary considerations, should this technology become of interest. Finally, we conclude with **Chapter 10**, briefly discussing the various investigations that may follow this body of work.

Chapter 1: Introduction and Summary

Investigations of health and human disease have long relied on adept analytical methodologies for biomolecular analysis. Decades of concerted focus and advanced training have bolstered the importance of understanding the molecular drivers of normal and aberrant biological processes and have progressively increased the need for more advanced technology and methods providing greater information breadth. A sector of contemporary biomolecular analysis is the study of proteins, the participants executing cellular functions necessary to sustain life. The study of protein expression, modification, and structure is ubiquitously referred to as 'proteomics.' Among the litany of analytical techniques available, mass spectrometry (MS) is unquestionably the method-of-choice as it facilitates the breadth of measurements needed for holistic proteomic analyses and provides the only high-throughput technique for peptide and protein identification[1]. Expectedly, the increasing appreciation for and reliance on MS-based analyses have brought about the development of nuanced variations that provide researchers with both greater experimental flexibility as well as the need to evaluate the efficiency of a chosen approach.

Broadly, MS-based proteomics may be divided into four distinct areas: sample preparation, chromatographic separation, data acquisition, and data processing. Technological advances have provided extreme diversity in approach for each of these four areas, with each researcher likely utilizing a different combination to facilitate proteomic measurements. Despite being heterogenous in approach, a high-level view of proteomics identifies some general consistencies shared across most applications. Argued subjectively, chromatographic separation is the area most conserved across time and research settings. Reversed-phase liquid chromatography (RPLC) is the standard mode of chromatographic separation with unquestionable efficiency and broad utility[2-4]. This separation, which takes advantages of differences in analyte hydrophobicity allows for facile separation of peptide and protein mixtures, making it suitable for most analyses. However,

literature has routinely suggested the limitations of reversed-phase chromatography due to the inability to effectively retain hydrophilic analytes. As well, analytes of hydrophilic character that are retained display very poor partitioning between the solid and aqueous phase, resulting in poor liquid-phase resolution and inhibiting detection and limiting utility in downstream analysis. Further complicating this point, hydrophilic analytes are often characterized by low ionization efficiency both due to the lower propensity to carry charge and being outcompeted for ionization by longer, often more basic hydrophobic peptides. These realities are further exacerbated when considering potential peptide modifications such as glycosylation and phosphorylation that impart additional hydrophobicity and increase the inherent complexity within complex mixtures[5, 6]. Considered holistically, the ubiquitous use of RPLC inherently limits the proteomic information revealed in routine analyses, significantly biasing our understanding of the proteome towards those biomolecules most compatible with this separation.

There are several alternative chromatography paradigms that may be used in place of RPLC. Ion-exchange, size exclusion, and affinity chromatography have all been extensively employed in proteomics analyses, utilizing the innate or inspired character of peptides to distinguish complex mixtures. However, these separations are relatively agnostic to peptide hydrophobicity and only serve to contrast RPLC separations if specific physiochemical properties are present. Hydrophilic interaction chromatography (HILIC) is the most common separation regime utilized to target hydrophilic analytes and has been routinely employed for analysis of peptides and small molecules alike[7-9]. A recent addition to the chromatographic toolbox is electrostatic repulsion-hydrophilic interaction chromatography (ERLIC)[10-13], which combines benefits of ion-exchange and HILIC to improve retention of polar analytes[14-17]. However, both HILIC and ERLIC present limitations that inhibit their utility in proteomic analyses. HILIC utilizes

gradients that move from a high concentration of organic buffer to a high concentration of aqueous buffer. As such, peptide mixtures must be resuspended in an organic solution prior to analysis, which can prove challenging for some hydrophilic analytes. Additionally, both HILIC and ERLIC require the use of salt-containing buffers to mitigate charge effects, which can hinder analyte detection via MS. Taken together, there is an immediate need to develop or validate easily implemented chromatography paradigms that display broad biomolecular compatibility while addressing the limitations presented in RPLC analyses.

Porous graphitic carbon (PGC) has gained significant attention in recent years for its ability to retain polar analytes, with specific attention drawn to metabolites and free glycans[18-31].. Whereas RPLC separations utilize a silica resin support to which long carbon chains may be affixed, PGC is constructed from planar graphene sheets, which promote analyte retention based on polarity and electrostatic interactions. This separation paradigm is compatible with the typical aqueous-organic gradients used in RPLC analysis while still exhibiting solvent-flexibility for unique applications, indicating it may be directly implemented in a majority of analytical settings. In addition, recent reports have demonstrated PGC separations present the unique capacity to separate biomolecular isomers at higher temperatures. This phenomenon is principally highlighted in the analysis of glycans, where at high temperatures the oligosaccharide moiety adopts an elongated conformation, allowing greater separation of biomolecules with unique connectivity. Importantly, a single report detailed the ability of PGC to distinguish glycopeptide isomers through targeted analysis of glycoprotein standards. This report enabled us to hypothesize that the polar and electrostatic retention of PGC separation is not completely incompatible with peptide backbones and may therefore be evaluated as a potential compliment to RPLC separations and may play display even further benefit in untargeted glycopeptide analysis.

In Chapter 2, we assess this hypothesis by fabricating custom nano-flow capillary columns to directly compare PGC and RPLC separations. An additional biphasic PGC-RPLC column was also fabricated to evaluate any combinatorial advantages. Using these columns to analyze complex cell line digests, and enriched glycopeptide mixtures, we demonstrate a significant complement in proteomic and glycoproteomic coverage. Though RPLC still exceeds the other separations in number of peptide and proteins identified, our analyses resulted in a 23% improvement in peptide identifications and a 14% improvement in protein identification when column temperatures are optimized. We were encouraged by these initial findings as they represent some of the first reports detailing the utility of PGC in proteomic analyses. However, we felt this shotgun approach may inherently limit the true potential of PGC separations and later sought to overcome these obstacles (see Chapter 5). In addition to these complementary peptide and protein identifications, we demonstrated that PGC analysis of glycopeptide mixtures results in greater retention of small, hydrophilic glycopeptides dominated in character by the attached glycan. As well, PGC-specific glycopeptides were found to have greater charge density than those found in RPLC, directly addressing the challenges of ionization inefficiency. We also observe greater recovery of site microheterogeneity and reveal facile isomeric separation at elevated temperatures. Taken together, this report details the potentially advantageous nature of incorporating PGC separations into routine analyses. Though sample coverage is complementary and not orthogonal, we present substantial improvements in the breadth and depth of proteomic information retrieved how this paradigm may be further developed for unique structural analyses.

However, these analyses did present some topical limitations. We observed significant disparities in the quantity and type of glycopeptides identified at higher temperatures – information that was not available until this report. Observing that high mannose and sialylated glycans were

the classes of glycans most directly affected – and noting the contrasting impact – we hypothesized this was a universal trend that would sufficiently impact identification and quantitation of glycopeptides. In Chapter 3 we investigate this hypothesis by performing follow-up analyses on enriched glycopeptide mixtures. Using a similar separation approach as that seen in Chapter 2, we reveal that increases in column temperature to 45°C largely benefit the identification of all glycopeptides while further increases to 60°C impart differential effects on various glycopeptide classes. At highest temperatures, mannose-type glycopeptides (i.e., high mannose and paucimannose glycopeptides) were not impacted, exhibiting greater liquid-phase resolution and higher reporting signal. Sialylated glycopeptides – and to a lesser extent complex and fucosylated glycopeptides – displayed greatest resolution but lowest reporting signal and identification rates at 60°C. Though we were not able to fully elucidate the cause of this observation, we speculate it is due progressive deterioration of glycopeptides caused by the acidic conditions used in LC-MS analyses and the introduction of additional energy in the form of heat. In the same report, we use spectral counting for quantitation and reveal that even though identifications are lowest at high temperatures, quantitative accuracy generally improves. Taken together, this report expands on the utility of PGC for glycopeptide analysis, providing heuristic guidance for how experimental conditions can, and should, be tailored to fit analytical need.

As preceding chapters have introduced the need for enhanced glycopeptide identification and also the relevance of PGC analyses in perform quantitative investigations, Chapter 4 provides a timely description of the most recent efforts in glycan and glycopeptide quantitation. Whether utilizing metabolic incorporation, isotopic labeling, isobaric labeling, or label free approaches, numerous detailed reports exist to provide guidance as to the best means of quantitation for a given application. These approaches may be easily incorporated into PGC-based glycopeptide analyses

to not only provide quantitation across samples, but also to provide quantitation of unique glycopeptide isomers distinguished at high temperatures. Ultimately, we consider the combination of high-throughput quantitation and PGC analyses an area of emerging interest and suggests a meaningful future direction (see Chapter 10).

As introduced previously, Chapter 5 further expands the role of PGC in proteomic analyses. Though we briefly revealed the advantages of PGC separations for peptide and protein identification in Chapter 2, we neither sought to explore the relevance of increased protein identifications nor contextualize the importance complementary peptide identifications. Within this report, we utilize a prefractionation approach to further compare the performance of PGC and RPLC separations. Using this strategy, we elevate the complementary performance of PGC, increasing peptide and protein identifications by 43% and 24%, respectively. More than this, we detail that although the quantity of peptides identified was lower than those in RPLC, PGC separations identified significantly more proteins overall, suggesting there is significant redundancy in RPLC analyses. We further explain that the additional proteins identified do not significantly differ in abundance than those identified in RPLC analyses. This finding is significant as it acutely displays the quantity and breadth of proteomic information lost during routine proteomic analyses. We go on to validate the importance of including additional peptide identifications in protein abundance estimations, revealing that protein quantitation ascertained through RPLC analyses may not be reflective of physiological concentration. Overall, this report further cements the use of PGC analyses in an array of MS-based proteomics investigations.

With the importance of chromatographic separation development squarely in place, we can begin to turn our attention towards the ways in which proteomic analyses may be used for biomolecular discovery and investigation of human disease. Chapter 6 introduces quantitative

proteomics in the effort to elucidate biomolecular changes seen in various cancers. Through understanding the various avenues for protein quantitation and their applicability towards biological discovery, we were able to ascertain appropriate experimental design for our investigation of prostate cancer (PCa) in Chapter 7.

Prostate cancer is the second leading form of cancer in men and the cancer with the highest age-adjusted rate of incidence overall. While the importance of illuminating strategies for disease diagnosis goes without stating, the rapidly aging demographics in population-dense and high-cancer-rate countries further impress this need and mandate immediate attention. Prostate cancer presents numerous limitations in diagnosis and treatment as common urological screenings, digital rectal exams (DREs), only serve to suggest the presence of prostate cancer. Ultimately PCa is diagnosed through biopsy and histological assessment, as well as through biofluid screening to determine prostate specific antigen (PSA) levels – an inefficient biomarker. This biomarker is repeatedly used to monitor PCa progression and treatment, indicating there are severe shortcomings in our current reliance on a single protein. For this reason, we hypothesize we can leverage MS-based proteomics to uncover proteins or protein communities that bear significance in PCa progression and may prove more useful in clinical diagnosis, PCa stratification, and therapeutic monitoring.

Previous works introduced here have exclusively utilized data-dependent acquisition. This type of data acquisition is based on a priority queue, whereby ionized analytes detected by mass spectrometry are ranked according to their abundance and then sent for fragmentation. This acquisition results in clean, easily processed data but is inherently biased towards those highly abundant analytes. Data-independent acquisitions provides a remedy to this limitation by isolating all ions in a given mass window and fragmenting them simultaneously. In this way, we can achieve

greater sample coverage, more robust quantitation, and higher identification rates at cost of more complex data analysis. Based on these considerations, we elected to utilize data-independent acquisition (DIA) in the discovery of protein fingerprints related to prostate cancer.

In Chapter 7, we introduce the idea of library-free DIA-MS analysis and utilize this technique to profile a novel, progressive prostate cancer cell model. Using label-free quantitation, we illuminate the biomolecular profiles observed as PCa advances from non-tumorigenic to aggressive tumorigenic to aggressive metastatic tumorigenic. Quantifying 6,614 proteins with 1,242 found to be dysregulated in malignant cell lines, we demonstrate the unique proteomic fingerprints of these cell lines easily distinguish cancer phenotypes from one another. Furthermore, we identify 7 distinct protein clusters that bear similarity in direction and magnitude of dysregulation, allowing us to easily identify protein communities related to PCa progression or those that serve as markers of malignancy. Within these analyses, we highlight the importance of various biological pathways such as mitosis, interferon signaling, and small molecule processing within PCa.

The work in Chapter 7 utilizes a library free approach, assigning peptide precursors according to their theoretical fragment ions and expected retention time. However, library free analyses only extend so far, being unavailable for use in post-translational modification (PTM analysis) and often containing theoretical information not analogous to that which observed. We posit the in-depth proteomic analyses, which account for PTMs and provide empirical measurements may be re-used by other researchers to advance their DIA-MS analyses. Within Chapter 8, we demonstrate that large, comprehensive spectral libraries containing data not collected by a given user can be calibrated to new experimental conditions and used for peptide and protein identification. We utilize an open-source machine learning architecture to train a model

that learns to predict retention time from our empirical data. This model is then introduced to unseen peptide sequences – those found in the MassIVE knowledgebase, which are then given an expected retention time, allowing us to identify peptides and proteins that would otherwise go unobserved. Using this approach, we quantify >9,300 protein groups across neurological disease cohorts, revealing 1,642 to be significantly dysregulated against healthy control. This information not only illuminates the proteomic reorganization specific to Alzheimer's Disease, but also highlights the ability to accurately reuse empirical proteomic measurements for biological discovery.

Finally, as proteomics encompasses protein identification, quantitation, and structural characterization, Chapter 9 provides a review of emerging MS technologies most suitable to a given approach. Ion mobility (IM)-MS is a powerful technique that may be leveraged for increased sample throughput or to provide structural information. We review the four main IM modalities – drift tube, traveling wave, trapped ion, and high field asymmetric waveform ion mobility mass spectrometry – for their capacity within various omics applications and their ability to provide structural elucidation. As seen through our in-depth comparison, IM technology not only presents a promising analytical technique across numerous applications, they also must be carefully selected and appropriately matched to a given application. In all, this review should serve both as preliminary information to new IM researchers, as well as providing logical guidance to future investigators seeking to utilize this technique for coming analyses.

References

1. Gillet, L. C.; Leitner, A.; Aebersold, R., Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annual Review of Analytical Chemistry* **2016,** *9* (1), 449-472.

2. Mao, Z.; Hu, C.; Li, Z.; Chen, Z., A reversed-phase/hydrophilic bifunctional interaction mixed-mode monolithic column with biphenyl and quaternary ammonium stationary phases for capillary electrochromatography. *Analyst* **2019,** *144* (14), 4386-4394.

3. Liang, Y.; Zhang, L.; Zhang, Y., Monolithic Materials-Based RPLC-MS for Proteoform Separation and Identification. In *Proteoform Identification: Methods and Protocols*, Sun, L.; Liu, X., Eds. Springer US: New York, NY, 2022; pp 43-53.

4. Rozing, G., Micropillar array columns for advancing nanoflow HPLC. *Microchemical Journal* **2021,** *170*, 106629.

5. Badgett, M. J.; Boyes, B.; Orlando, R., Predicting the Retention Behavior of Specific O-Linked Glycopeptides. *J Biomol Tech* **2017,** *28* (3), 122-126.

6. Zhu, R.; Zacharias, L.; Wooding, K. M.; Peng, W.; Mechref, Y., Chapter Twenty-One - Glycoprotein Enrichment Analytical Techniques: Advantages and Disadvantages. In *Methods in Enzymology*, Shukla, A. K., Ed. Academic Press: 2017; Vol. 585, pp 397-429.

7. Boersema, P. J.; Mohammed, S.; Heck, A. J. R., Hydrophilic interaction liquid chromatography (HILIC) in proteomics. *Analytical and Bioanalytical Chemistry* **2008,** *391* (1), 151-159.

8. Sun, Z.; Ji, F.; Jiang, Z.; Li, L., Improving deep proteome and PTMome coverage using tandem HILIC-HPRP peptide fractionation strategy. *Analytical and Bioanalytical Chemistry* **2019,** *411* (2), 459-469.

9. Badgett, M. J.; Boyes, B.; Orlando, R., Peptide retention prediction using hydrophilic interaction liquid chromatography coupled to mass spectrometry. *Journal of Chromatography A* **2018,** *1537*, 58-65.

10. Yeh, T.-T.; Ho, M.-Y.; Chen, W.-Y.; Hsu, Y.-C.; Ku, W.-C.; Tseng, H.-W.; Chen, S.-T.; Chen, S.-F., Comparison of different fractionation strategies for in-depth phosphoproteomics by liquid chromatography tandem mass spectrometry. *Analytical and Bioanalytical Chemistry* **2019,** *411* (15), 3417-3424.

11. Cui, Y.; Tabang, D. N.; Zhang, Z.; Ma, M.; Alpert, A. J.; Li, L., Counterion Optimization Dramatically Improves Selectivity for Phosphopeptides and Glycopeptides in Electrostatic Repulsion-Hydrophilic Interaction Chromatography. *Analytical Chemistry* **2021,** *93* (22), 7908-7916.

12. Cui, Y.; Yang, K.; Tabang, D. N.; Huang, J.; Tang, W.; Li, L., Finding the Sweet Spot in ERLIC Mobile Phase for Simultaneous Enrichment of N-Glyco and Phosphopeptides. *Journal of the American Society for Mass Spectrometry* **2019,** *30* (12), 2491-2501.

13. Alpert, A. J., Electrostatic Repulsion Hydrophilic Interaction Chromatography for Isocratic Separation of Charged Solutes and Selective Isolation of Phosphopeptides. *Analytical Chemistry* **2008,** *80* (1), 62-76.

14. Zhen, J.; Kim, J.; Zhou, Y.; Gaidamauskas, E.; Subramanian, S.; Feng, P., Antibody characterization using novel ERLIC-MS/MS-based peptide mapping. *mAbs* **2018,** *10* (7), 951-959.

15. Yan, J.; Ding, J.; Jin, G.; Duan, Z.; Yang, F.; Li, D.; Zhou, H.; Li, M.; Guo, Z.; Chai, W.; Liang, X., Profiling of Human Milk Oligosaccharides for Lewis Epitopes and Secretor Status by Electrostatic Repulsion Hydrophilic Interaction Chromatography Coupled with Negative-Ion Electrospray Tandem Mass Spectrometry. *Analytical Chemistry* **2019,** *91* (13), 8199-8206.

16. Qing, G.; Yan, J.; He, X.; Li, X.; Liang, X., Recent advances in hydrophilic interaction liquid interaction chromatography materials for glycopeptide enrichment and glycan separation. *TrAC Trends in Analytical Chemistry* **2020,** *124*, 115570.

17. Bermudez, A.; Pitteri, S. J., Enrichment of Intact Glycopeptides Using Strong Anion Exchange and Electrostatic Repulsion Hydrophilic Interaction Chromatography. In *Mass Spectrometry of Glycoproteins: Methods and Protocols*, Delobel, A., Ed. Springer US: New York, NY, 2021; pp 107-120.

18. West, C.; Elfakir, C.; Lafosse, M., Porous graphitic carbon: A versatile stationary phase for liquid chromatography. *Journal of Chromatography A* **2010,** *1217* (19), 3201-3216.

19. Bapiro, T. E.; Richards, F. M.; Jodrell, D. I., Understanding the Complexity of Porous Graphitic Carbon (PGC) Chromatography: Modulation of Mobile-Stationary Phase Interactions Overcomes Loss of Retention and Reduces Variability. *Analytical Chemistry* **2016,** *88* (12), 6190-6194.

20. Stavenhagen, K.; Hinneburg, H.; Kolarich, D.; Wuhrer, M., Site-Specific N- and O-Glycopeptide Analysis Using an Integrated C18-PGC-LC-ESI-QTOF-MS/MS Approach. In *High-Throughput Glycomics and Glycoproteomics: Methods and Protocols*, Lauc, G.; Wuhrer, M., Eds. Springer New York: New York, NY, 2017; pp 109-119.

21. Xu, J.; Liu, X.; Zhou, H., Recent advances in separation methods for post-translational modification proteomics. *Sepu* **2016,** *34* (12), 1199-1205.

22. Zhou, S.; Huang, Y.; Dong, X.; Peng, W.; Veillon, L.; Kitagawa, D. A. S.; Aquino, A. J. A.; Mechref, Y., Isomeric Separation of Permethylated Glycans by Porous Graphitic

Carbon (PGC)-LC-MS/MS at High Temperatures. *Analytical Chemistry* **2017,** *89* (12), 6590-6597.

23. Zhou, S.; Dong, X.; Veillon, L.; Huang, Y.; Mechref, Y., LC-MS/MS analysis of permethylated N-glycans facilitating isomeric characterization. *Analytical and Bioanalytical Chemistry* **2017,** *409* (2), 453-466.

24. Huang, Y.; Zhou, S.; Zhu, J.; Lubman David, M.; Mechref, Y., LC-MS/MS isomeric profiling of permethylated N-glycans derived from serum haptoglobin of hepatocellular carcinoma (HCC) and cirrhotic patients. *ELECTROPHORESIS* **2017,** *38* (17), 2160-2167.

25. Ashwood, C.; Lin, C.-H.; Thaysen-Andersen, M.; Packer, N. H., Discrimination of Isomers of Released N- and O-Glycans Using Diagnostic Product Ions in Negative Ion PGC-LC-ESI-MS/MS. *Journal of the American Society for Mass Spectrometry* **2018,** *29* (6), 1194-1209.

26. Hinneburg, H.; Chatterjee, S.; Schirmeister, F.; Nguyen-Khuong, T.; Packer, N. H.; Rapp, E.; Thaysen-Andersen, M., Post-Column Make-Up Flow (PCMF) Enhances the Performance of Capillary-Flow PGC-LC-MS/MS-Based Glycomics. *Analytical Chemistry* **2019,** *91* (7), 4559-4567.

27. Ashwood, C.; Pratt, B.; MacLean, B. X.; Gundry, R. L.; Packer, N. H., Standardization of PGC-LC-MS-based glycomics for sample specific glycotyping. *The Analyst* **2019,** *144* (11), 3601-3612.

28. Ashwood, C.; Waas, M.; Weerasekera, R.; Gundry, R. L., Reference glycan structure libraries of primary human cardiomyocytes and pluripotent stem cell-derived cardiomyocytes reveal cell-type and culture stage-specific glycan phenotypes. *Journal of Molecular and Cellular Cardiology* **2020,** *139*, 33-46.

29. Wei, J.; Tang, Y.; Bai, Y.; Zaia, J.; Costello, C. E.; Hong, P.; Lin, C., Toward Automatic and Comprehensive Glycan Characterization by Online PGC-LC-EED MS/MS. *Analytical Chemistry* **2020,** *92* (1), 782-791.

30. Chen, C.-H.; Lin, Y.-P.; Ren, C.-T.; Shivatare, S. S.; Lin, N.-H.; Wu, C.-Y.; Chen, C.-H.; Lin, J.-L., Enhancement of fucosylated N-glycan isomer separation with an ultrahigh column temperature in porous graphitic carbon liquid chromatography-mass spectrometry. *Journal of Chromatography A* **2020,** *1632*, 461610.

31. Riley, N. M.; Bertozzi, C. R.; Pitteri, S. J., A Pragmatic Guide to Enrichment Strategies for Mass Spectrometry-based Glycoproteomics. *Molecular & Cellular Proteomics* **2021**, 100029.

Chapter 2: Complementary proteome and glycoproteome access revealed through comparative analysis of reversed phase and porous graphitic carbon chromatography.

Delafield, D. G., Miles, H. N., Liu, Y., Ricke, W. A., & Li, L. (2022) Complementary proteome and glycoproteome access revealed through comparative analysis of reversed phase and porous graphitic carbon chromatography. *Analytical and Bioanalytical Chemistry*. 414, 5461–5472.

Abstract

Continual development in instrumental and analytical techniques have aided in establishing rigorous connections between protein glycosylation and human illness. These illnesses, such as various forms of cancer, are often associated with poor prognoses, prompting the need for more comprehensive characterization of the glycoproteome. While innovative instrumental and computational strategies have largely benefited glycoproteomic analyses, less attention is given to benefits gained through alternative, optimized chromatographic techniques. Porous graphitic carbon (PGC) chromatography has gained considerable interest in glycomics research due to its mobile phase flexibility, increased retention of polar analytes and improved structural elucidation at higher temperatures. PGC has yet to be systematically compared against or in tandem with standard reversed phase liquid chromatography (RPLC) in high-throughput bottom-up glycoproteomics experiments, leaving the potential benefits unexplored. Performing comparative analysis of single and biphasic separation regimes at a range of column temperatures illustrates complementary advantages for each method. PGC separation is shown to selectively retain shorter, more hydrophilic glycopeptide species, imparting higher average charge, and exhibiting greater microheterogeneity coverage for identified glycosites. Additionally, we demonstrate that liquid-phase separation of glycopeptide isomers may be achieved through both single and biphasic PGC separations, providing a means towards facile, multidimensional glycopeptide characterization. Beyond this, we demonstrate how utilization of multiple separation regimes and column temperatures can aid in profiling the glycoproteome in tumorigenic and aggressive prostate cancer cells. RAW MS proteomics and glycoproteomics datasets have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD024196 (10.6019/PXD024196) and PXD024195, respectively.

Introduction

Glycoproteins, a unique class of post-translationally modified biomolecules, have long been an area of dedicated investigation within proteomic research communities. The high degree of modification complexity[1] has revealed the numerous roles glycoproteins play in physiological processes such as cell trafficking[2-4], signaling pathways[5-7] and protein folding[8-10], among others. Aided by continual improvements in mass spectrometry (MS)-based biomolecule discovery and identification, researchers have turned their attention towards understanding the relationship between these protein modifications and human disease. The dynamic nature of modification site occupancy and microheterogeneity[11] has established connections between glycoproteins and neurodegenerative diseases[12-17], autoimmune disorders[18-20] and numerous forms of cancer[21-23].

The growing relevance of the human glycoproteome has prompted analytical innovation, stimulating curiosity in strategies best suited for detecting, identifying, and characterizing glycoproteins. Specific to bottom-up glycoproteomic investigations, glycopeptide enrichment strategies and instrument capabilities are often the areas of most intense focus. Numerous novel enrichment strategies[24-26] have been developed to compensate for the low abundance of glycopeptides within complex proteolytic mixtures, while novel dissociation techniques[27-30] are pursued to provide more effective single-run sequencing of peptide backbones. Beyond this, ion separation regimes are often the primary focus when seeking to provide structural information of diverse glycan modifications[31-33].

Though these analytical developments have provided unique avenues towards more comprehensive glycoproteomic analysis, the experiments validating their efficacy almost

exclusively rely on traditional reversed-phase liquid chromatography (RPLC). As there has been no extensive utilization of alternative separation regimes, with the exception of hydrophilic interaction chromatography (HILIC), the potential benefits and complements offered through novel chromatography have been left uncharacterized. Porous graphitic carbon (PCG) has shown increased utility in recent years for glycan identification and characterization[34-37] experiments but has not been evaluated in high-throughput glycopeptide analyses. As Zhu *et al*.[38] recently demonstrated the utility of PGC in elucidating glycopeptide modification structure, we question to what extent PGC may be employed to deepen the coverage of the human glycoproteome and complement the robust strategies that currently exist.

Presented here is a systematic comparison of proteome and glycoproteome profiling enabled through traditional RPLC C18, PGC, and biphasic RPLC-PGC separation regimes. We build on current literature that highlights the differences in analyte retention at elevated column temperatures, illustrating the impact it has in complex glycopeptide profiling experiments. This study brings to light the complementary proteome coverage provided by PGC, selectively retaining those shorter, more hydrophilic analytes that go unretained in RPLC analyses. PGC is also shown to provide greater coverage for glycan microheterogeneity and enable higher charge states on the retained hydrophilic glycopeptides. Interestingly, we also demonstrate both PGC and biphasic separation regimes provide higher isomeric resolution as column temperature increases, indicating potential utility in characterization experiments. Finally, we examine two complex, human-derived prostate cancer cell lines, demonstrating the utility of complementary separation regimes and characterizing the benefits and drawbacks of altering column temperatures in complex glycopeptide profiling experiments.

Materials and methods

Reagents and Materials

Bovine fetuin (F3004), bovine ribonuclease B (R7884), and bovine a1-Acid Glycoprotein (G3643) standards, as well as 1,4-dithiothreitol (DTT, D9779), 2-iodoacetamide (IAA, I6125), and methanol (MeOH, 34860) were purchased from MilliporeSigma (Burlington, MA). Urea (U15), tris-base (BP152), hydrochloric acid (A144SI), chloroform (C297), and acetonitrile (ACN, A998) were purchased from Fisher Scientific (Waltham, MA). Trypsin was purchased from Promega (V5113, Madison, WI). Capillary tubing (1068150019) was purchased from PolyMicro. RPLC packing material (4451GP) was purchased from Osaka Soda Co. (Osaka, Japan). PGC packing material was obtained by extracting stationary phase from PGC guard columns (35003-014001) purchased from ThermoFisher Scientific (Waltham, MA). PolyHYDROXYETHYL A packing material was purchased from PolyLC (Columbia, MD). Materials for preparing frits were purchased from Next Advance (Troy, NY).

Pancreatic Cancer Cell (PANC1) Culture

The commercial pancreatic cancer cell lines PANC-1 obtained from American Type Culture Collection (ATCC, Manassas, VA, USA) were cultured and maintained in DMEM:F12 (Hyclone, GE Healthcare Life Sciences, Logan, Utah, USA) containing 10% fetal bovine serum (FBS) (Gibco, Origin: Mexico), 1% penicillin-streptomycin solution (Gibco, Life Technologies Corporation, Grand Island, NY, USA). Cells cultured in a 37°C moisture incubator supplied with 5% CO2. PANC-1 cultures between passage 3 and 15 were used for all experiments. Cells at 70%-90% confluence were trypsinized using 0.25% trypsin EDTA solution (Corning, Mediatech, Inc., Manassas, VA, USA). Cell suspension was centrifuged at 800g for 5 minutes and the medium was

discarded. Cells were resuspended in phosphate-buffered saline (PBS) (Corning, Mediatech, Inc., Manassas, VA, USA) and washed 3 times with PBS, flash frozen in dry ice, and stored at −80°C.

Prostate Cancer Cell (BCaP, LNCaP) Culture

The BPH1 to cancer progression (BCaP)[39] and LNCaP[40] cell lines were grown and maintained in phenol-free DMEM/Ham's F-12 (Gibco) supplemented with 5% fetal bovine serum (HyClone) and 1% penicillin-streptomycin solution (Gibco). T175 culture flasks were placed in an incubator at 5% CO_2 and 98% humidity. Cells were grown to 90% confluency, washed with 1X phosphate-buffered saline (Cytiva) and harvested using a cell scraper. Cell pellets were washed twice using phosphate-buffered saline and stored at -80°C for subsequent processing.

Protein Extraction and Digestion

Evaluating various preparation protocols, an adapted FASP approach was chosen for extraction and digestion of proteins from PANC1, LNCaP, BCaP-NT1, and BCaP-T10 cell lines. This strategy, which does not provide targeted membrane protein enrichment, proved to enable highest recovery of secretory glycoproteins and access to a small number of heavily glycosylated[41-43] membrane-bound species. Dedicated strategies are required to target membrane proteins and may be pursued in later analyses. Briefly, cells were resuspended in and washed with PBS, centrifuging at 14,000 rcf between washes. Cells were then resuspended in lysis buffer (8M Urea, protease inhibitor, 20mM HEPES) and lysed via ultrasonication. Cell debris was removed, and proteins were reduced and alkylated with centrifuge filtering between each step. Proteins were digested with trypsin (1:50) overnight at 37°C and desalted via reverse phase desalting cartridges. Volume was reduced under vacuum and concentration of peptides was estimated via Pierce Peptide Assay. Standard glycoproteins were reduced alkylated and digested similarly; a detailed preparation workflow is described in the electronic supplemental information.

Glycopeptide Enrichment

The enrichment procedure is fully detailed in the electronic supplemental information. Briefly, 200µL pipette tips were blocked with 3mg sterile cotton, and loaded with a hydrated polyHYDROXYETHYL-A resin. Weight of beads is adjusted according to the mass of peptides used. Peptide mixture was added and allowed to bind through repeated centrifugation. Non-specific peptides were washed away using 80% ACN+0.1% TFA with enriched glycopeptides eluted in 10% ACN+5% FA. Prior to analysis, glycopeptides were dried under vacuum and resuspended in LC/MS-grade water+0.1% FA.

LC-MS Parameters

Samples were analyzed on an Thermo Scientific QE-HF mass spectrometer coupled with a NanoUltimate 3000 chromatography stack. Follow-up analyses were performed on an Thermo Scientific Fusion Lumos with the same chromatography system. Custom capillary columns were fabricated and coupled as described (electronic supplemental information, Figure S2.1) and temperatures were controlled by a 30 cm pencil column heater (Phoenix S&T). Flow rate was established for each column by achieving a stable pressure drop of 450 bar at 30°C and was held constant as temperature was varied (C18: 475µL/min, PGC: 750µL/min, BP: 525µL/min). Using water+0.1%FA and ACN+0.1%FA as buffers A and B LC gradients were held consistent between columns: a trapping phase from 0-16 minutes at 3% B moving to 40% at minute 100, 75% B at from minutes 102.5-105, 97% B from minutes 105.1-113, and 3% B from minutes 115-125.

Mass spectrometry settings were configured to balance profiling depth with mass resolution. Using a full MS/dd-MS2 method, MS1 settings were as follows: polarity, positive; default charge state, 2; microscans, 1; resolution, 60K; AGC target, 1e6; maximum injection time, 250 ms; m/z range, 150-1500; spectrum type, centroid. MS2 settings: microscans, 1; resolution,

15K; AGC target, 2e5; maximum AGC 8e3; maximum injection time, 120 ms; loop count, 15; isolation window, 2 m/z, isolation offset, 0 m/z; fixed first mass, 100.0 m/z; collision energy, stepped HCD (25, 35, 45); spectrum type, centroid; dynamic exclusion, 30s; rejected charge states, 1+, 8+ >8+.

Data Analysis and Availability

Preliminary results from analysis of PANC1 cell lysate digest were identified in PEAKS X (Bioinformatics Solutions Inc.) searching against the UniProt reviewed human proteome. Glycopeptide data was annotated via Byonic (Protein Metrics). Standard glycopeptides were run against a custom database of UniProt protein sequences while LNCaP and BCaP samples were searched against the human proteome. The parameters used for PEAKS X and Byonic searching are described in the electronic supplemental information.

RAW MS data for proteomics and glycoproteomics datasets have been deposited to the ProteomeXchange Consortium via the PRIDE[44] partner repository with the dataset identifier PXD024196 (10.6019/PXD024196) and PXD024195, respectively. Peptide, glycopeptide, and protein identifications are also provided in supplemental tables S1-7. Byonic output files are available at https://figshare.com/account/home#/projects/98459. Custom data scraping, analysis, calculations, and visualization was done primarily in Python using the Altair[45] library with additional components done in R, D3, and vega-lite. All code is available at https://github.com/lingjunli-research/pgc-parallel-comparison.

Results and Discussion

While PGC is noted to be amenable to an array of solvent systems, a key benefit is its compatibility with typical water/ACN buffer systems. In hand with the principle of contrasting retention mechanisms compared to RPLC, this indicates a facile means of complementary

proteomic/glycoproteomic analysis and multidimensionality compared to organic-based separations such as HILIC. Though a multidimensional C18-PGC strategy was reported for pronase-digested glycopeptides [46], the profiling depths of these two separation regimes have not yet been compared for complex proteomic or glycoproteomic analyses.

We address this knowledge gap through comparative analyses of complex proteolytic and enriched glycopeptide mixtures across three custom-fabricated nanoflow separation regimes run at varying temperatures (Figure 2.1). Columns were prepared as described (electronic supplemental information) and coupled to a Q-Exactive-HF nano-electrospray source as shown (Figure S2.1). As a note, 3mm diameter packing materials are used for each stationary phase as, at the time of writing, this was the smallest available PGC packing material. Smaller diameter RPLC material may be employed to increase liquid-phase resolution and is under evaluation for future research.

Proteomic Performance

To evaluate the performance in shotgun-like proteomic investigations, proteins were extracted from pancreatic cancer cells (PANC1), digested with trypsin, and subjected to LC-MS/MS at varying temperatures (30°C, 45°C, 60°C). RPLC demonstrated the best overall performance in peptide and protein identifications, followed by PGC and biphasic (BP) separations (

Figure 2.2**,** Figure S2.2**,** Table S1). While temperature variations had negligible effect on BP performance, both C18 and PGC experiments displayed noticeable improvements in number of identified species when run at 45°C; this observation is consistent with later analyses and is discussed within. Across all temperatures, our results produced the expected overlap in highly

abundant identified species, though RPLC and PGC demonstrate the most distinct differences in number of unique identifications (

Figure 2.2). This observation supports the idea of complementary proteome access enabled through PGC separation, indicating more holistic coverage may be obtained when incorporated alongside RPLC. We further investigated the species found to be unique in RPLC or PGC separations to establish any preferential retention of subcellular species. Our data demonstrate that there are noticeable differences in the most prevalent subcellular locations and cellular compartments associated with proteins unique to each separation regime (Figure S2.1**,** Table S2). While this finding supports the gain of complimentary proteome access through respective separations, it must be considered carefully. Rather than definitively conclude that each separation regime targets specific protein subclasses, these data only present trends that may be further defined after in-depth discovery proteomics experiments.

Hypothesizing that differences in peptide character are the primary factor in driving this complementary proteome access, we first evaluated the distribution of retained peptide lengths. Considering the high number of peptides identified in each experiment (between 6,000 and 11,000) it is not surprising the median peptide lengths are largely conserved across all experiments (

Figure 2.2). However, at all temperatures, PGC demonstrates the lowest median length, a more concise interquartile range, and shorter maximums and outlier species. More importantly is the evaluation of hydrophilic character of retained peptides identified in each experiment. Using GRAVY (grand average of hydropathy) score to infer hydrophilic character, RPLC separations

retained those peptide species with the greatest overall hydrophobic character at all gradient timepoints, as evidenced by the higher, positive GRAVY scores (

Figure 2.2). Conversely, PGC separations yield peptides with overall hydrophilic character (lower, more negative GRAVY scores). Noting intermediate trend of biphasic separations compared to the two higher performing regimes, this knowledge, combined with the presence of unique peptide identifications in this separation strategy, indicates that further optimization of biphasic separations could provide synergistic retention and identification. The utility of biphasic regimes is further expanded in glycopeptide analyses, as seen below.

To establish the reliability of our custom separation strategies, it was imperative we evaluate inter-run reproducibility. Our follow-up replicate analyses of each separation regime (Table S3) established high reproducibility in peptide and protein identifications (Figure S2.4), as well as confirmed the trends seen in peptide length and hydropathy (Figure S2.5). In total, these results point toward two notable conclusions. First, PGC separations can be successfully employed in high-throughput analyses of tryptic proteolytic mixtures. As previous experiments have been confined to analyses of small peptides - often using sequential or non-specific digestion - this observation stands alone as the first of its kind and implies high utility in proteomic analyses. Second, our results demonstrate the complementary access to the human proteome that is granted when PGC separations are employed. Whereas PGC selectively retains those shorter, more hydrophilic peptides, as well as being amenable to standard LC mobile phases, this regime presents a useful approach to proteome profiling that may be readily implemented in research settings where RPLC is employed. As well, greater benefit may be found in multidimensional chromatography experiments and instances where fractionation is used to reduced sample complexity.

Trends in Glycopeptide Identification

To understand the performance of these complementary separation regimes in glycopeptide identification and characterization experiments, we performed HILIC enrichment on three glycopeptide-rich samples: one mixture of standard glycoprotein digests and two BPH1 to cancer progression (BCaP) cell line digests. The former sample was employed to provide a robust, reproducible dataset that may be validated in other research settings, while the latter two are meant to serve as real, complex glycoproteomic mixtures. Within our analyses, we sought to profile the characteristic differences of glycopeptides retained in RPLC, PGC, and biphasic separations while evaluating the extent to which incorporation of PGC separations would benefit glycoproteomic workflows, expand microheterogeneity profiling, and offer benefits in structural elucidation.

After compiling annotated glycopeptides identified in all samples, we again observed a greater number of total and unique glycopeptides in RPLC experiments with PGC as the next best performer (Figure 2.3), with these two regimes also displaying the greatest number of unique glycopeptides across all experiments (Figure 2.3). Comparing glycopeptide identifications when sample constituents and column temperature are held constant confirms the difference in retention mechanism is the primary cause of unique identifications (Figure 2.3**,** Figure S2.6). However, our data also confirms that that there are temperature-dependent differences in glycopeptide identification (Figure 2.3**,** Figure S2.7), an observation further validated in subsequent analyses (Figure S2.8**,** Tables S4-5). Knowing this, we then sought to understand what differences exist between glycopeptides retained on each stationary phase. As a note, we again observe a noticeable improvement in analyte identification when RPLC and PGC columns are operated at 45°C compared to other conditions, which is discussed below.

Peptide-Level Differences in Retained Glycopeptides

To illuminate the peptide-level differences existing between RPLC and PGC separations, we consolidated the dataset to contain only those glycopeptides that were identified in a single separation regime. Given the hydrophobicity-driven separation mechanism of RPLC, we had anticipated this regime would vastly outpace the others in number of unique glycopeptide backbones. In contrast to this hypothesis, holistic and sample-specific data (Figure 2.4, Figure S2.9**,** Table S6) revealed comparable numbers of unique species in RPLC and PGC with the former regime identifying only slightly more peptide backbones. Additionally, RPLC-specific glycopeptides were found to be significantly longer than those identified in PGC ($p=1.44E^{-32}$, Wilcoxin rank-sum) and notably different from those in BP experiments ($p=0.102$, Wilcoxin rank-sum) while simultaneously providing a greater distribution of charge states (Figure 2.4). These stark discrepancies in RPLC experiments result from a higher prevalence of basic amino acid residues, due in large part to missed cleavage sites.

These observations must be carefully considered in the context of glycoproteomics. Our data suggest that RPLC separations provide the highest identification rates of glycopeptides, agreeing with the general utility of this technique across current literature. However, RPLC separations very clearly benefit from instances where the hydrophobic character of the peptide outpaces the often-dominating hydrophilicity of the glycan; such is the case for long peptides and those with missed cleavage sites. These results highlight that the incorporation of PGC separations into existing experiments directly complements, rather than detracts from current glycoproteomics methodologies. As each regime provides meaningful access to unique components of the glycoproteome, these two regimes may be leveraged in tandem to increase glycoproteomic coverage.

Furthermore, the differences in charge state distribution across all experiments may be further contextualized for potential benefits and drawbacks. Within glycoproteomic analyses, longer peptides with higher charge states could be considered beneficial in some analytical applications as they may be leveraged for modification site localization through electron transfer dissociation (ETD) and directly combats traditional limitations in ionization efficiency of glycopeptides. However, this can also result in precursor reporting signal being distributed across several channels, lowering overall signal intensity, and potentially limiting precursor selection in data-dependent analyses. Interestingly, glycopeptides identified in PGC analyses display a consolidated charge state distribution, primarily displaying 2+ and 3+ ions at rates higher than or equal to those seen in RPLC analyses. Normalizing this charge to the length of the peptide backbone reveals that PGC-specific glycopeptides demonstrated higher charge density among all separation regimes (Figure 2.4**,** Figure S2.9). This increased charge density displayed in short, hydrophilic analytes is a direct consequence of not selectively retaining those hydrophobic species that traditionally outcompete during ionization. As such, we may conclude that for applications where charge-mediated fragmentation is desired, RPLC separations are more beneficial as they will retain longer, highly charged glycopeptides. At the same time, PGC separations may find utility in applications where researchers wish to remedy ionization inefficiency hydrophilic analytes – the need for which is exacerbated when considering glycan-level differences of retained glycopeptides.

Glycan-Level Differences

Turning to the differences exhibited at the glycan level, the benefits of PGC separations become more pronounced. Topically, PGC analyses retained the broadest array of unique glycan structures, only falling behind RPLC in the standard glycopeptide sample (Figure 2.5**,** Figure

S2.10). Interestingly, when comparing the extent of glycosylation (inferred from the number of monosaccharide residues within each glycan structure), initial observations revealed that PGC retained smaller glycans compared to other separation regimes, though this discrepancy is confined to the standard peptide sample and is likely due to microheterogeneity differences at retained glycosites.

Across all experiments, we did not observe notable bias towards any glycopeptide type, as all identifications were found to be exclusively high-mannose or complex. More notable however, is the relative size of the glycan in relation to the peptide backbone. Across all samples, PGC-specific peptides displayed the highest monosaccharide-to-amino-acid ratio, indicating the character of these glycopeptides is largely dominated by the large, hydrophilic glycan. Rearticulating the above-mentioned observation that PGC selectively retains glycopeptides with shorter, more hydrophilic backbones, this revelation further emphasizes the utility of PGC separations in accessing those portions of the glycoproteome that would traditionally go undetected in RPLC analysis.

As a primary goal in glycoprotein profiling investigations is to characterize site microheterogeneity, it is of topical concern to evaluate the number of glycans associated with each identified peptide backbone. Across all experiments, PGC regularly outpaced the other separations, providing a higher median number of glycans per glycosite, with higher maximums found in all but one sample (Figure 2.5). This observation further illustrates the value of PGC separation in glycoproteomic analyses as it not only accesses those portions of the glycoproteome not found in RPLC analyses, but it also provides improved microheterogeneity profiling depth. This reality may be further leveraged to provide unparalleled glycoproteome profiling, if used

alongside established methods of sequential or non-specific enzymatic digestion to provide a broad array of PGC-compatible glycopeptides.

Improved Isomeric Resolution

While analytical methods capable of assigning glycan location and composition occupy a large portion of glycoproteomics investigations, facile structural elucidation remains on the outset of widespread investigation. Previous studies have detailed the ability of PGC to provide separation of isomeric glycans at higher temperatures [34, 38, 47] postulating the expanded glycan morphology that results from increased temperature interacts more readily with the stationary phase, improving liquid-phase resolution. Though this PGC-based approach was recently demonstrated in collections of tryptic/chymotryptic glycopeptides [38], it has never been explored in complex samples that display high diversity in peptide sequence and glycan composition.

Similar to these previous results, our studies show that isomeric glycopeptides can be resolved through PGC separations with benefits often being further pronounced at higher temperatures (Figure 2.6**,** Figure S2.11-Figure S2.12). However, as previous studies utilized commercial columns of shorter length (L=10cm, *i.d.*=75µm), we predicted our longer stationary phase would impart greater resolution of constitutional isomers prior to MS. Contrary to this hypothesis, a subset of glycopeptides that were partially or fully resolved in PGC experiments also displayed comparable resolution in BP separations, with some more well-resolved in the latter regime (Figure S2.13). This realization, in tandem with the poorer demonstrated performance of BP in glycopeptide retention indicate a novel approach may be to couple higher resolution RPLC stationary phases (L=15cm, *d.p.*=1.7µm) with a 15cm PGC stationary component. This method may provide a more effective means of deep profiling and structural elucidation.

In addition, a survey of those species with altered liquid phase retention yields two putative trends. The less surprising observation is that isomeric resolution is greater for those containing high mannose and complex glycans with unequal antennae, especially those found to be singly sialylated. Second, as liquid-phase separation was more readily achieved in glycopeptides from our standard protein mixture and in species identified on all separation regimes, we hypothesize this observation is concentration-dependent. This is also supported in theory. Considering the unequal distribution between major and minor glycan conformations, it is logical to assume some minor species may not be preserved after extensive sample handling.

Taken together, our results serve to further emphasize the potential benefits of incorporating PGC separation strategies into routine glycoproteomics investigations. Whereas previous studies often employ more expensive commercial columns – many of which are no longer available for purchase – our study highlights the facility of custom fabrication and customization of analytical columns capable of discerning structural information. Comprehensive structural characterization is not a primary goal of many bottom-up glycoproteomics experiments, we present a low-cost, facile means of deepening glycoproteome coverage that may be further developed to provide substantial gains in biological information.

Glycoproteome Profiling

Though the results have thus far demonstrated the marked differences in proteome access provided through PGC and RPLC separations, we must assess any potential bias in glycopeptide retention presented in PGC analysis and evaluate the potential impact that higher temperature plays on glycoproteome profiling. Focusing on the two complex samples, BPH1 to cancer progression (BCaP) NT1 and T10 cell lines [48], it is immediately noticeable that high mannose glycans present

the highest proportion of glycan species identified (Figure 2.7). As this observation is shared between RPLC and PGC separations, and because analyses of our standard mixture deviate from this trend (Figure S2.14), we can hypothesize this is not a result of preferential retention of high mannose glycans and is instead related to higher prevalence of these species in the samples analyzed. Future quantitative investigations, performed across numerous cell lines, are needed to confirm this hypothesis.

Mammalian N-glycoproteins are known to be dominated by complex glycans, with high mannose modifications being considered "immature" within the synthetic pathway [1]. Recently, a growing number of studies have revealed high abundances of high mannose glycans in cancer cells [49, 50] with these moieties noted as being involved in cell-cell interactions [51, 52]. Of note, a recent study revealed that extended high mannose glycans directly contribute to bile duct cancer metastasis, noting the importance of specific cell surface glycosylation [53]. Importantly, this study notes the importance of terminal α-1,2 mannose residues, bolstering the importance of providing structural elucidation in glycoproteomic investigations. Based on our results that demonstrate resolution of isomeric glycopeptides in BP and PGC separations, we anticipate this type of analysis would provide a novel path toward in-depth, targeted glycoproteomics. As the importance given to glycan isomers continues to grow, and the ability to easily distinguish these species without specialized instrumentation becomes invaluable, the incorporation of a PGC separation component will play a vital role in future investigations.

With respect to the observed prevalence of high mannose glycan modifications, we also observe an obvious difference in the number of species identified at 45°C versus those at any other temperature. As this trend was also conserved in both RPLC and PGC experiments, we surveyed

the composition of high mannose glycans identified at all temperatures and found no significant disparities at the peptide level suggesting a cause for increased identifications (Figure S2.15-Figure S2.16).

Speculating the increase of high mannose glycopeptides may be due to species only being retained at 45°C, extracting the masses of identified glycopeptides demonstrated that most precursors could be identified across all temperatures, and that very few glycopeptides displayed temperature-dependent retention. A more consistent observation, however, is that different column temperatures result in different MS^1 peak heights/areas, likely impacting their selection in data-dependent experiments. The prevailing trend is for species to exhibit highest median peak heights at 45°C, either increasing or decreasing in magnitude at 60°C (Figure S2.17). We did not observe notable differences in MS^2 fragmentation across unique temperatures (Figure S2.18**)**, indicating MS^1 level differences are likely driving altered identification of glycopeptides. More intensive analysis is needed to determine the exact cause of this observation, though we hypothesize that 45°C may provide a more optimal balance between droplet desolvation and liquid-phase resolution, resulting in higher identification rates overall.

One potential concern of raising column temperatures is the additional energy provided to species prior to ionization. A consistent observation in both the complex and standard glycopeptide samples is that lower temperatures were more successful in identifying sialylated glycans (Figure 2.7**,** Figure S2.14). In general, we observed a decay in reporting signal as temperature increased (Figure S2.19) though a wider collection of sialylated species may be needed to confirm the universality of this trend. However, it cannot be ignored that in a small number of cases, little or no MS^1 signal could be observed for sialylated glycopeptides at 60°C. Considering the labile nature of glycosidic bonds, it is possible the increased energy prompted premature dissociation of

glycosidic bonds and therefore has direct adverse effects on identification. These hypotheses are currently under investigation.

These potential limitations notwithstanding, our data demonstrate both RPLC and PGC separations are benefited by moderate increases in running temperatures. Raising column temperature even higher may result in better isomeric resolution, as suggested by previous studies, but may cause undesired signal decay. This tradeoff must be carefully balanced depending on experimental objectives.

Conclusions

Taken together, our results demonstrate the incorporation of a PGC stationary phase grants complementary, though not orthogonal, access to the proteome and glycoproteome. While traditional RPLC separation regimes provide the greatest overall retention of peptides and glycopeptides, our results demonstrate RPLC preferentially retain longer, more hydrophobic species within solution. In contrast, PGC expands overall sample coverage by retaining those smaller, more hydrophilic peptides and glycopeptides that would otherwise go unidentified. Furthermore, we demonstrate that the glycopeptides retained by PGC have a more consolidated charge state distribution than those in RPLC experiments, generating more charge per length of peptide regardless of dominant hydrophilic character. As well, PGC separations provided greater microheterogeneity profiling depth by identifying more glycans per retained peptide. Our results also build on previous studies that show the capacity for liquid-phase separation of isomeric glycopeptides at higher temperatures. However, our results show a long PGC stationary phase is not exclusively necessary to provide improved liquid-phase separation. This knowledge indicates the combination of high-resolution RPLC stationary components in tandem with a PGC stationary phase could provide a useful strategy in high-throughput characterization studies. Finally, we

demonstrate that running both RPLC and PGC separations at higher temperatures provides altered peptide and glycopeptide identification rates, with the most optimal temperature shown to be 45°C. In summary, PGC enables pronounced benefits in proteomic and glycoproteomic analyses and provides a means towards facile, high-throughput characterization. As such, this regime is sure to exhibit utility in applications targeting unique post-translational modifications and as a separation strategy for structural interrogation.

Acknowledgments

The authors thank Dr. Qinying Yu, Danqing Wang and Dr. Dustin Frost in the Li Research Group for their helpful suggestions in the areas of glycopeptide enrichment and custom LC separations This work was funded in part by the National Institutes of Health (NIH) grants RF1 AG052324, U01CA231081, and R01 DK071801. The QE-HF Orbitrap instrument was generously provided by Thermo Fisher Scientific. LL would like to acknowledge NIH grant support R21AG065728, NCRRS10RR029531, and S10OD025084, a Pancreas Cancer Pilot grant from the University of Wisconsin Carbone Cancer Center (233-AAI9632), as well as a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Figures

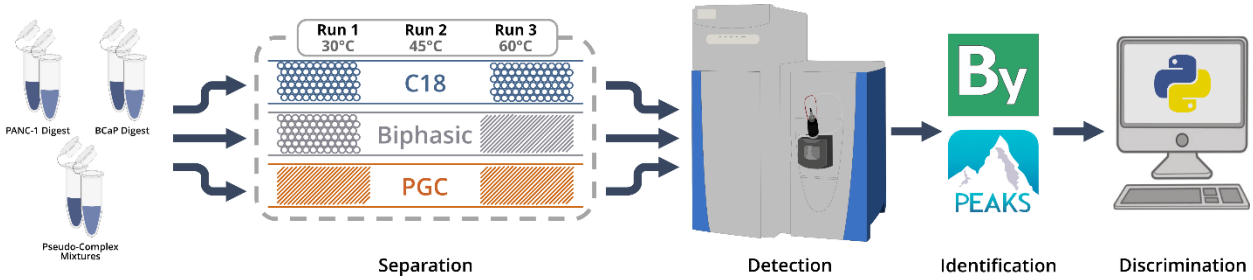


Figure 2.1 General workflow utilized throughout the experiment.

Samples were run sequentially at different temperatures (run 1: 30°C, run 2: 45°C, run 3: 60°C). Data was collected on a Thermo Q-Exactive-HF orbitrap mass sepctrometer, with peptide and glycopeptide annotations provided by PEAKS and Byonic prior to custom data analysis.

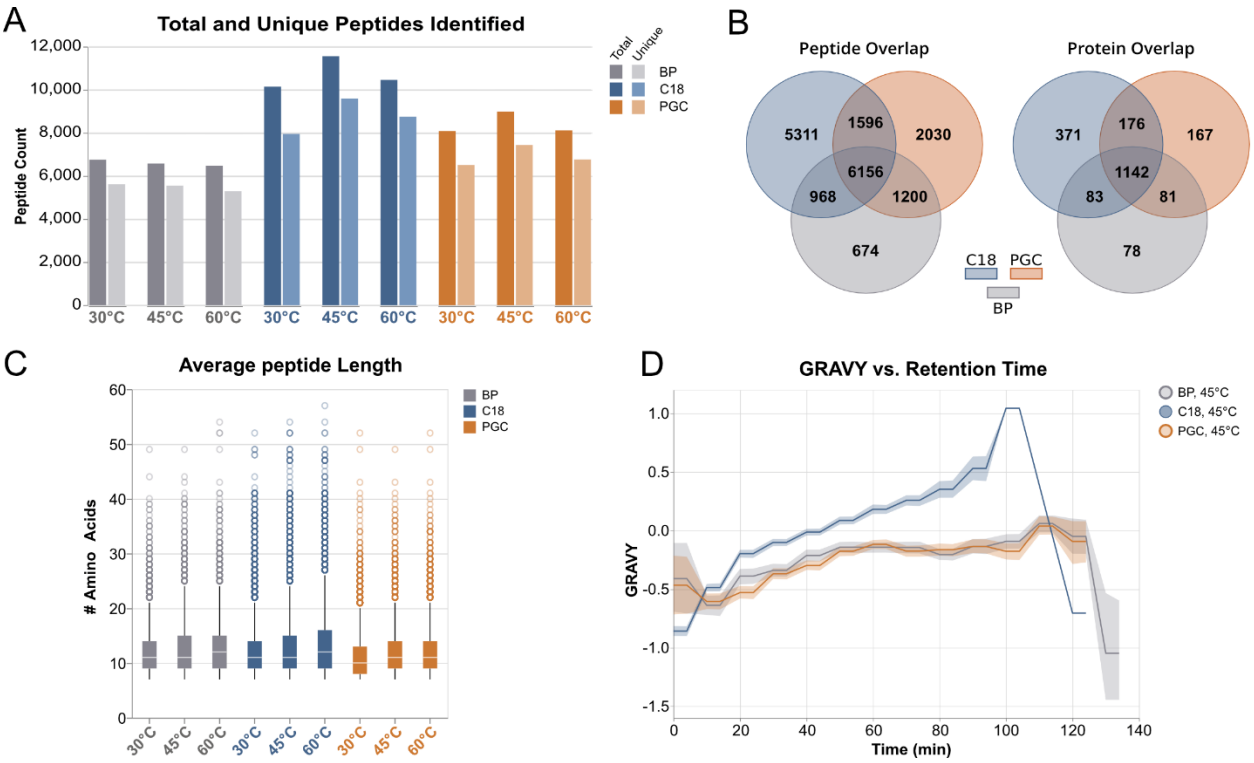


Figure 2.2 High-level overview of identified peptides resulting from analysis of PANC1 cell lysate digests.

A) Comparison of total and unique peptides identified between all columns at varying temperatures. **B**) Overlap of identified peptides (left) and identified proteins (right), demonstrating the clear distinction in the species retained in PGC and RPLC separations. **C**) Comparison of identified peptide lengths. **D**) Hydrophilicity of identified peptides as a function of time, demonstrating the complementary retention between PGC and RPLC. Lesser values are associated with hydrophilic character while greater values are associated with hydrophobic character. The line represents average GRAVY score, shaded area represents 95% confidence interval.

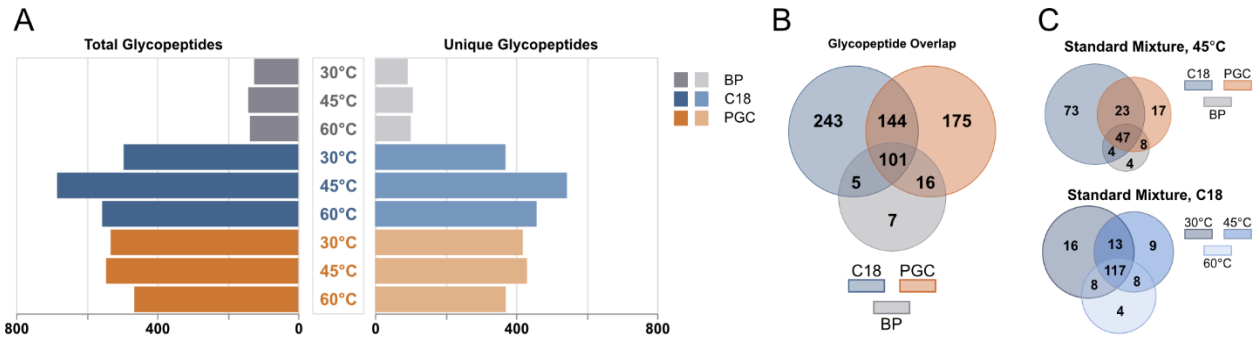


Figure 2.3 Overview of glycopeptides identified between all separation regimes.

Overview of glycopeptides identified between all separation regimes. **A**) Comparison of total (left) and unique (right) glycopeptides identified across all samples, shown according to separation strategy and column temperature. **B**) Overlap of all glycopeptides identified in each separation phase across all samples. **C**) (top) Overlap of glycopeptides identified in the standard mixture when column temperature is held constant and column is varied, (bottom) overlap of glycopeptides from standard mixture when separation regime is constant and temperature is varied.

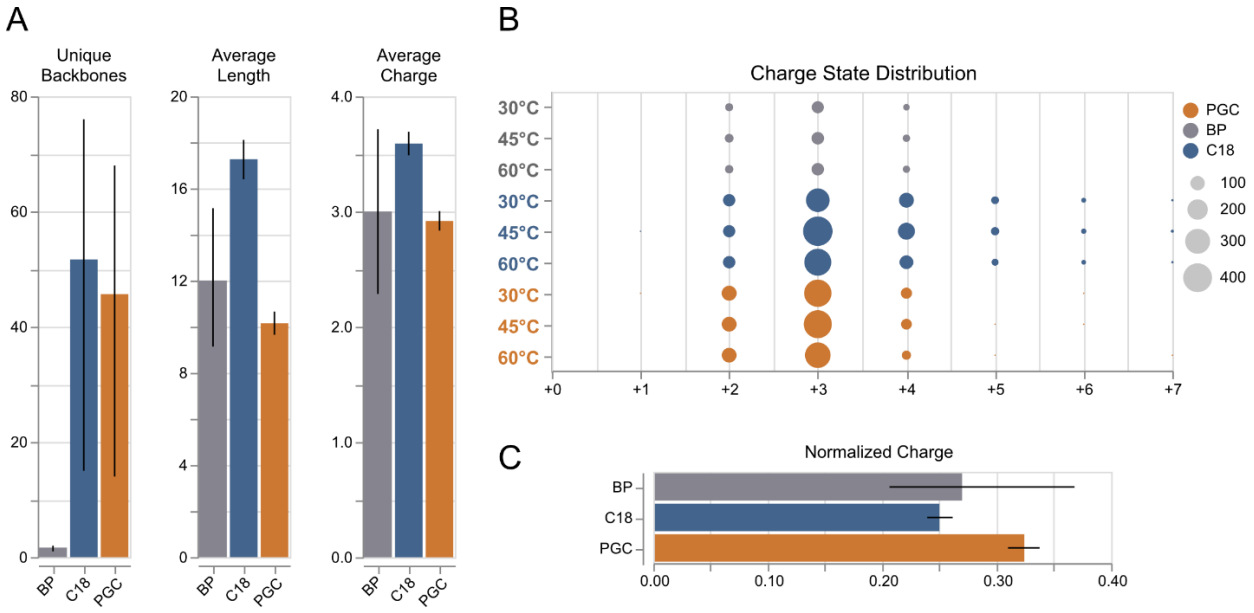


Figure 2.4 Peptide level differences between glycopeptides identified across all experiments. **A)** Distribution of number unique backbones identified, average glycopeptide length, and average glycopeptide charge between all separation regimes. Error bars: 95% confidence inteval, *n*=3 (each sample). **B**) Distribution of glycopeptide charge states between all separations and temperatures. **C**) Normalized charge per glycopeptide, shown as a function of average charge/amino acid residue. Error bars: 95% confidence inteval, *n*=3 (each sample).

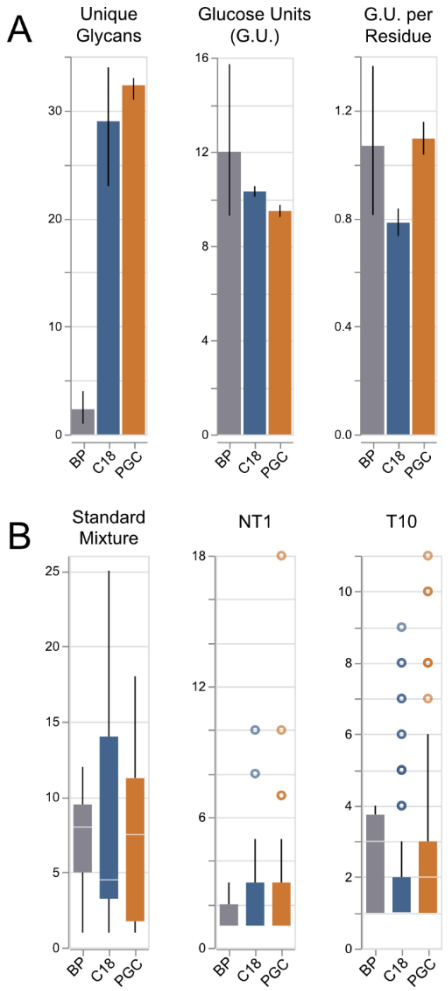


Figure 2.5 Comparison of glycopeptide characteristics.

A) Distribution of # unique glycans, glucose units per glycan, and glucose units per amino acid residue. Error bars: 95% confidence inteval, *n*=3 (each sample). **B**) Boxplots displaying the number of glycans associated with retained peptide bacbones. Note: for samples "NT1" and "T10", the median values of RPLC are found to be 1, making them difficult to view here.

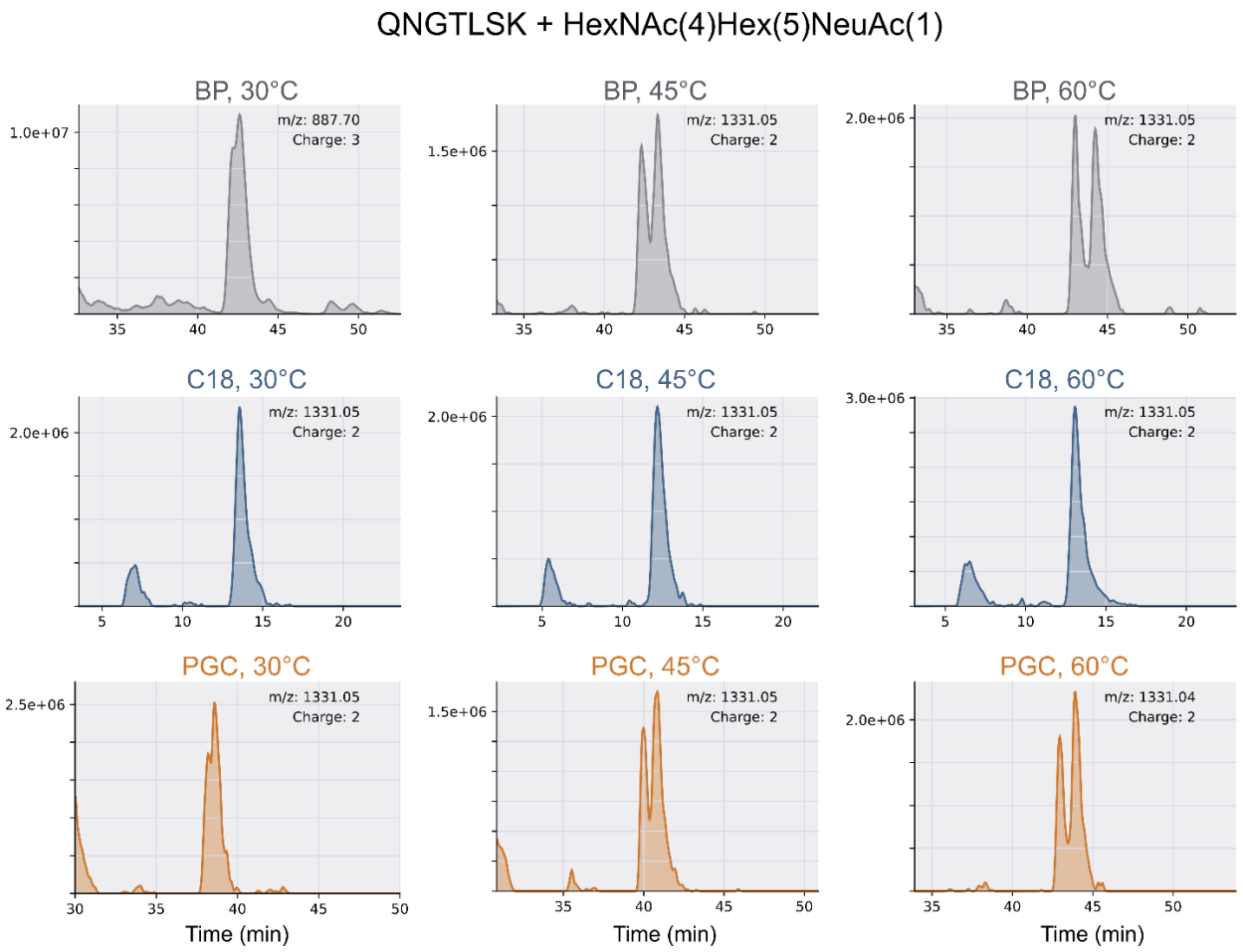


Figure 2.6 Extract ion chromatogram (EIC) of a representative glycopeptide identified on all columns at all temperatures.

As the retention mechanism of RPLC is driven by hydrophobic interactions, the lack of any liquid-phase resolution is expected. However, both a longer (L=30cm) and short (L=15cm) PGC stationary phases were shown to provide adequate liquid phase resolution of isomeric glycopeptides. Note, the 3+ charge state in BP, 30°C has been manually confirmed as correct.

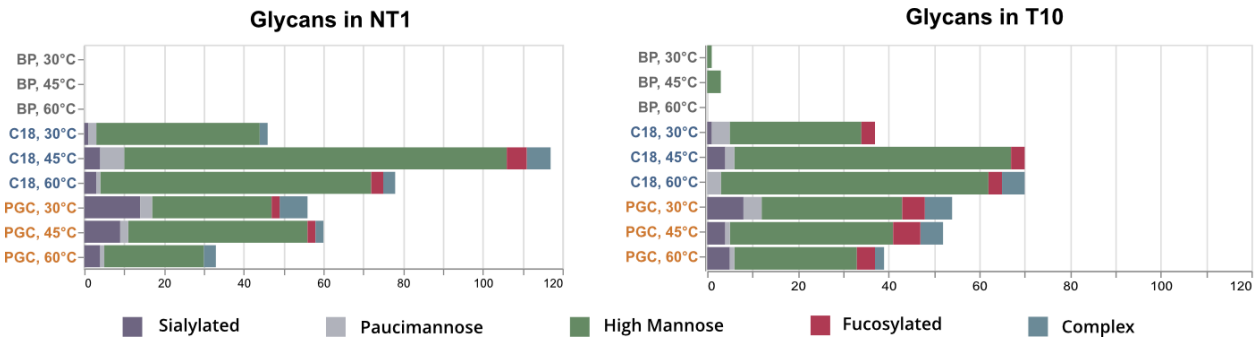


Figure 2.7 Distribution of glycan types identified within prostate cancer cell lines NT1 and T10.

Glycan class (i.e., High Mannose, Complex, etc.) are colored to highlight the changing quantities of glycans as a function of column temperature.

References

1. Varki, A., *Essentials of glycobiology*. Third edition. ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 2017; p xxix, 823 pages.

2. Sperandio, M.; Gleissner, C. A.; Ley, K., Glycosylation in immune cell trafficking. *Immunol Rev* **2009,** *230* (1), 97-113.

3. Hobbs, S. J.; Nolz, J. C., Regulation of T Cell Trafficking by Enzymatic Synthesis of O-Glycans. *Frontiers in Immunology* **2017,** *8* (600).

4. Sackstein, R.; Merzaban, J. S.; Cain, D. W.; Dagia, N. M.; Spencer, J. A.; Lin, C. P.; Wohlgemuth, R., Ex vivo glycan engineering of CD44 programs human multipotent mesenchymal stromal cell trafficking to bone. *Nature Medicine* **2008,** *14* (2), 181-187.

5. Dong, Y.; Li, W.; Gu, Z.; Xing, R.; Ma, Y.; Zhang, Q.; Liu, Z., Inhibition of HER2-Positive Breast Cancer Growth by Blocking the HER2 Signaling Pathway with HER2-Glycan-Imprinted Nanoparticles. *Angew Chem Int Ed Engl* **2019,** *58* (31), 10621-10625.

6. Xuan, L.; Luan, G.; Wang, Y.; Lan, F.; Zhang, X.; Hao, Y.; Zheng, M.; Wang, X.; Zhang, L., MicroRNAs regulating mucin type O-glycan biosynthesis and transforming growth factor β signaling pathways in nasal mucosa of patients with chronic rhinosinusitis with nasal polyps in Northern China. *Int Forum Allergy Rhinol* **2019,** *9* (1), 106-113.

7. Au - Bagavant, H.; Au - Trzeciak, M.; Au - Papinska, J.; Au - Biswas, I.; Au - Dunkleberger, M. L.; Au - Sosnowska, A.; Au - Deshmukh, U. S., A Method for the Measurement of Salivary Gland Function in Mice. *JoVE* **2018,** (131), e57203.

8. Chaffey, P. K.; Guan, X.; Wang, X.; Ruan, Y.; Li, Y.; Miller, S. G.; Tran, A. H.; Koelsch, T. N.; Pass, L. F.; Tan, Z., Quantitative Effects of O-Linked Glycans on Protein Folding. *Biochemistry* **2017,** *56* (34), 4539-4548.

9. Liu, Y.-S.; Guo, X.-Y.; Hirata, T.; Rong, Y.; Motooka, D.; Kitajima, T.; Murakami, Y.; Gao, X.-D.; Nakamura, S.; Kinoshita, T.; Fujita, M., N-Glycan–dependent protein folding and endoplasmic reticulum retention regulate GPI-anchor processing. *Journal of Cell Biology* **2017,** *217* (2), 585-599.

10. Kuribara, T.; Totani, K., Structural insights into N-linked glycan-mediated protein folding from chemical and biological perspectives. *Current Opinion in Structural Biology* **2021,** *68*, 41-47.

11. Čaval, T.; Heck, A. J. R.; Reiding, K. R., Meta-heterogeneity: evaluating and describing the diversity in glycosylation between sites on the same glycoprotein. *Molecular & Cellular Proteomics* **2021**, 100010.

12. Losev, Y.; Frenkel-Pinter, M.; Abu-Hussien, M.; Viswanathan, G. K.; Elyashiv-Revivo, D.; Geries, R.; Khalaila, I.; Gazit, E.; Segal, D., Differential effects of putative N-glycosylation sites in human Tau on Alzheimer's disease-related neurodegeneration. *Cellular and Molecular Life Sciences* **2020**.

13. Losev, Y.; Paul, A.; Frenkel-Pinter, M.; Abu-Hussein, M.; Khalaila, I.; Gazit, E.; Segal, D., Novel model of secreted human tau protein reveals the impact of the abnormal N-glycosylation of tau on its aggregation propensity. *Scientific Reports* **2019,** *9* (1), 2254.

14. Frenkel-Pinter, M.; Shmueli, M. D.; Raz, C.; Yanku, M.; Zilberzwige, S.; Gazit, E.; Segal, D., Interplay between protein glycosylation pathways in Alzheimer's disease. *Science Advances* **2017,** *3* (9), e1601576.

15. Videira, P. A. Q.; Castro-Caldas, M., Linking Glycation and Glycosylation With Inflammation and Mitochondrial Dysfunction in Parkinson's Disease. *Frontiers in Neuroscience* **2018,** *12* (381).

16. Thompson, J. W.; Sorum, A. W.; Hsieh-Wilson, L. C., Deciphering the Functions of O-GlcNAc Glycosylation in the Brain: The Role of Site-Specific Quantitative O-GlcNAcomics. *Biochemistry* **2018,** *57* (27), 4010-4018.

17. Reindl, M.; Waters, P., Myelin oligodendrocyte glycoprotein antibodies in neurological disease. *Nature Reviews Neurology* **2019,** *15* (2), 89-102.

18. Stack, J. R.; Madigan, A.; Helbert, L.; Dunne, E.; Gardiner, E. E.; Andrews, R. K.; Finan, R.; Smyth, E.; Kenny, D.; McCarthy, G. M., Soluble glycoprotein VI, a specific marker of platelet activation is increased in the plasma of subjects with seropositive rheumatoid arthritis. *PLOS ONE* **2017,** *12* (11), e0188027.

19. Salem, D.; Subang, R.; Kuwana, M.; Levine, J. S.; Rauch, J., T cells from induced and spontaneous models of SLE recognize a common T cell epitope on β2-glycoprotein I. *Cell Mol Immunol* **2019,** *16* (8), 685-693.

20. Rauch, J.; Salem, D.; Subang, R.; Kuwana, M.; Levine, J. S., β2-Glycoprotein I-Reactive T Cells in Autoimmune Disease. *Frontiers in Immunology* **2018,** *9* (2836).

21. Mereiter, S.; Balmaña, M.; Campos, D.; Gomes, J.; Reis, C. A., Glycosylation in the Era of Cancer-Targeted Therapy: Where Are We Heading? *Cancer Cell* **2019,** *36* (1), 6-16.

22. Holst, S.; Belo, A. I.; Giovannetti, E.; van Die, I.; Wuhrer, M., Profiling of different pancreatic cancer cells used as models for metastatic behaviour shows large variation in their N-glycosylation. *Scientific Reports* **2017,** *7* (1), 16623.

23. Ideo, H.; Kondo, J.; Nomura, T.; Nonomura, N.; Inoue, M.; Amano, J., Study of glycosylation of prostate-specific antigen secreted by cancer tissue-originated spheroids reveals new candidates for prostate cancer detection. *Scientific Reports* **2020,** *10* (1), 2708.

24. Riley, N. M.; Bertozzi, C. R.; Pitteri, S. J., A Pragmatic Guide to Enrichment Strategies for Mass Spectrometry-based Glycoproteomics. *Molecular & Cellular Proteomics* **2021**, 100029.

25. Xue, Y.; Xie, J.; Fang, P.; Yao, J.; Yan, G.; Shen, H.; Yang, P., Study on behaviors and performances of universal N-glycopeptide enrichment methods. *Analyst* **2018,** *143* (8), 1870-1880.

26. Qing, G.; Yan, J.; He, X.; Li, X.; Liang, X., Recent advances in hydrophilic interaction liquid interaction chromatography materials for glycopeptide enrichment and glycan separation. *TrAC Trends in Analytical Chemistry* **2020,** *124*, 115570.

27. Riley, N. M.; Malaker, S. A.; Driessen, M. D.; Bertozzi, C. R., Optimal Dissociation Methods Differ for N- and O-Glycopeptides. *Journal of Proteome Research* **2020**.

28. Riley, N. M.; Hebert, A. S.; Westphall, M. S.; Coon, J. J., Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nature Communications* **2019,** *10* (1), 1311.

29. Escobar, E. E.; King, D. T.; Serrano-Negrón, J. E.; Alteen, M. G.; Vocadlo, D. J.; Brodbelt, J. S., Precision Mapping of O-Linked N-Acetylglucosamine Sites in Proteins Using Ultraviolet Photodissociation Mass Spectrometry. *Journal of the American Chemical Society* **2020,** *142* (26), 11569-11577.

30. Wei, J.; Tang, Y.; Ridgeway, M. E.; Park, M. A.; Costello, C. E.; Lin, C., Accurate Identification of Isomeric Glycans by Trapped Ion Mobility Spectrometry-Electronic Excitation Dissociation Tandem Mass Spectrometry. *Analytical Chemistry* **2020,** *92* (19), 13211-13220.

31. Pathak, P.; Baird, M. A.; Shvartsburg, A. A., High-Resolution Ion Mobility Separations of Isomeric Glycoforms with Variations on the Peptide and Glycan Levels. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (7), 1603-1609.

32. Mookherjee, A.; Guttman, M., Bridging the structural gap of glycoproteomics with ion mobility spectrometry. *Current Opinion in Chemical Biology* **2018,** *42*, 86-92.

33. Chen, Z.; Glover, M. S.; Li, L., Recent advances in ion mobility–mass spectrometry for improved structural characterization of glycans and glycoconjugates. *Current Opinion in Chemical Biology* **2018,** *42*, 1-8.

34. Zhou, S.; Huang, Y.; Dong, X.; Peng, W.; Veillon, L.; Kitagawa, D. A. S.; Aquino, A. J. A.; Mechref, Y., Isomeric Separation of Permethylated Glycans by Porous Graphitic Carbon (PGC)-LC-MS/MS at High Temperatures. *Analytical Chemistry* **2017,** *89* (12), 6590-6597.

35. Wei, J.; Tang, Y.; Bai, Y.; Zaia, J.; Costello, C. E.; Hong, P.; Lin, C., Toward Automatic and Comprehensive Glycan Characterization by Online PGC-LC-EED MS/MS. *Analytical Chemistry* **2020,** *92* (1), 782-791.

36. Ashwood, C.; Lin, C.-H.; Thaysen-Andersen, M.; Packer, N. H., Discrimination of Isomers of Released N- and O-Glycans Using Diagnostic Product Ions in Negative Ion PGC-LC-ESI-MS/MS. *Journal of the American Society for Mass Spectrometry* **2018,** *29* (6), 1194-1209.

37. Ashwood, C.; Waas, M.; Weerasekera, R.; Gundry, R. L., Reference glycan structure libraries of primary human cardiomyocytes and pluripotent stem cell-derived cardiomyocytes reveal cell-type and culture stage-specific glycan phenotypes. *Journal of Molecular and Cellular Cardiology* **2020,** *139*, 33-46.

38. Zhu, R.; Huang, Y.; Zhao, J.; Zhong, J.; Mechref, Y., Isomeric Separation of N-Glycopeptides Derived from Glycoproteins by Porous Graphitic Carbon (PGC) LC-MS/MS. *Analytical Chemistry* **2020,** *92* (14), 9556-9565.

39. Liu, T. T.; Ewald, J. A.; Ricke, E. A.; Bell, R.; Collins, C.; Ricke, W. A., Modeling human prostate cancer progression in vitro. *Carcinogenesis* **2018,** *40* (7), 893-902.

40. Horoszewicz, J. S.; Leong, S. S.; Kawinski, E.; Karr, J. P.; Rosenthal, H.; Chu, T. M.; Mirand, E. A.; Murphy, G. P., LNCaP Model of Human Prostatic Carcinoma. *Cancer Research* **1983,** *43* (4), 1809-1818.

41. Chandler, K. B.; Leon, D. R.; Kuang, J.; Meyer, R. D.; Rahimi, N.; Costello, C. E., N-Glycosylation regulates ligand-dependent activation and signaling of vascular endothelial growth factor receptor 2 (VEGFR2). *Journal of Biological Chemistry* **2019,** *294* (35), 13117-13130.

42. Chandler, K. B.; Costello, C. E., Glycomics and glycoproteomics of membrane proteins and cell-surface receptors: Present trends and future opportunities. *Electrophoresis* **2016,** *37* (11), 1407-1419.

43. Lin, C.-Y.; Lee, C.-H.; Chuang, Y.-H.; Lee, J.-Y.; Chiu, Y.-Y.; Wu Lee, Y.-H.; Jong, Y.-J.; Hwang, J.-K.; Huang, S.-H.; Chen, L.-C.; Wu, C.-H.; Tu, S.-H.; Ho, Y.-S.; Yang, J.-M., Membrane protein-regulated networks across human cancers. *Nature Communications* **2019,** *10* (1), 3131.

44. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A., The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **2019,** *47* (D1), D442-d450.

45. VanderPlas, J.; Granger, B.; Heer, J.; Moritz, D.; Wongsuphasawat, K.; Satyanarayan, A.; Lees, E.; Timofeev, I.; Welsh, B.; Sievert, S., Altair: Interactive Statistical Visualizations for Python. The Open Journal: 2018; p 1057.

46. Stavenhagen, K.; Plomp, R.; Wuhrer, M., Site-Specific Protein N- and O-Glycosylation Analysis by a C18-Porous Graphitized Carbon–Liquid Chromatography-Electrospray Ionization Mass Spectrometry Approach Using Pronase Treated Glycopeptides. *Analytical Chemistry* **2015,** *87* (23), 11691-11699.

47. Zhou, S.; Dong, X.; Veillon, L.; Huang, Y.; Mechref, Y., LC-MS/MS analysis of permethylated N-glycans facilitating isomeric characterization. *Analytical and Bioanalytical Chemistry* **2017,** *409* (2), 453-466.

48. Liu, T. T.; Ewald, J. A.; Ricke, E. A.; Bell, R.; Collins, C.; Ricke, W. A., Modeling human prostate cancer progression in vitro. *Carcinogenesis* **2019,** *40* (7), 893-902.

49. de Leoz, M. L.; Young, L. J.; An, H. J.; Kronewitter, S. R.; Kim, J.; Miyamoto, S.; Borowsky, A. D.; Chew, H. K.; Lebrilla, C. B., High-mannose glycans are elevated during breast cancer progression. *Mol Cell Proteomics* **2011,** *10* (1), M110.002717.

50. Lattová, E.; Skřičková, J.; Hausnerová, J.; Frola, L.; Křen, L.; Ihnatová, I.; Zdráhal, Z.; Bryant, J.; Popovič, M., N-Glycan profiling of lung adenocarcinoma in patients at different stages of disease. *Modern Pathology* **2020,** *33* (6), 1146-1156.

51. Brown Chandler, K.; E. Costello, C.; Rahimi, N., Glycosylation in the Tumor Microenvironment: Implications for Tumor Angiogenesis and Metastasis. *Cells* **2019,** *8* (6), 544.

52. An, H. J.; Gip, P.; Kim, J.; Wu, S.; Park, K. W.; McVaugh, C. T.; Schaffer, D. V.; Bertozzi, C. R.; Lebrilla, C. B., Extensive Determination of Glycan Heterogeneity Reveals an Unusual Abundance of High Mannose Glycans in Enriched Plasma Membranes of Human Embryonic Stem Cells*. *Molecular & Cellular Proteomics* **2012,** *11* (4), M111.010660.

53. Park, D. D.; Phoomak, C.; Xu, G.; Olney, L. P.; Tran, K. A.; Park, S. S.; Haigh, N. E.; Luxardi, G.; Lert-itthiporn, W.; Shimoda, M.; Li, Q.; Matoba, N.; Fierro, F.; Wongkham, S.; Maverakis, E.; Lebrilla, C. B., Metastasis of cholangiocarcinoma is promoted by extended high-mannose glycans. *Proceedings of the National Academy of Sciences* **2020,** *117* (14), 7633-7644.

Supplemental Methods

Cell preparation

The needed solutions were prepared as follows:

Reagent	Instructions/Recipe
1x PBS Buffer	From purchased stock
200 mM HEPES, pH 8.0	476.6 mg HEPES dissolved in ultrapure water Adjust pH to 8.0 using HCl or NaOH
Protease inhibitor	Collect 1 tablet each of Roche protease (5892970001) and phosphatase inhibitor (4906845001) Dissolve 1 tablet in 1 mL of ultrapure water to yield 10x solution (can be stored up to 12 weeks in -20°C) Dilute to 1x using 9:1 water: inhibitor ratio
Urea Lysis Buffer	**Must be made fresh** 10 mL: 1 mL 200 mM HEPES (pH 8.0) + 4.8 g Urea + 8 mL ddH_2O + 1 mL protease inhibitor After adding urea, volume of water used to dissolve is approximate. Final volume should be 10mL, even though volume of water added is not exactly 8mL
200 mM HEPES Buffer	Dissolve 476.6 mg HEPES in 10 mL ddH_2O
500 mM DTT	Dissolve 77.125 mg DTT in 1mL 200 mM HEPES
500 mM IAA	Dissolve 92.48 mg IAA in 1 mL 200 mM HEPES
50 mM Tris-HCl	Dissolve 39.53 mg ABC in 10 mL dd H_2O
Trypsin	Purchased from Promega (V5113)

Cell pellets (~50-100uL) were first washed three times with 500uL PBS. Each wash entailed adding PBS, mixing via pipette, and spinning down at 14,000rcf at 4C for 5 minutes. After extracting the final PBS wash, cell pellets were resuspended in 250uL urea lysis buffer with no additional mixing. Cell pellets were then lysed via ultrasonication using pulse 3 second pulses (3s on, 3s off) for a total sonication time of 1minute at the lowest amplitude. If any remaining cells could be observed, the cell pellets were left on ice for 10 minutes and sonicated again.

Following sonication, cell debris was removed by centrifugation at 14,000rcf for 15 minutes at 4C. The resulting supernatant was removed and placed in a 10kDa MWCO filter

(Millipore) and centrifuged at 14,000rcf for 10 minutes. Flow through was discarded and remaining sample volume was adjusted to 200uL with urea lysis buffer. DTT was added 1:50 (v:v) and incubated at room temperature for 60 minutes. Buffer was then exchanged by centrifuging at 14,000rcf for 10 minutes at 4C and readjusting volume to 200uL with urea lysis buffer. IAA was then added 1:25 (v:v) and incubated at room temperature for 30 minutes in the dark. Buffer was exchanged three times using the same method as previous two steps. Following this samples were washed with Tris-HCl three times using the same buffer exchange method. After the third wash, trypsin was added 1:50 (w:w) and incubated overnight (~18hr) at 37C. Cutoff filters were then placed in fresh tubes and peptide mixtures were collected via centrifugation at 14,000 rcf for 10 minutes. Filters were washed three times with Tris-HCl and flow through was kept at each step.

Standard glycoprotein preparation

Bovine fetuin, bovine ribonuclease B, and bovine a1-Acid Glycoprotein were reconstituted in ultrapure water. Concentrations were estimated via Pierce BCA Assay and two 500mg aliquots of each protein were dried to completion under vacuum. Urea, DTT, and IAA were prepared at concentrations of 8M, 100mM, and 200mM concentrations, respectively, with 50mM tris-HCl. Dried protein was reconstituted in 100µL Urea and mixed at room temperature for 5 minutes. DTT was added to yield 2:5 DTT:protein ratio (26.32µL), mixed briefly and incubated for 1 hour at room temperature. IAA was then added to yield a 2:3 IAA:protein ratio (42.67µL), mixed and incubated at room temperature for 30 minutes in the dark. Excess IAA was quenched by addition of DTT, mixed and left to stand for 5 minutes. Solutions were then diluted to lower urea concentration to 0.9M and mixed 1:50 with trypsin. Proteins were digested overnight at 37°C and desalted via reverse phase desalting cartridges.

Glycopeptide Enrichment

Before beginning, the following solutions were prepared: 1%TFA in ultrapure water, 80% ACN + 1% TFA in ultrapure water, and 10% ACN + 5% FA in ultrapure water.

polyHYDROXYETHYL-A beads were weighed out in a 30:1 beads:peptide mass ratio (e.g. 300mg peptides required 9mg beads). Beads were then resuspended in 1% TFA solution, using 200uL for every 10mg beads (e.g., 9mg beads were added to 180uL). Mixtures were vortexed for 15 minutes. 3.0 mg sterile cotton was inserted tightly in to a 200uL pipette tip and placed within a microtube as shown below. The bead slurry was then added to the pipette tip and centrifuged at 0.2kg for 2 minutes to remove liquid. Beads were then washed three times with 300uL aliquots of 1% TFA, centrifuging at 0.2kg for 2 minutes each time. Beads were conditioned with three aliquots of 300uL 80% ACN, centrifuging for 5 minutes at 0.1kg. Peptide samples were reconstituted in 80% ACN and loaded into pipette tips. Samples were flowed through via centrifugation (0.1kg, 5minutes). The flow-through was collected and readministered 5 times. Samples were washed with 6 aliquots of 80% ACN, with centrifugation at 0.2kg for 2 minutes. Flow through was separated at each step. Remaining glycopeptides were then eluted into fresh tube using three 150uL aliquots of 10% ACN via centrifugation at 0.2kg for 3 minutes.

Analytical column preparation

Three, 40cm lengths of capillary tubing (360µm o.d., 75µm i.d.) were cut and flushed with methanol. While drying, a frit solution of potassium silicate (Kasil) and formamide was prepared according to manufacturer instructions. Once dry, one end of each capillary was inserted into the frit solution for 10 seconds. Frits were dried at 100°C overnight, inspected under microscope and trimmed to a final length of ~2mm. RPLC C18 (3µm d.p.) and PGC (3µm d.p.) packing materials were reconstituted in chloroform and MeOH, respectively. Using a column packer (Next

Advance), packing solutions were pushed through at 1200psi, increasing pressure gradually as needed, until packed portions had reached 30cm. C18 columns were packed tight with 100% ACN for 20 minutes at 1500psi and PGC columns with 100% MeOH, each followed by flowing nitrogen over packed bed for 20min at 1500psi. Note: this process is consisted for the biphasic (BP) columns, only with packing phases stopped at 15cm.

Searching Parameters

PEAKS X

All PANC1 cell lysate datasets were searched against the UniProt Human Proteome (Reviewed only) using the following settings: Parent mass error, 20ppm; fragment mass error, 0.02 Da; precursor search type, monoisotopic; enzyme, trypsin; max missed cleavages, 2; digest mode, unspecific; fixed modifications, carbamidomethylation; variable modifications, oxidation (M), acetylation (N-term); max variable mod per peptide, 3.

Byonic

All standard glycopeptide mixtures were searched against a custom fasta database containing only the UniProt sequences belonging to the three proteins in use (Bovine Fetuin, Bovine a1-Acid Glycoprotein, Bovine Ribonuclease B) using the following parameters: Cleavage site: RK; cleavage side: C-terminal; digestion specificity, fully specific; missed cleavages, 2; precursor mass tolerance, 25; fragmentation type, QTOF/HCD; fragment mass tolerance, 0.02 Da; recalibration, none. Modifications: Carbamidomethyl (fixed), oxidation (M, variable, common1), acetyl (N-term, variable, common1); total common max, 4; total rare max, 1. Glycans: N-glycan 73 bovine as rare1 modifications. All other settings left as default.

NT1 and T10 glycopeptides were searched against the UniProt Human proteome using the following parameters: Cleavage site: RK; cleavage side: C-terminal; digestion specificity, fully specific; missed cleavages, 2; precursor mass tolerance, 25; fragmentation type, QTOF/HCD; fragment mass tolerance, 0.02 Da; recalibration, none. Modifications: Carbamidomethyl (fixed), oxidation (M, variable, common1), acetyl (N-term, variable, common1); total common max, 4; total rare max, 1. Glycans: N-glycan 309 mammalian no sodium as rare1 modifications. All other settings left as default.

Results were custom filtered, using the following criteria. Byonic score ≥ 200, Delta Mod ≥ 10, Log Prob ≥ 1.

Supplemental Tables

For brevity, the tables referred to in the manuscript have been omitted. These tables may be found alongside the online version of this manuscript (10.1007/s00216-022-03934-7).

Supplemental Figures

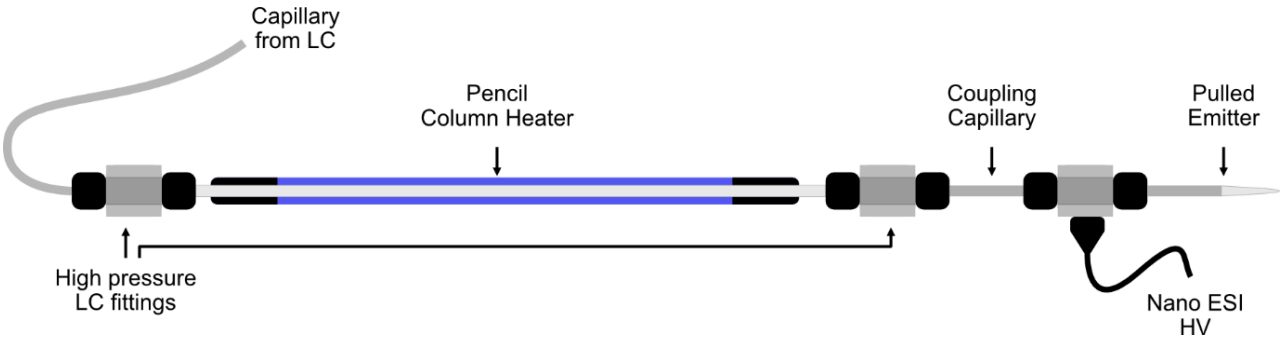


Figure S2.1 Graphical representation of LC connections coupling column to LC, heater and ESI.

The high-pressure LC fittings were purchased from IDEX Health & Science (UH-436) and accommodate 360um o.d. capillary tubing with zero dead volume. The coupling capillary (~5cm) was implemented to protect the stationary phase from direct HV application. The emitter was pulled from 360um o.d. capillary tubing and was ~4cm in length.

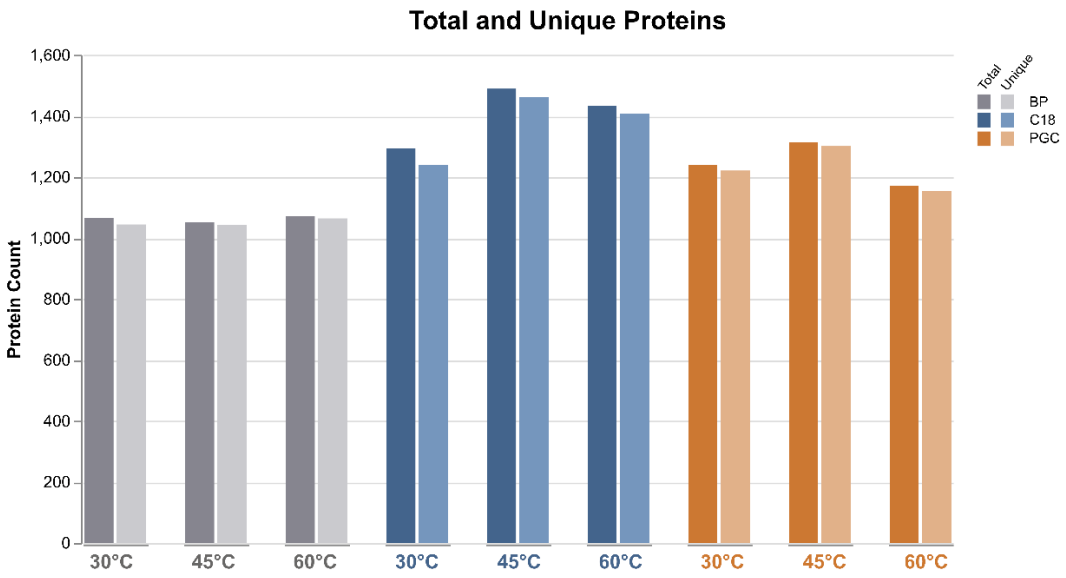


Figure S2.2 Total and unique proteins identified during analysis of PANC1 cell lysate digests.

Total and unique proteins identified during analysis of PANC1 cell lysate digests. PGC, C18 and BP analyses were performed sequentially with temperatures varied from low to high.

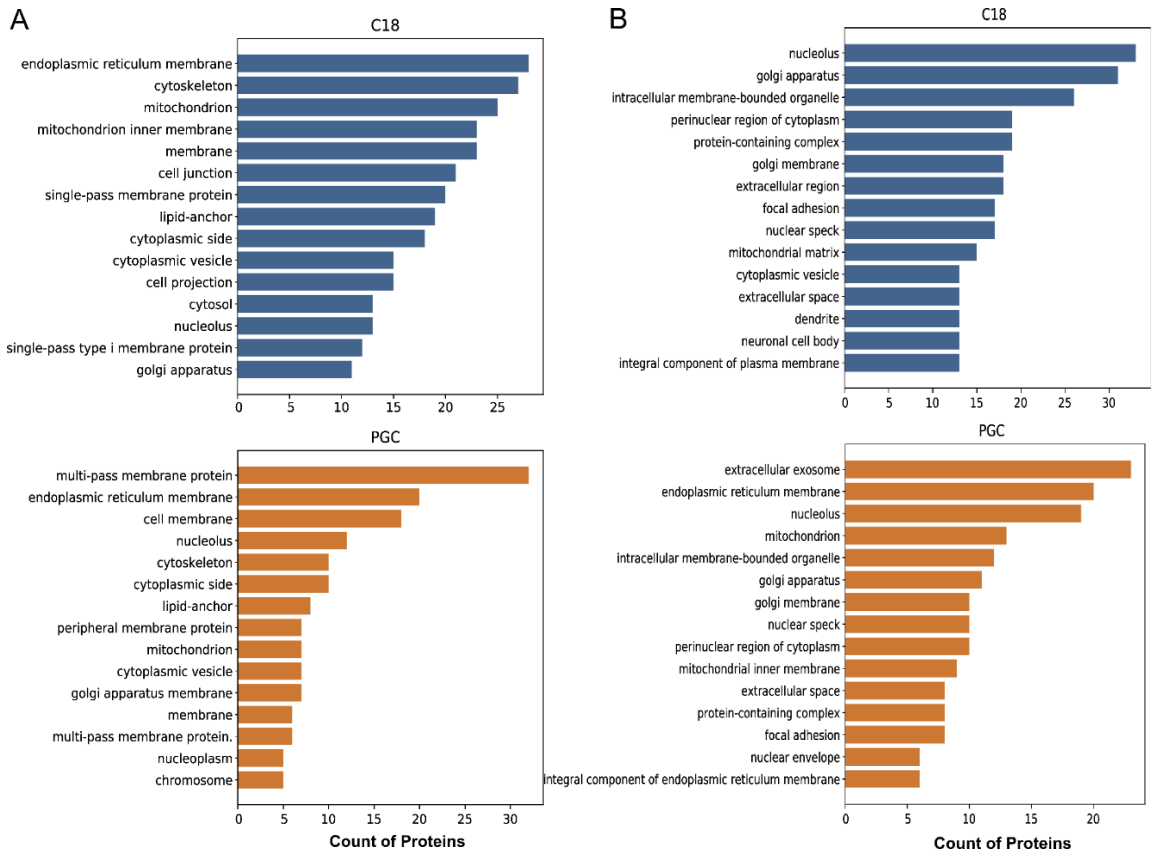


Figure S2.3 Subcellular location and compartment terms.

Count of proteins belonging to A) subcellular location and B) cellular compartment terms associated with unique proteins identified in C18 and PGC analyses. Terms are compiled from the UniProt Human Proteome database. This list of entries is comprised of the top 90% of counted terms. This top 90% provides a succinct, relevant list of terms useful for discussion and removes all cases where a term is counted only one time. When examining these data, PGC and C18 seem to provide directed retention of specific entities, *with respect to the proteins identified in this experimen*t. While these findings suggest some useful differences between each separation regime, we are careful not to present these as absolute. In future studies where a larger population of proteins are identified in each separation regime, it will be evident whether a separation strategy provides directed analysis of protein subclasses. But as of now, the dataset is limited in scope and these conclusions must remain as potential outcomes of later investigations.

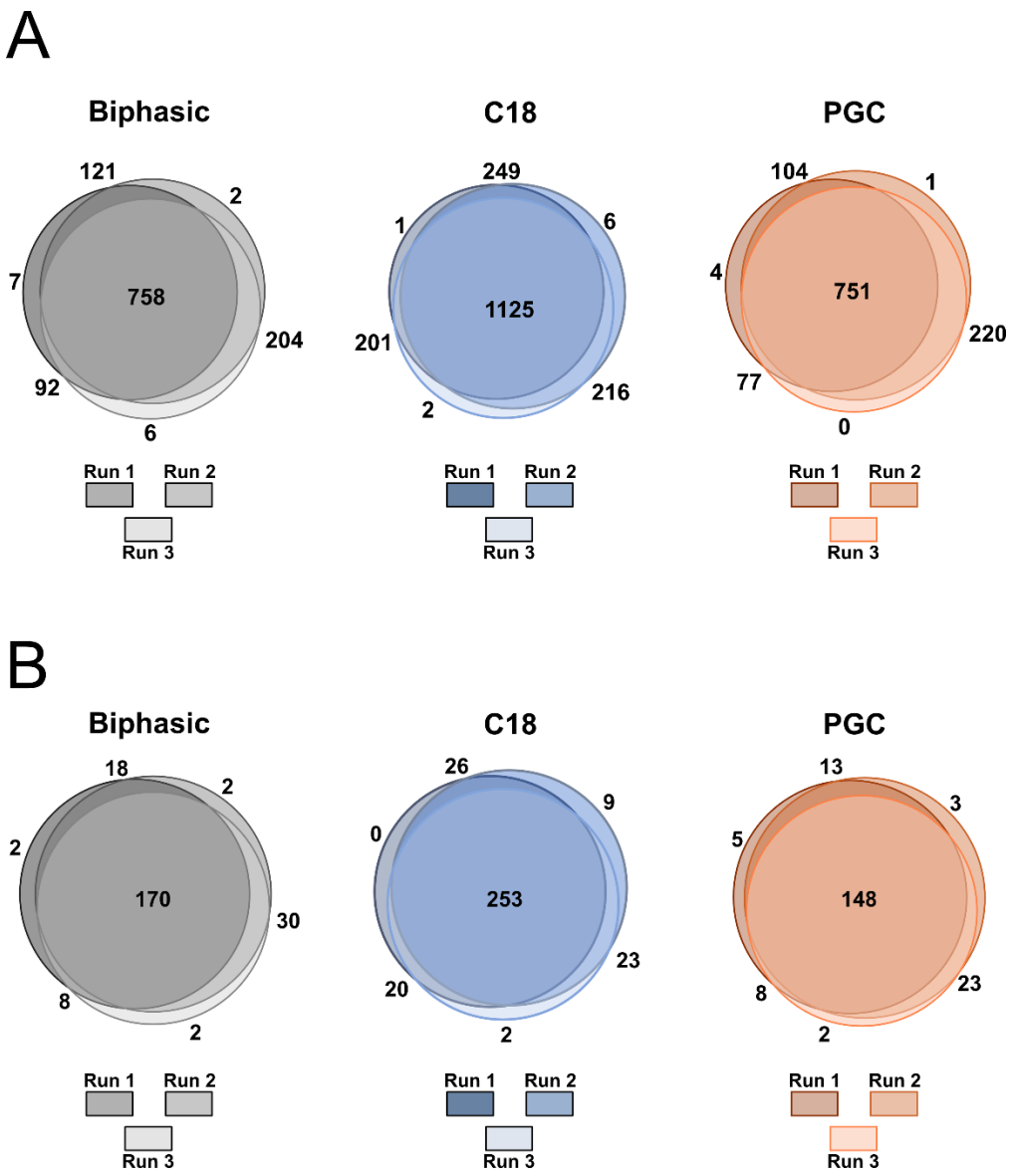


Figure S2.4 Glycopeptide and glycoprotein overlap.

To ensure reliability and reproducibility of our analyses, we performed follow-up analyses increase confidence in our initial observations. Examining the A) glycopeptide and B) protein overlap taken from triplicate analysis of LNCaP cell line digests, we demonstrate high inter-run reproducibility for each of our custom separation strategies. This reproducibility and the analysis of multiple unique biological samples within the main text increase our confidence that observations within

the main text truly stem from unique retention mechanisms and not sample/run variance. All triplicates were collected at 45°C.

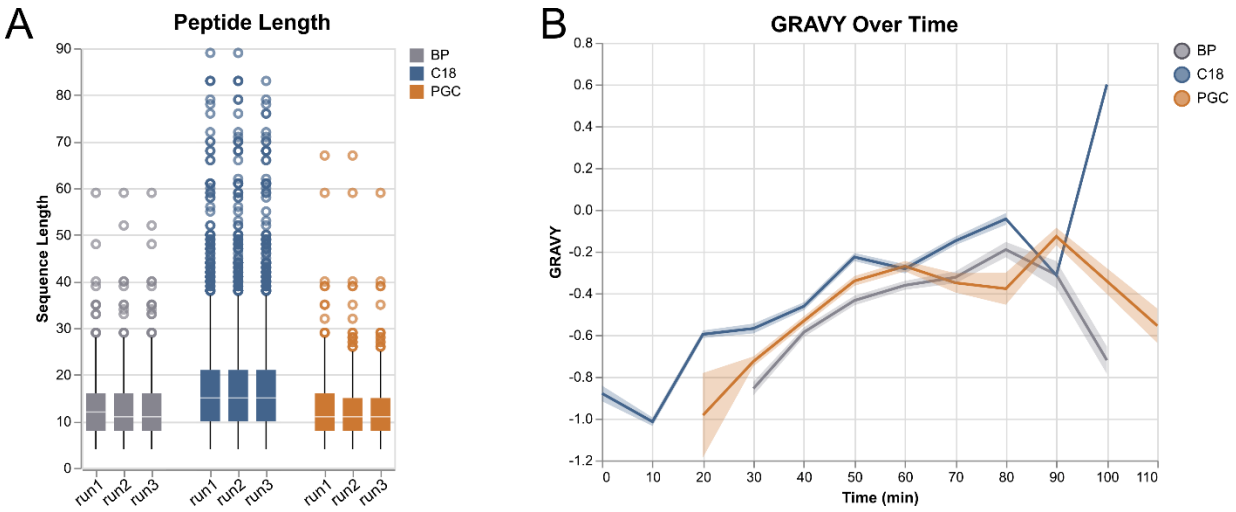


Figure S2.5 Peptide character across separations.

Our initial proteomic investigations revealed that the overall character of the peptide was the primary factor causing differential identification of peptides across each separation regime. These **A**) peptide length and **B**) hydropathy observations were further verified on follow-up triplicate analysis of LNCaP cell line digests. Error bands on plot **B** represent the 95% confidence interval. All triplicates were collected at 45°C.

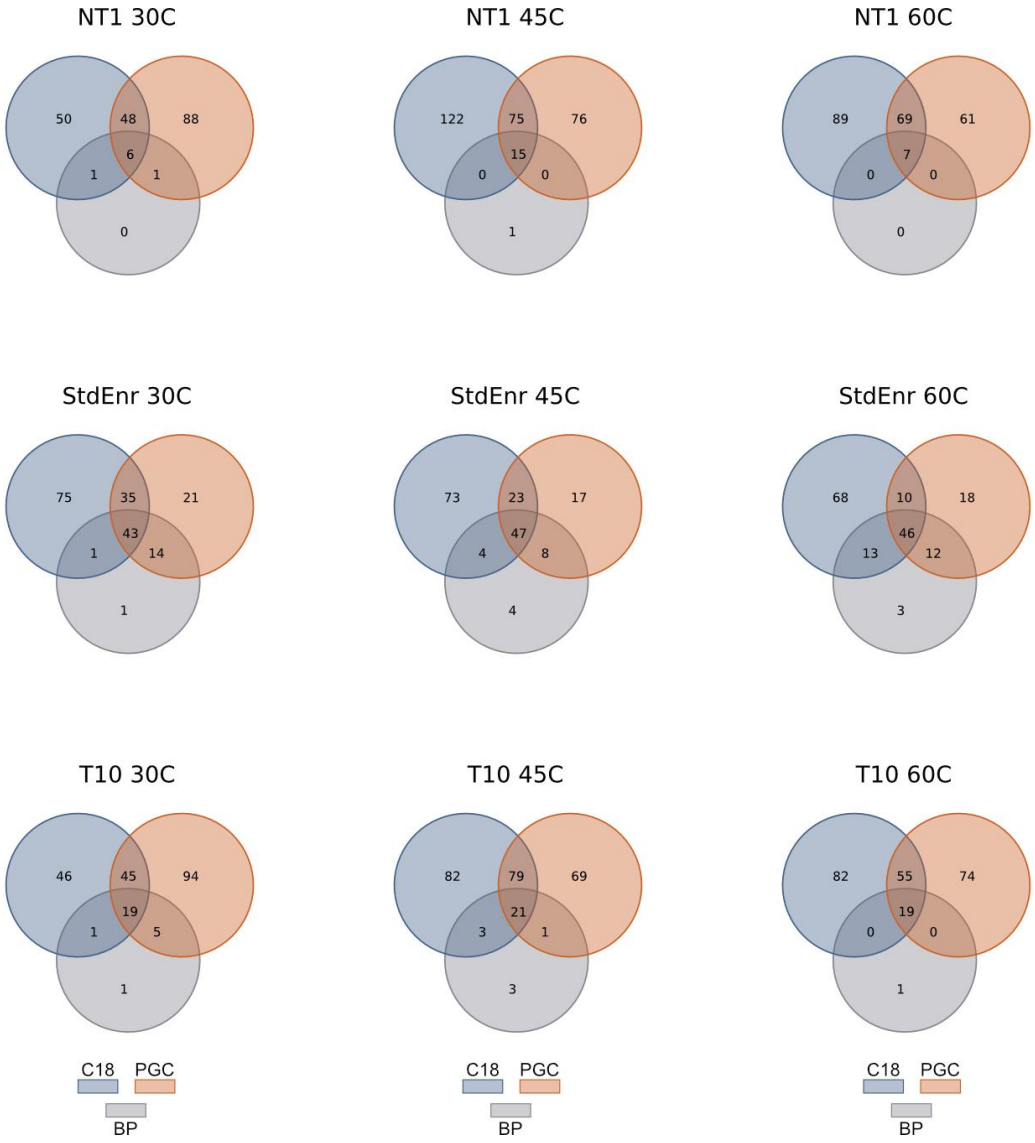


Figure S2.6 Glycopeptide overlap when column is varied.

Venn diagram of unique glycopeptides identified across all experiments, shown. Glycopeptides are stratified by sample and by column. Across all experiments, the largest quantities of unique glycopeptides are found in RPLC and PGC separations, indicating that complementary retention mechanism is the driving unique identification rates. ‘StdEnr’ represents the standard glycopeptide mixture discussed in the main text.

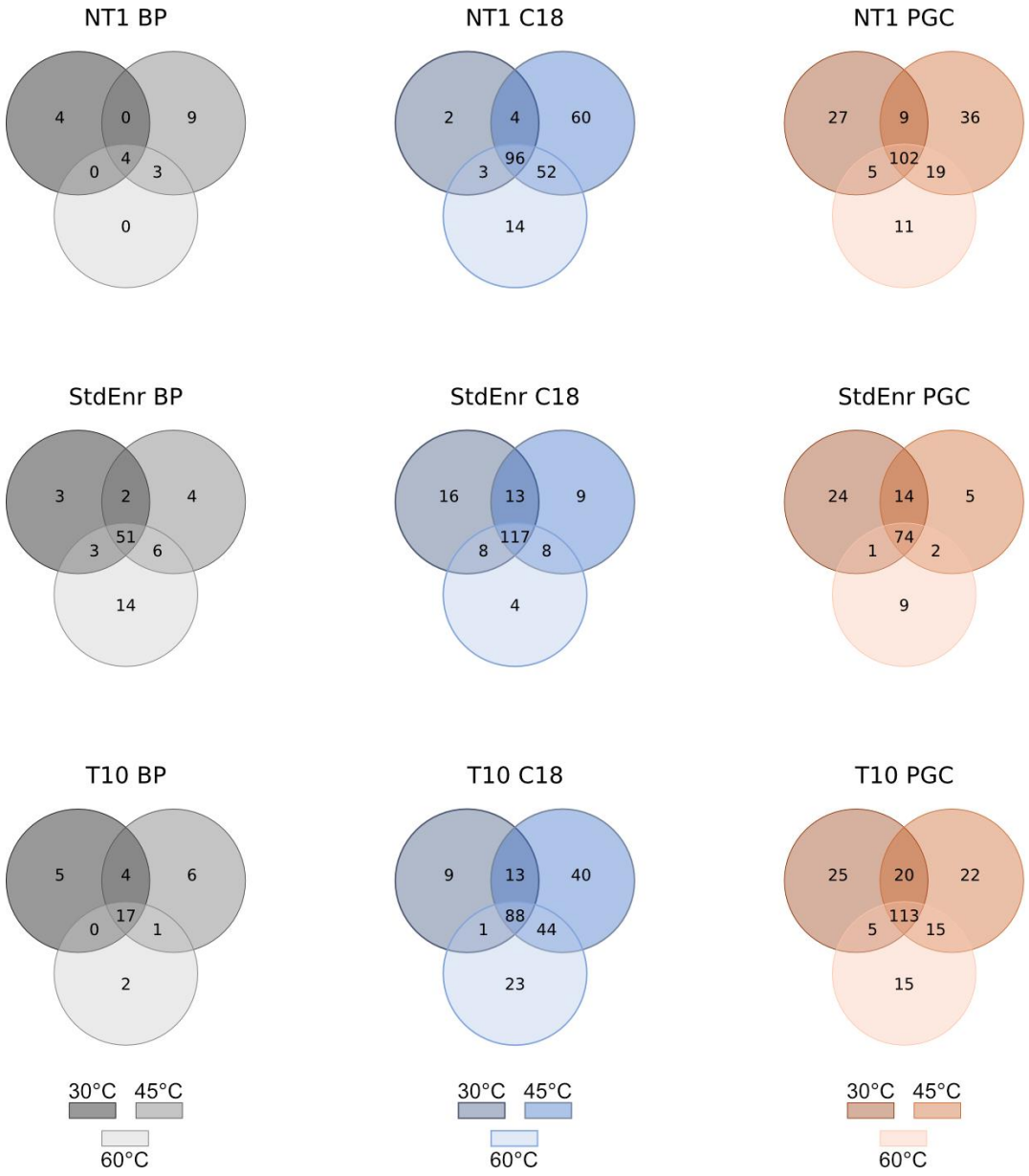


Figure S2.7 Glycopeptide overlap when temperature is varied.

Venn diagram of unique peptides identified across all experiments, illustrating the effects of temperature on glycopeptide identification. Glycopeptides are stratified by sample and by column. ‘StdEnr’ represents the standard glycopeptide mixture discussed in the main text.

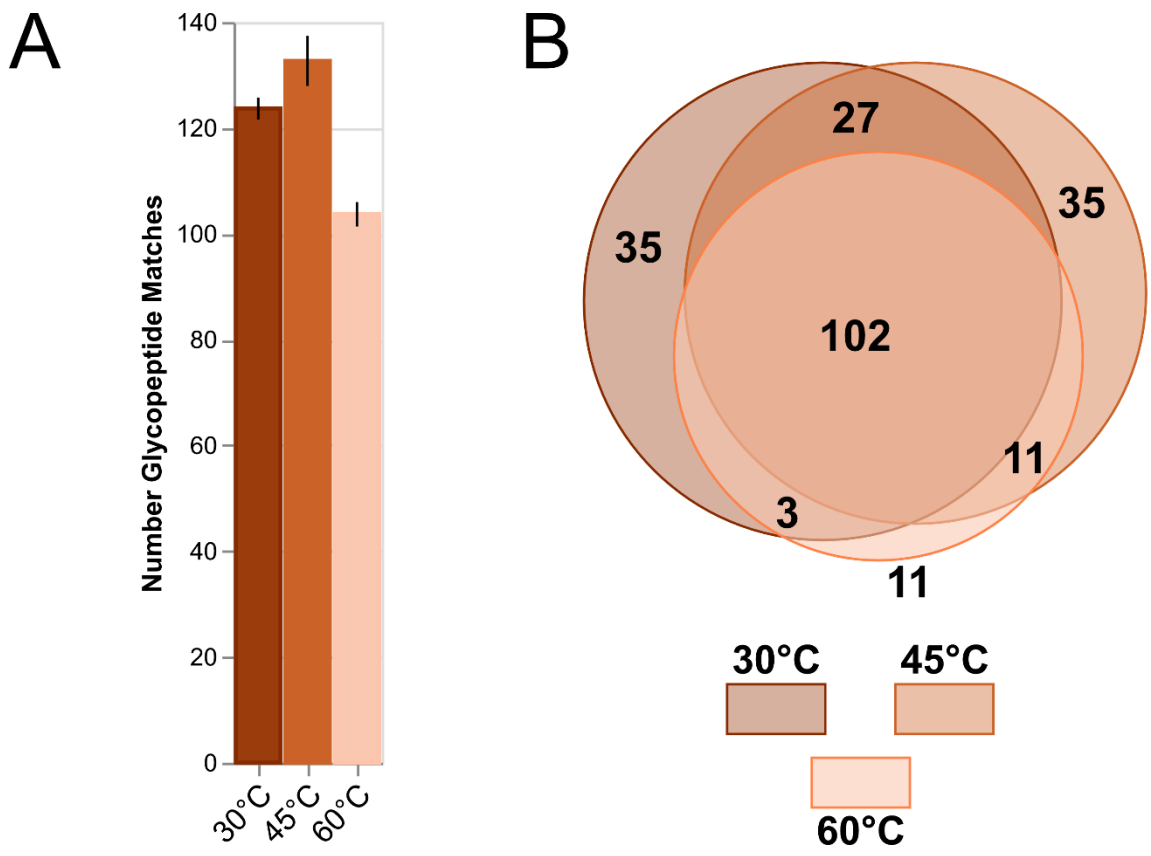


Figure S2.8 Triplicate confirmation of glycopeptide identifications through PGC.

The data presented within the main text allowed us to correlate differences in column temperature with differential retention an identification of glycopeptides. Reanalyzing the standard glycopeptide mixture in triplicate and using PGC separations as a representative, we continue to observe the same trends noted in the main body. **A**) As temperature increases, the overall number of glycopeptide deviates, continuing to display increased identifications at 45°C, and showing the greatest disparity at 60°C. Error bars represent standard deviation, n=3. **B**) Comparing the overlap of identified peptides (those identified in at least two runs at a given temperature), we again observe meaningful quantities of unique species identified at each temperature.

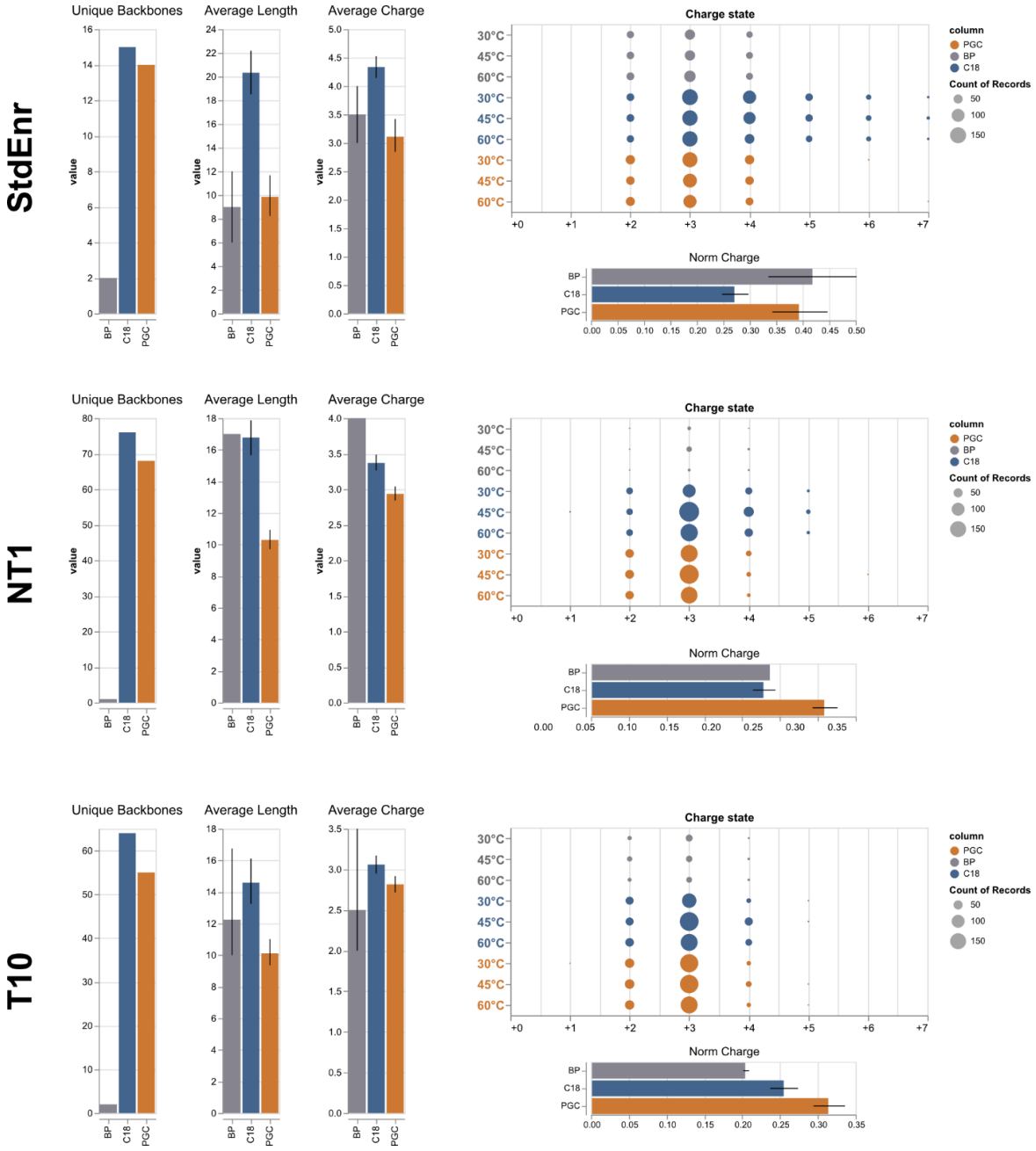


Figure S2.9 Comparison of glycopeptide backbones across columns.

Bar and dot plots of unique peptides identified across all experiments, illustrating the effects of temperature on glycopeptide identification. Glycopeptides are stratified by sample and by column. Error bars: 95% confidence inteval, *n*=3 (each unique sample). 'StdEnr' represents the standard glycopeptide mixture discussed in the main text.

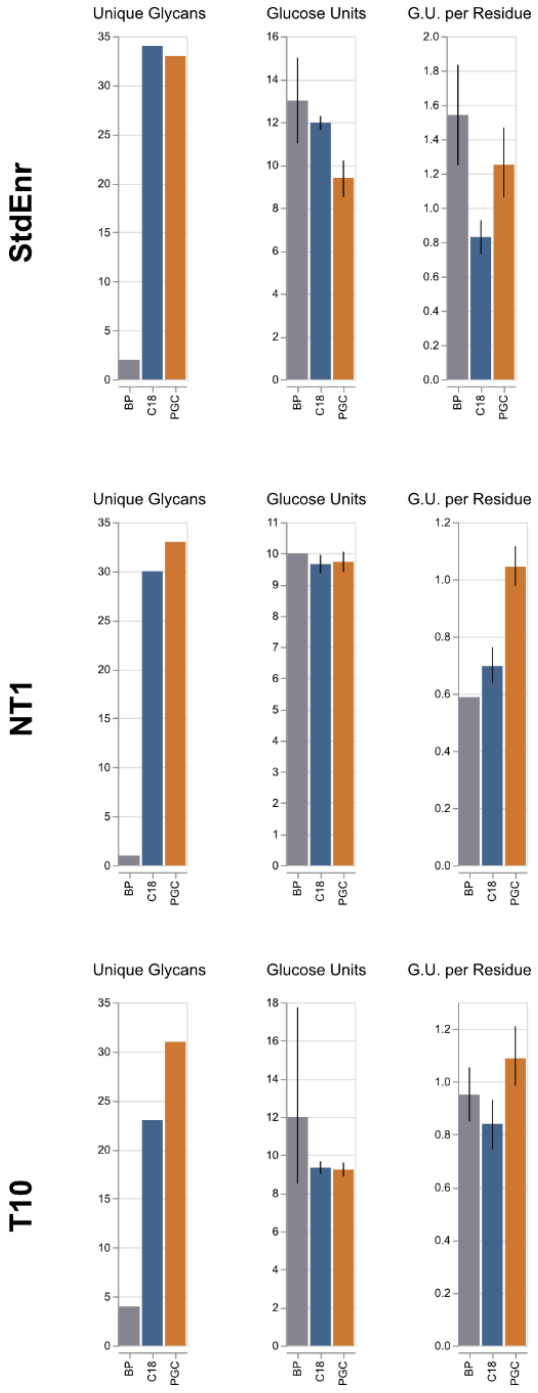
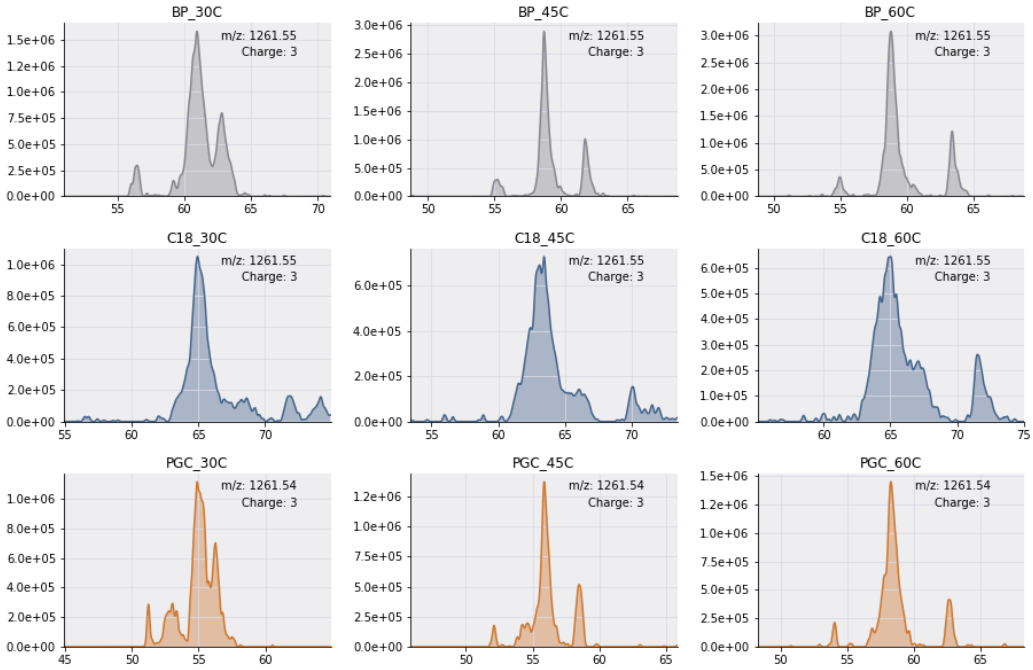


Figure S2.10 Comparison of glycopeptide modifications across columns.

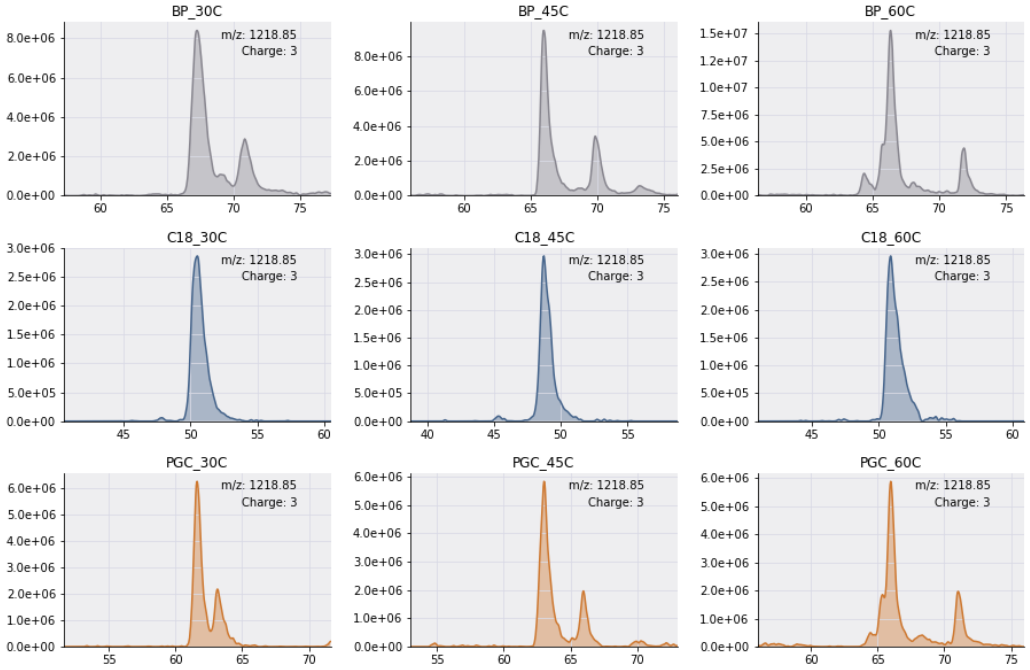
Distribution of number of unique glycans, glucose units per glycan and glucose units per amino acid residue from glycopeptides identified in each sample. Error bars: 95% confidence inteval, *n*=3

(each unique sample). ‘StdEnr’ represents the standard glycopeptide mixture discussed in the main text.

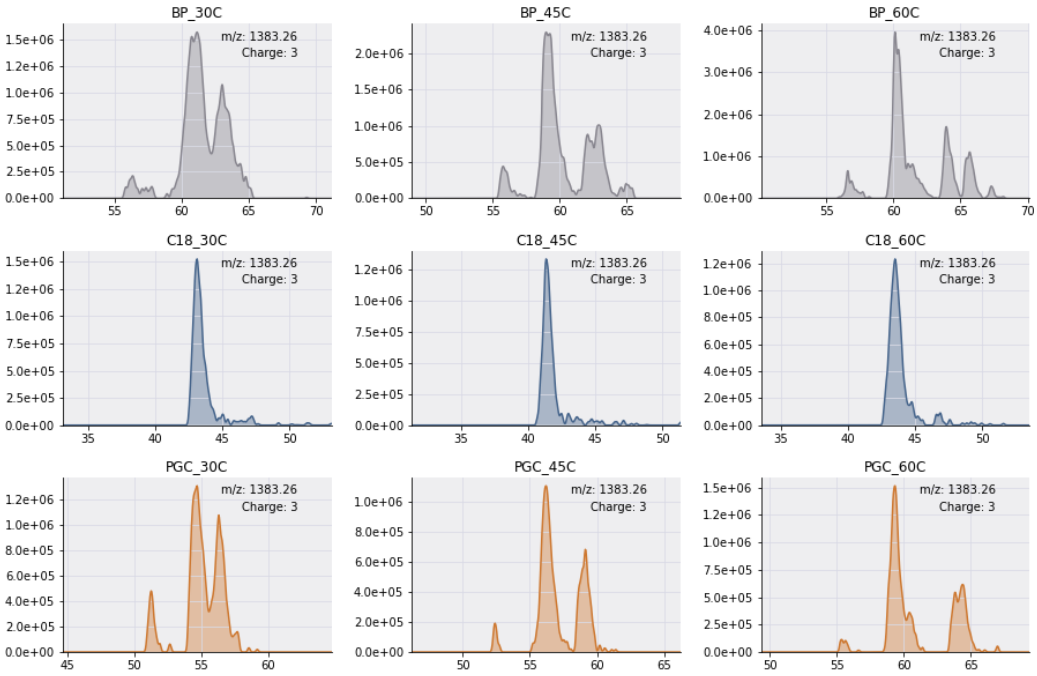
KLCPDCPLLAPLNDSR + HexNAc(4)Hex(5)NeuAc(1)



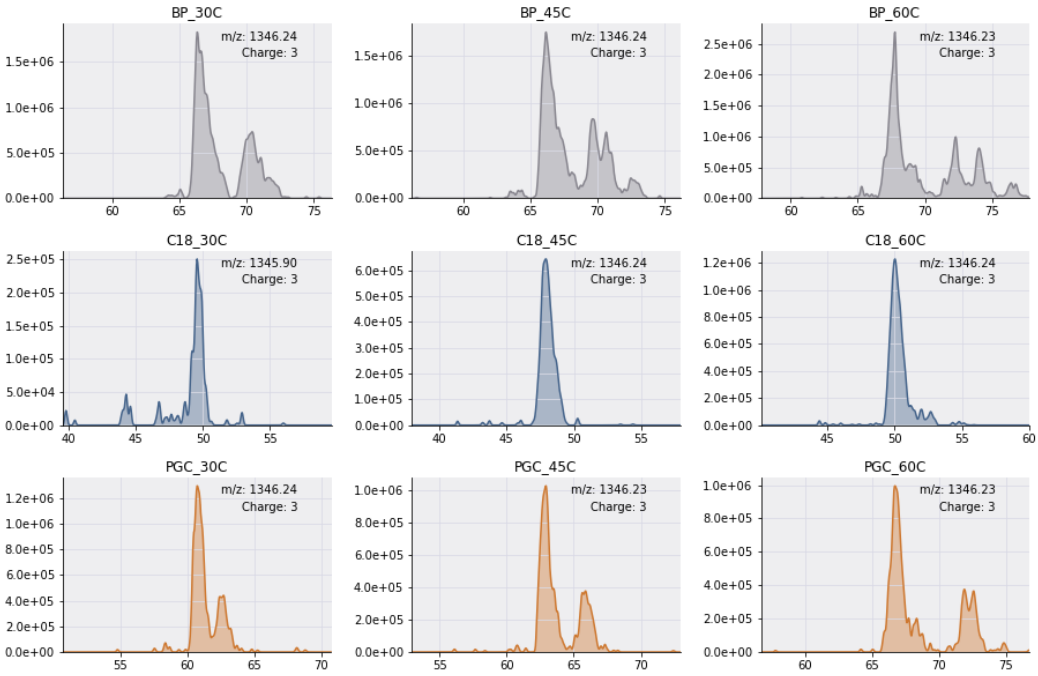
LCPDCPLLAPLNDSR + HexNAc(4)Hex(5)NeuAc(1)



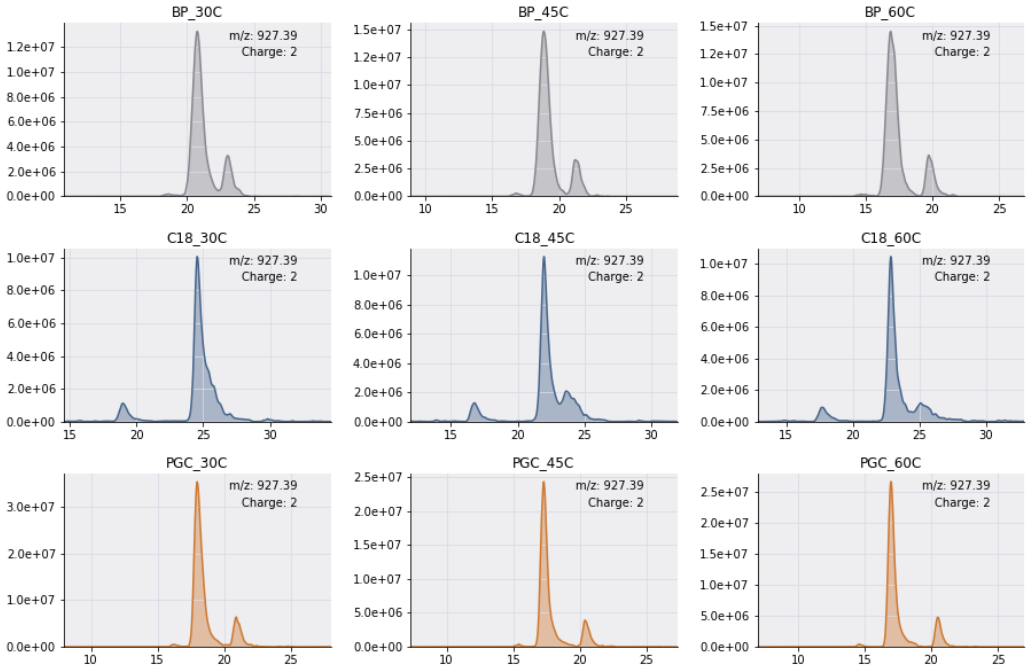
KLCPDCPLLAPLNDSR + HexNAc(5)Hex(6)NeuAc(1)



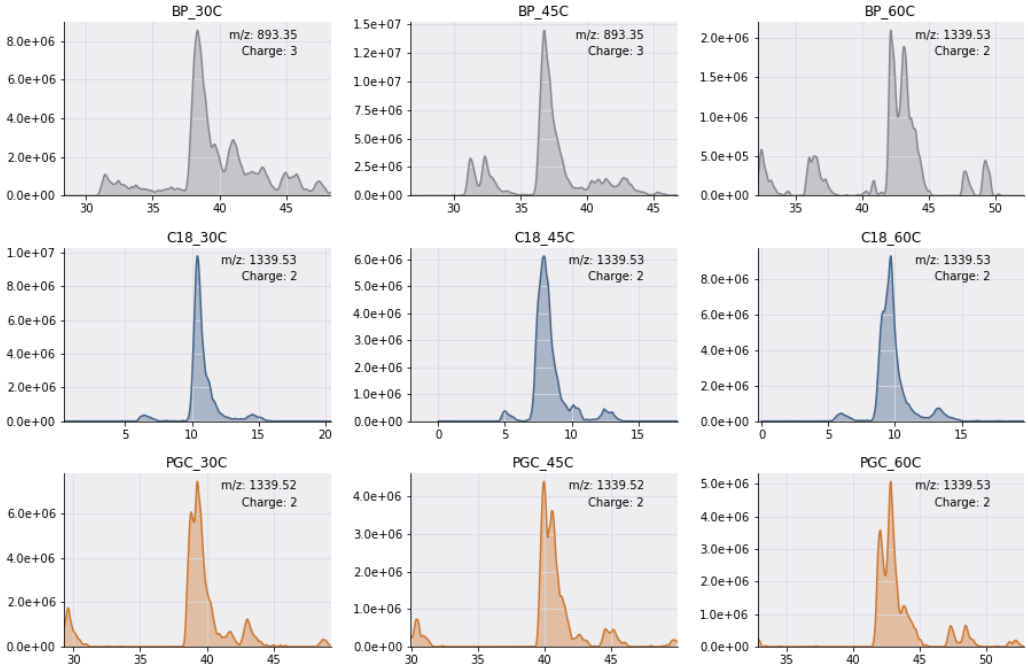
LCPDCPLLAPLNDSR + HexNAc(5)Hex(6)NeuGc(1)



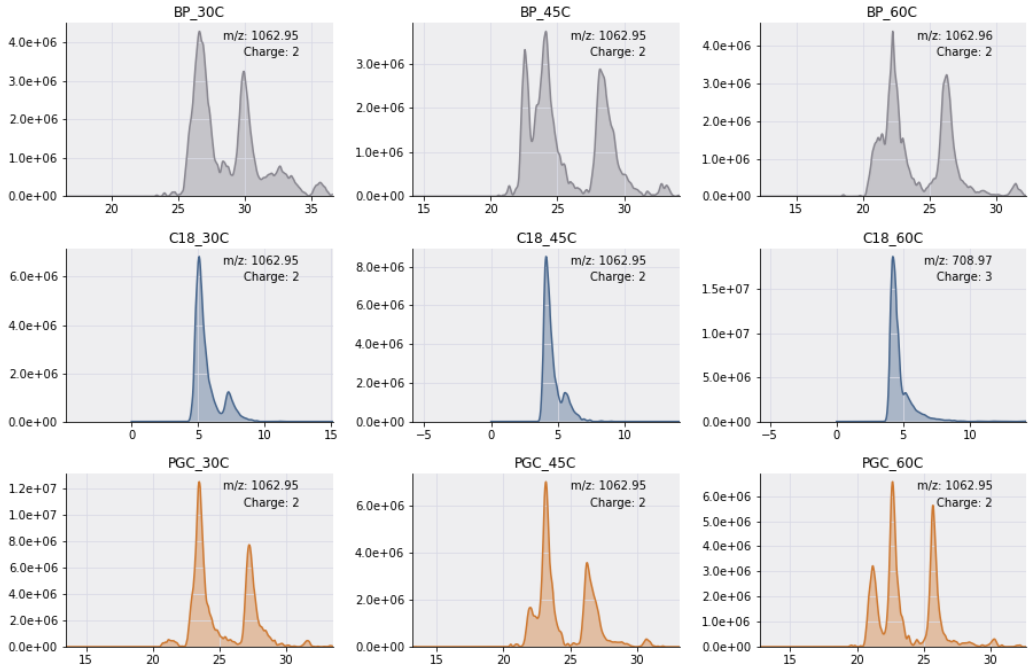
NLTK + HexNAc(2)Hex(6)



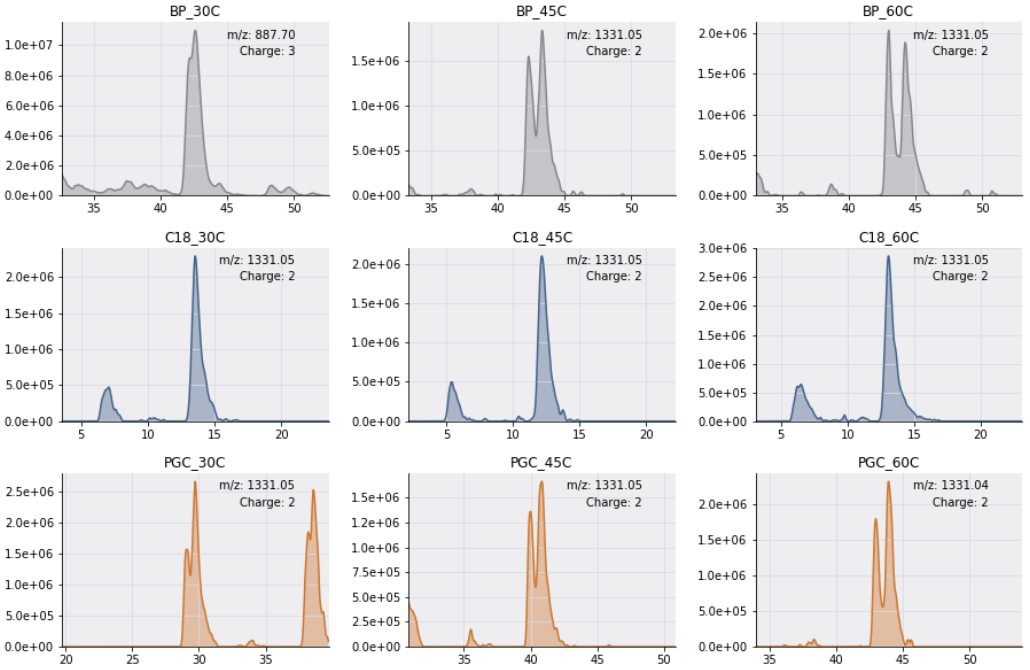
NPEYNK + HexNAc(4)Hex(5)NeuAc(1)



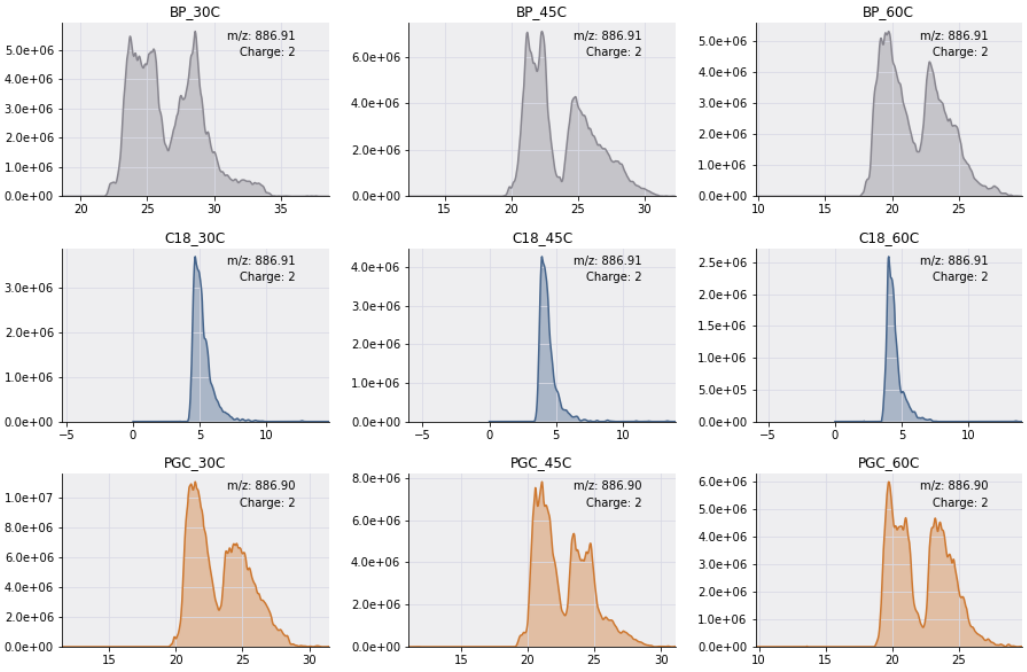
NLTKDR + HexNAc(2)Hex(6)



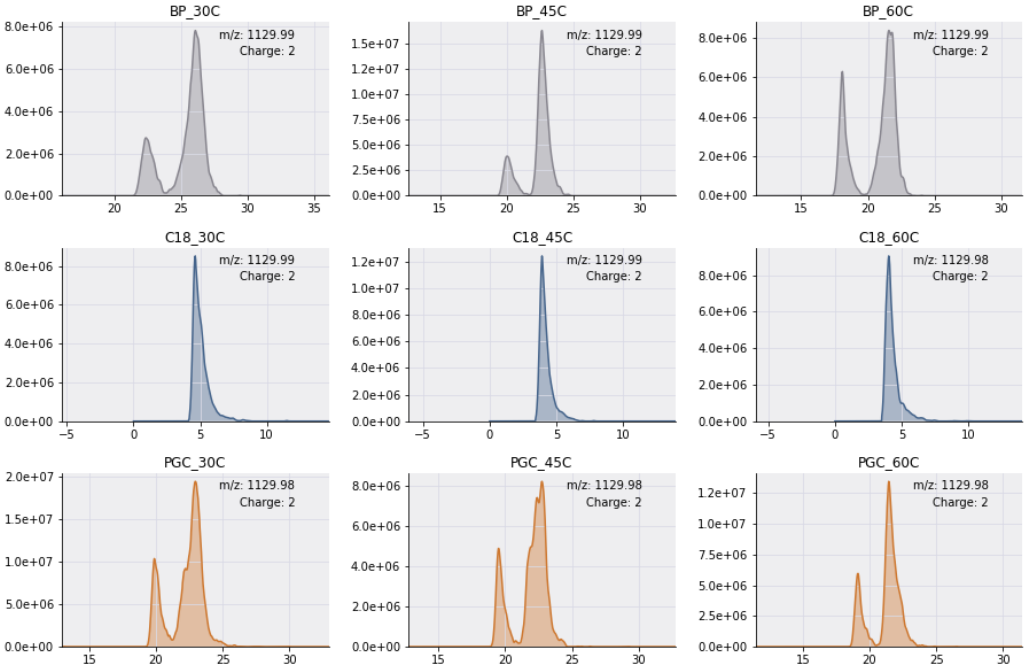
QNGTLSK + HexNAc(4)Hex(5)NeuAc(1)



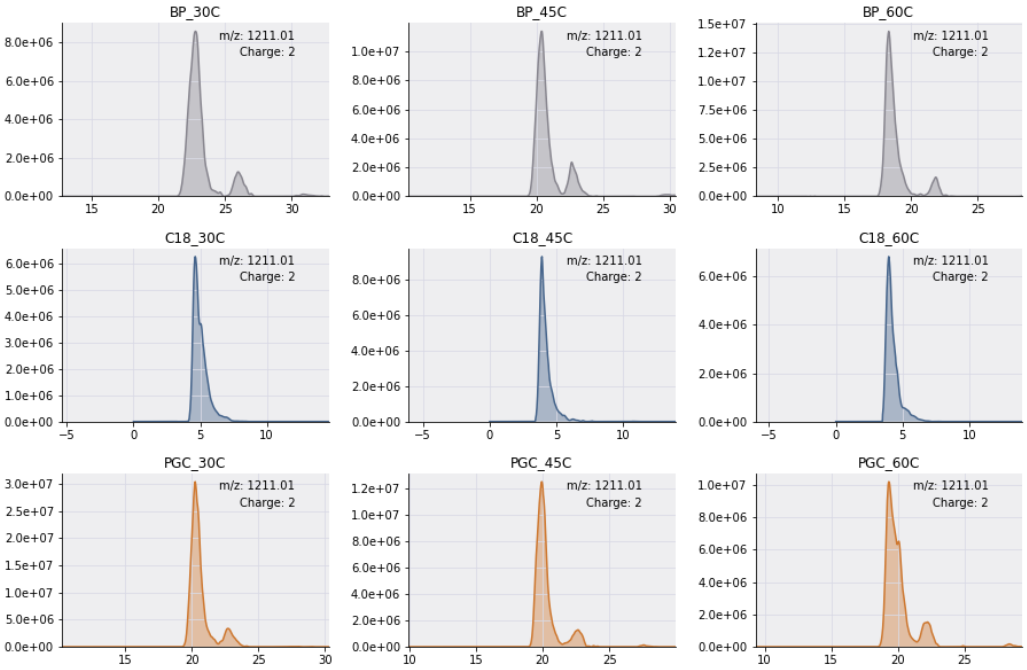
SRNLTK + HexNAc(2)Hex(4)



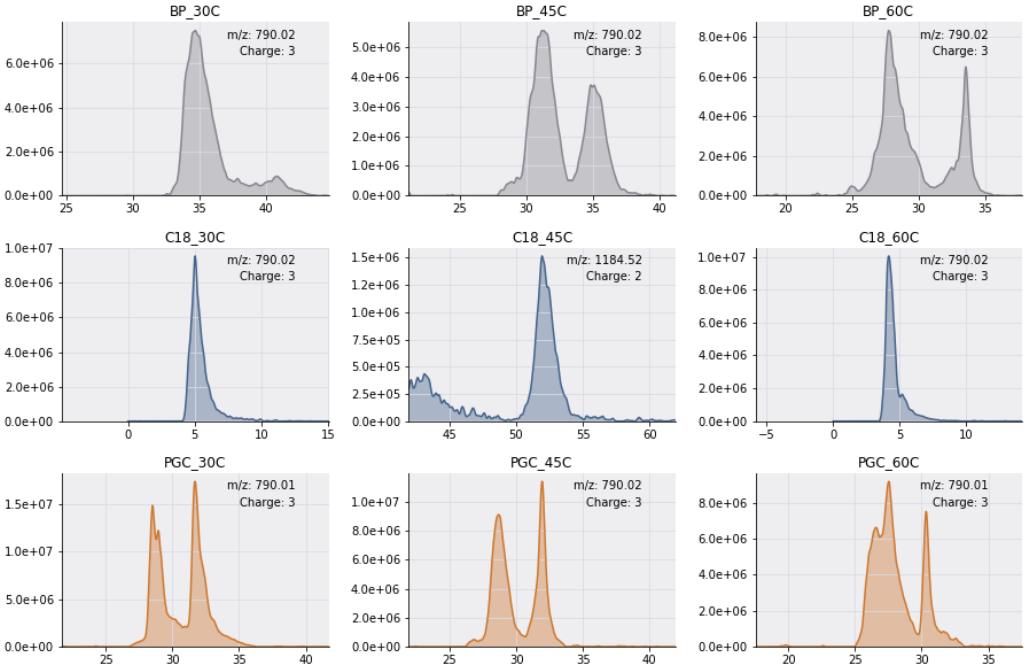
SRNLTK + HexNAc(2)Hex(7)



SRNLTK + HexNAc(2)Hex(8)



SRNLTKDR + HexNAc(2)Hex(6)



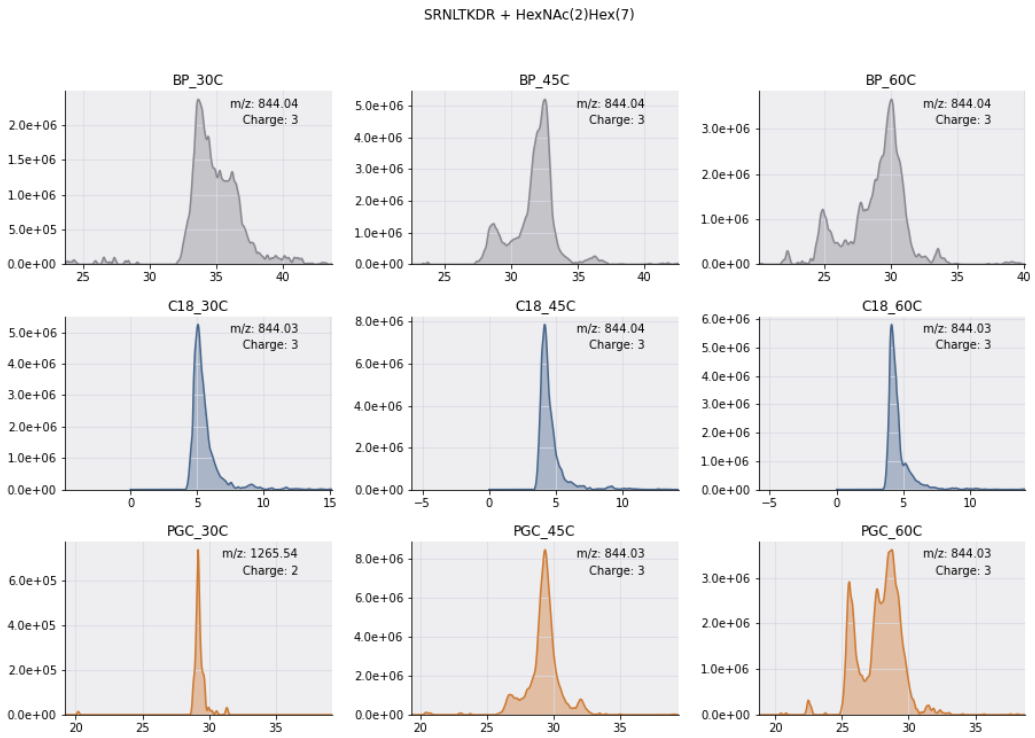
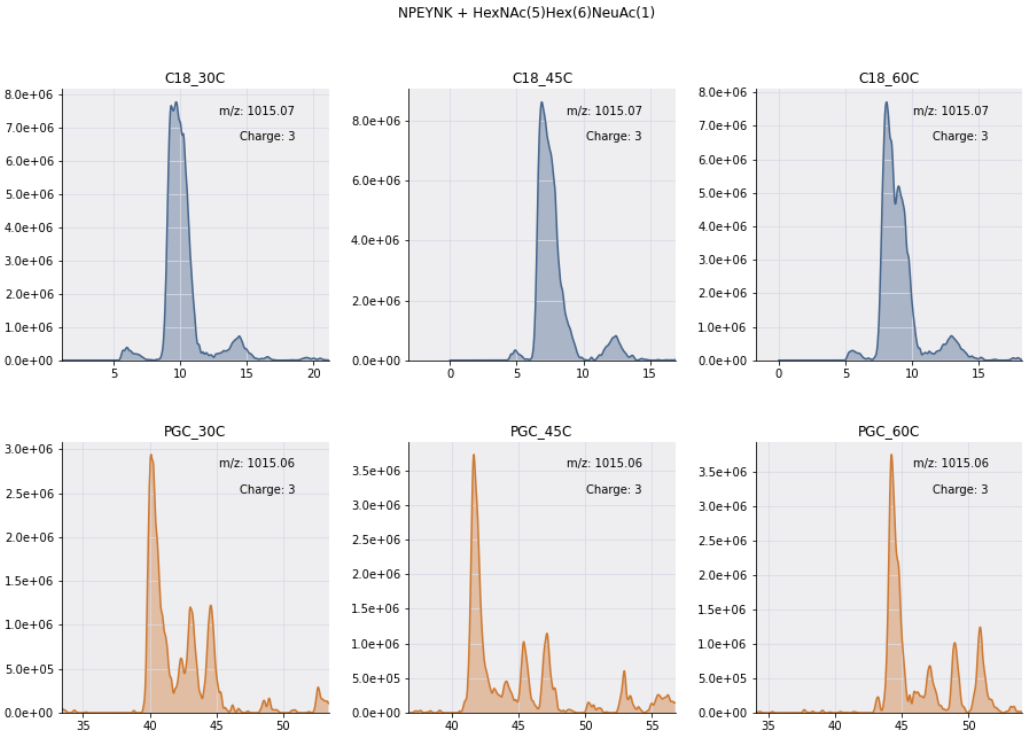
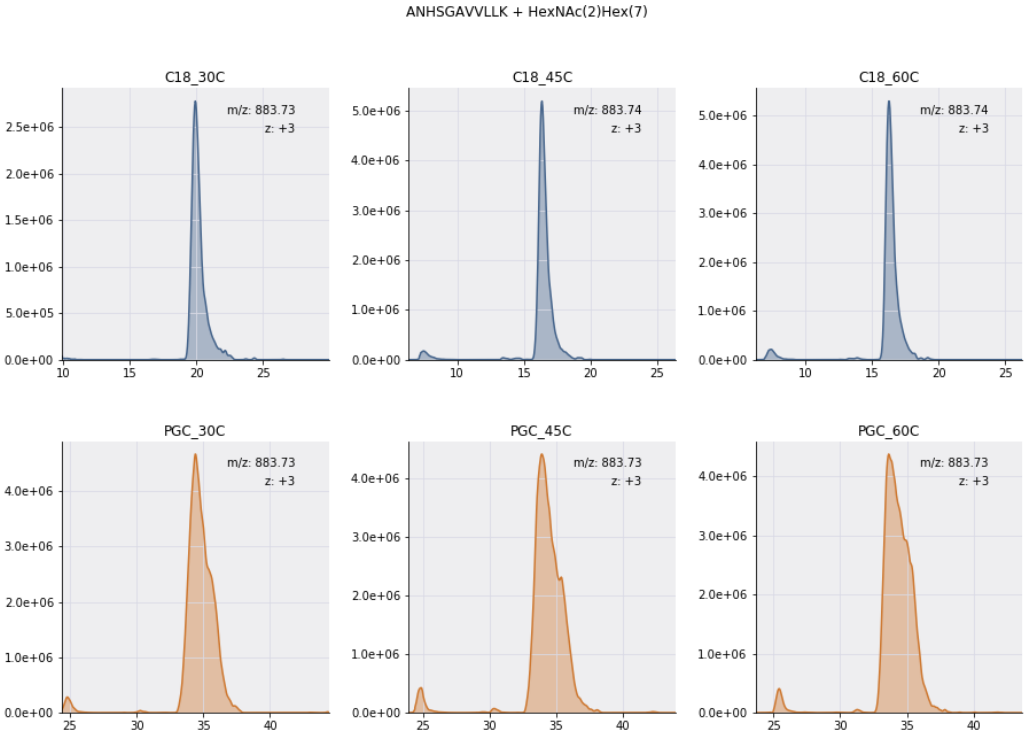


Figure S2.11 Extract ion chromatograms of species displaying altered resolution at higher temperatures.

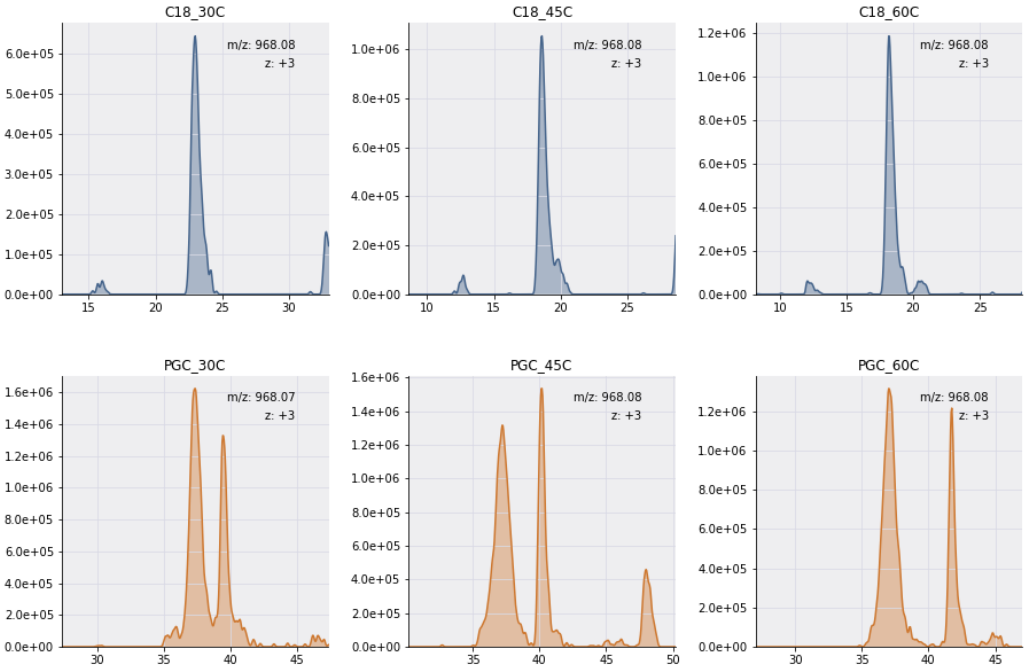
These figures contain EICs from each sample analyzed and only species identified in every individual run (i.e., BP at all temperatures, C18 at all temperatures and PGC at all temperatures) were extracted. After extracting EICs of all species meeting this criterion (mass tolerance = 10ppm), plots were filtered manually to display only those where liquid-phase resolution was apparent and unambiguous.


NPEYNK + HexNAc(5)Hex(6)NeuAc(1)
C18_30C
m/z: 1015.07
Charge: 3
C18_45C
m/z: 1015.07
Charge: 3
C18_60C
m/z: 1015.07
Charge: 3
PGC_30C
m/z: 1015.06
Charge: 3
PGC_45C
m/z: 1015.06
Charge: 3
PGC_60C
m/z: 1015.06
Charge: 3

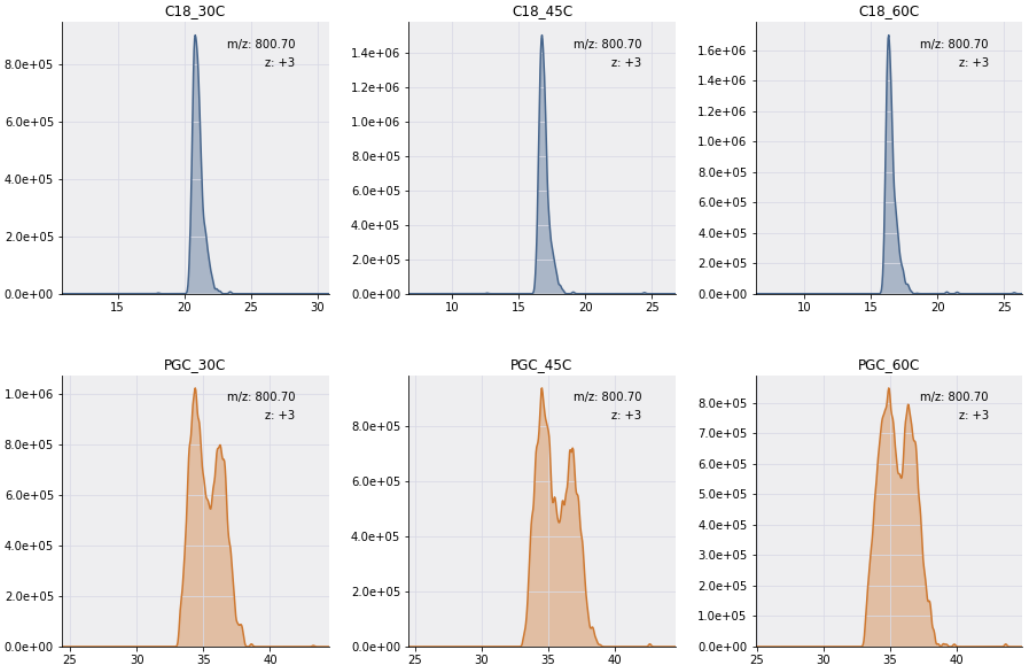

ANHSGAVVLLK + HexNAc(2)Hex(7)
C18_30C
m/z: 883.73
z: +3
C18_45C
m/z: 883.74
z: +3
C18_60C
m/z: 883.74
z: +3
PGC_30C
m/z: 883.73
z: +3
PGC_45C
m/z: 883.73
z: +3
PGC_60C
m/z: 883.73
z: +3

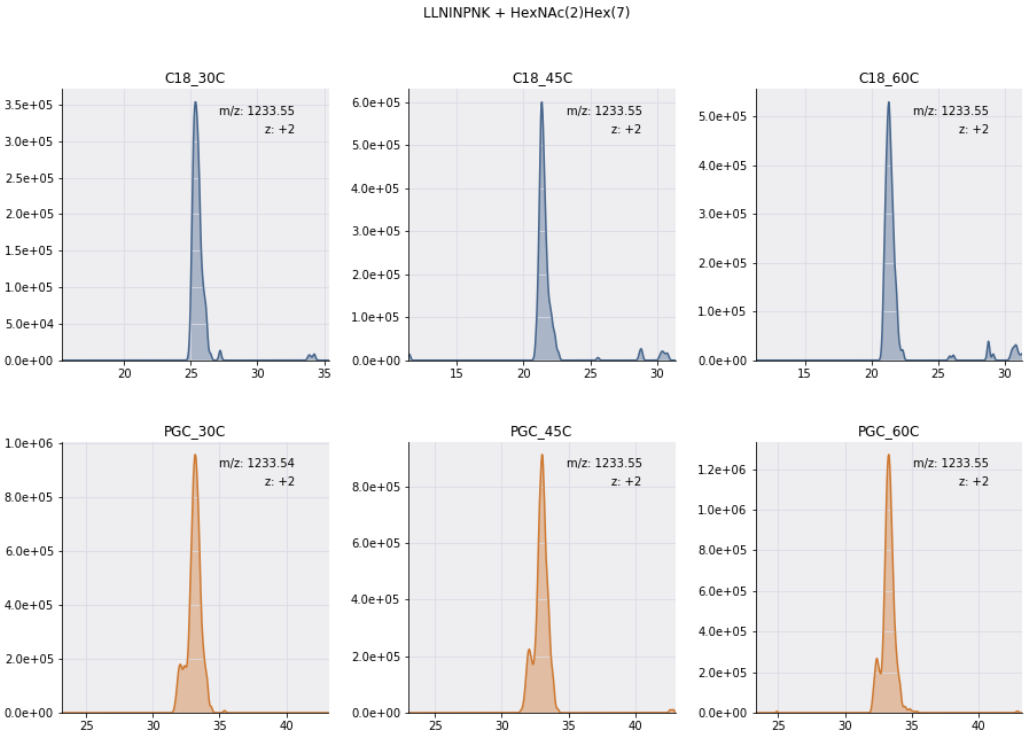
.

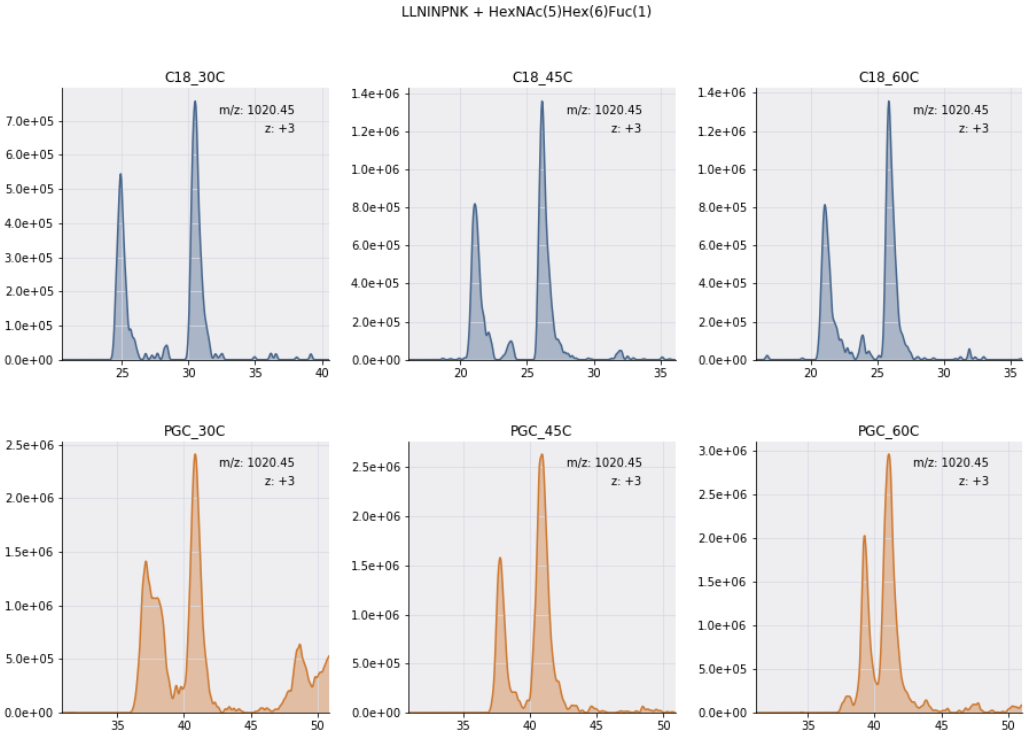
HTNMNLTQLK + HexNAc(2)Hex(8)



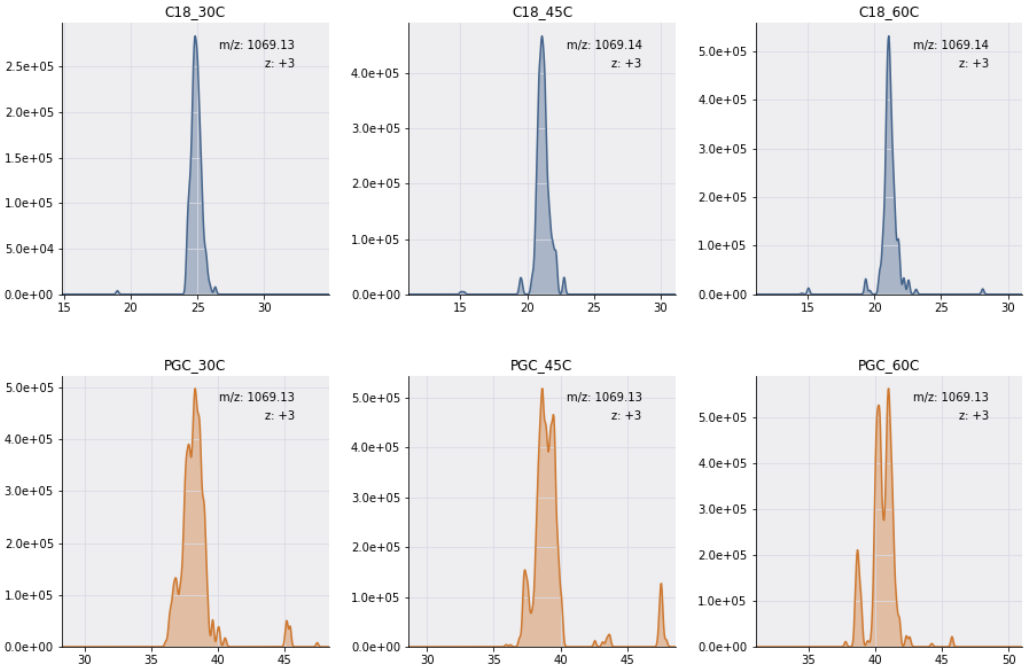
LKPLFNK + HexNAc(2)Hex(7)



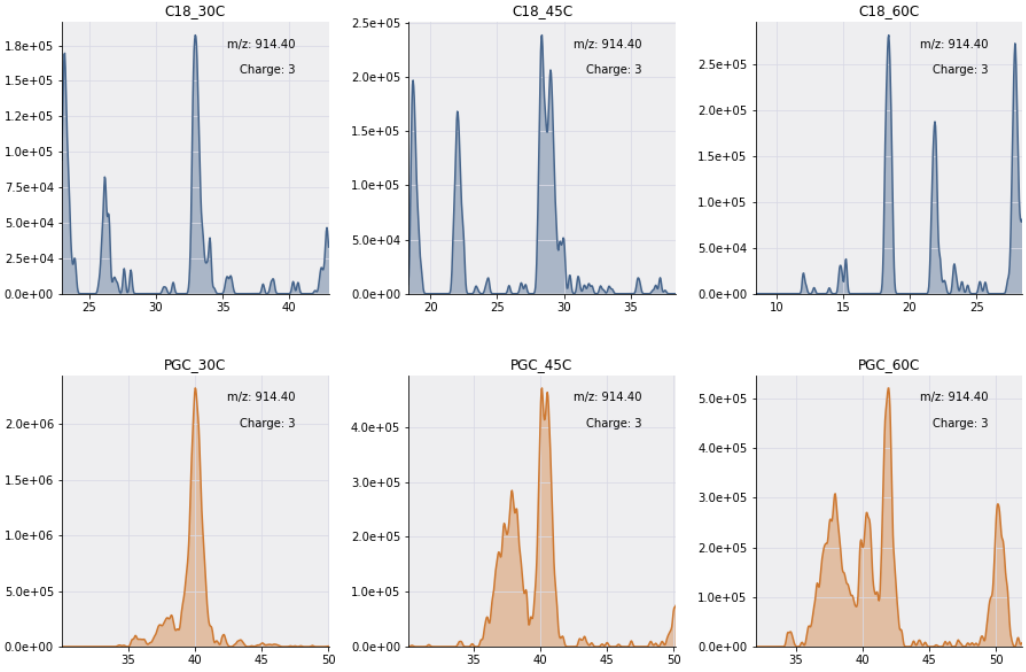

LLNINPNK + HexNAc(2)Hex(7)
C18_30C
m/z: 1233.55
z: +2
C18_45C
m/z: 1233.55
z: +2
C18_60C
m/z: 1233.55
z: +2
PGC_30C
m/z: 1233.54
z: +2
PGC_45C
m/z: 1233.55
z: +2
PGC_60C
m/z: 1233.55
z: +2


LLNINPNK + HexNAc(5)Hex(6)Fuc(1)
C18_30C
m/z: 1020.45
z: +3
C18_45C
m/z: 1020.45
z: +3
C18_60C
m/z: 1020.45
z: +3
PGC_30C
m/z: 1020.45
z: +3
PGC_45C
m/z: 1020.45
z: +3
PGC_60C
m/z: 1020.45
z: +3

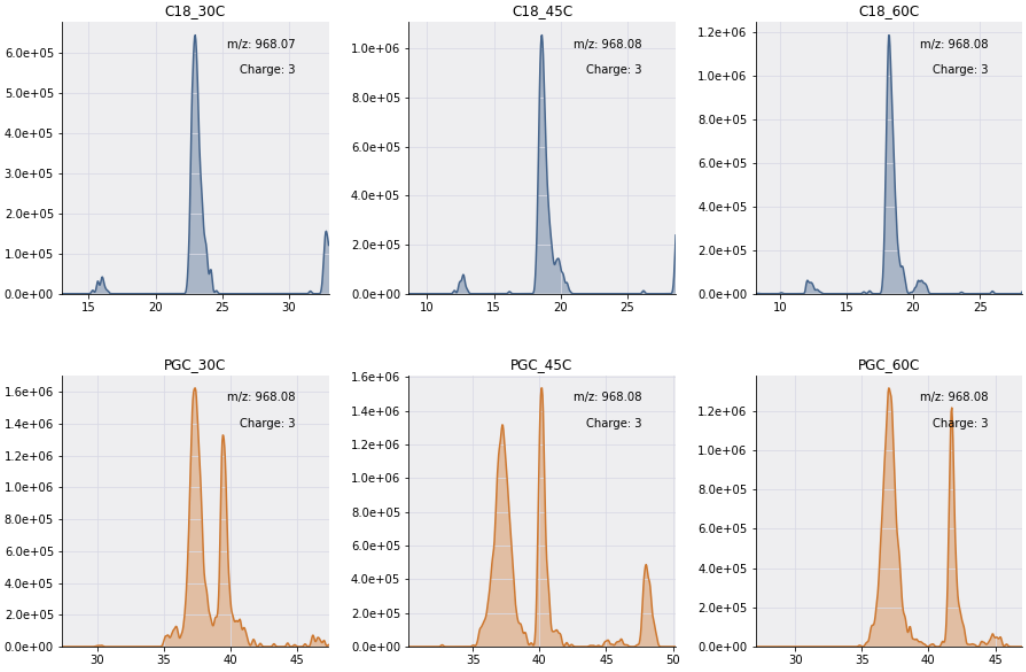
LLNINPNK + HexNAc(5)Hex(6)Fuc(2)



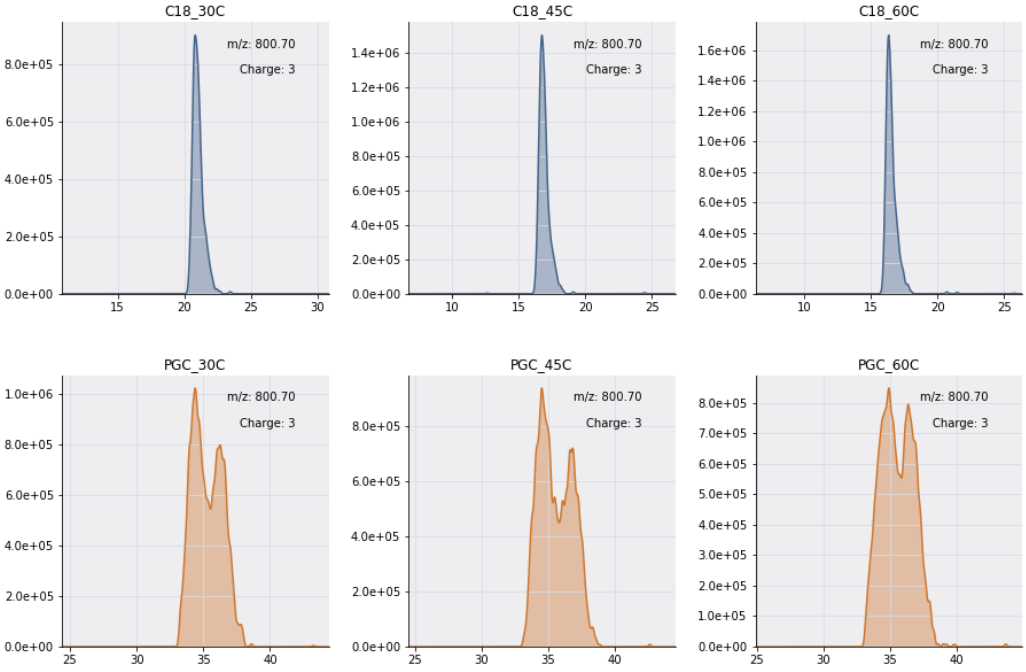
GHTLTLNFTR + HexNAc(3)Hex(6)



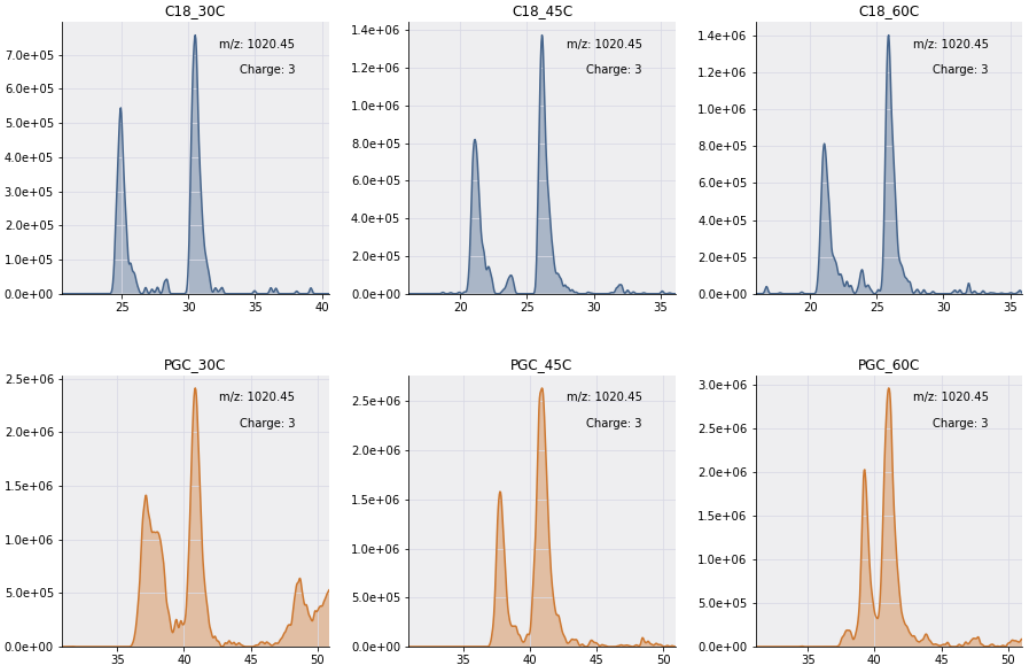
HTNMNLTQLK + HexNAc(2)Hex(8)



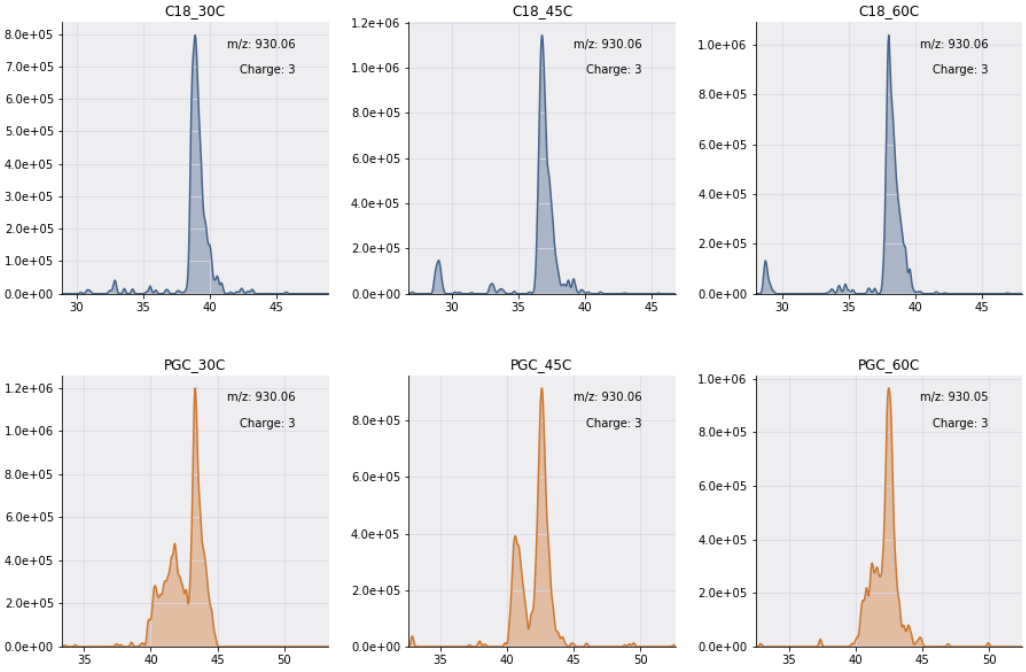
LKPLFNK + HexNAc(2)Hex(7)



LLNINPNK + HexNAc(5)Hex(6)Fuc(1)



VNVTVEDER + HexNAc(3)Hex(6)Fuc(1)



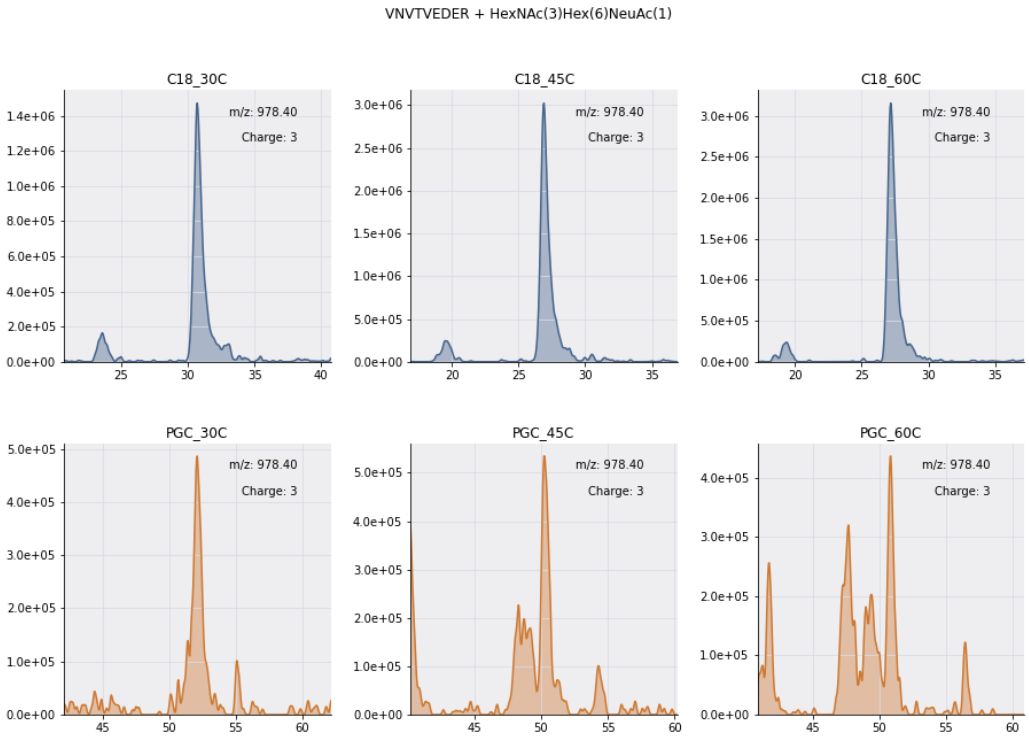


Figure S2.12 Extract ion chromatograms, C18 and PGC only.

This figure contains EICs from each sample analyzed and only species identified in every C18 and PGC run (i.e., C18 at all temperatures and PGC at all temperatures) were extracted. After extracting EICs of all species meeting this criterion (mass accuracy=10ppm), plots were filtered manually to display only those where liquid-phase resolution was apparent and unambiguous.

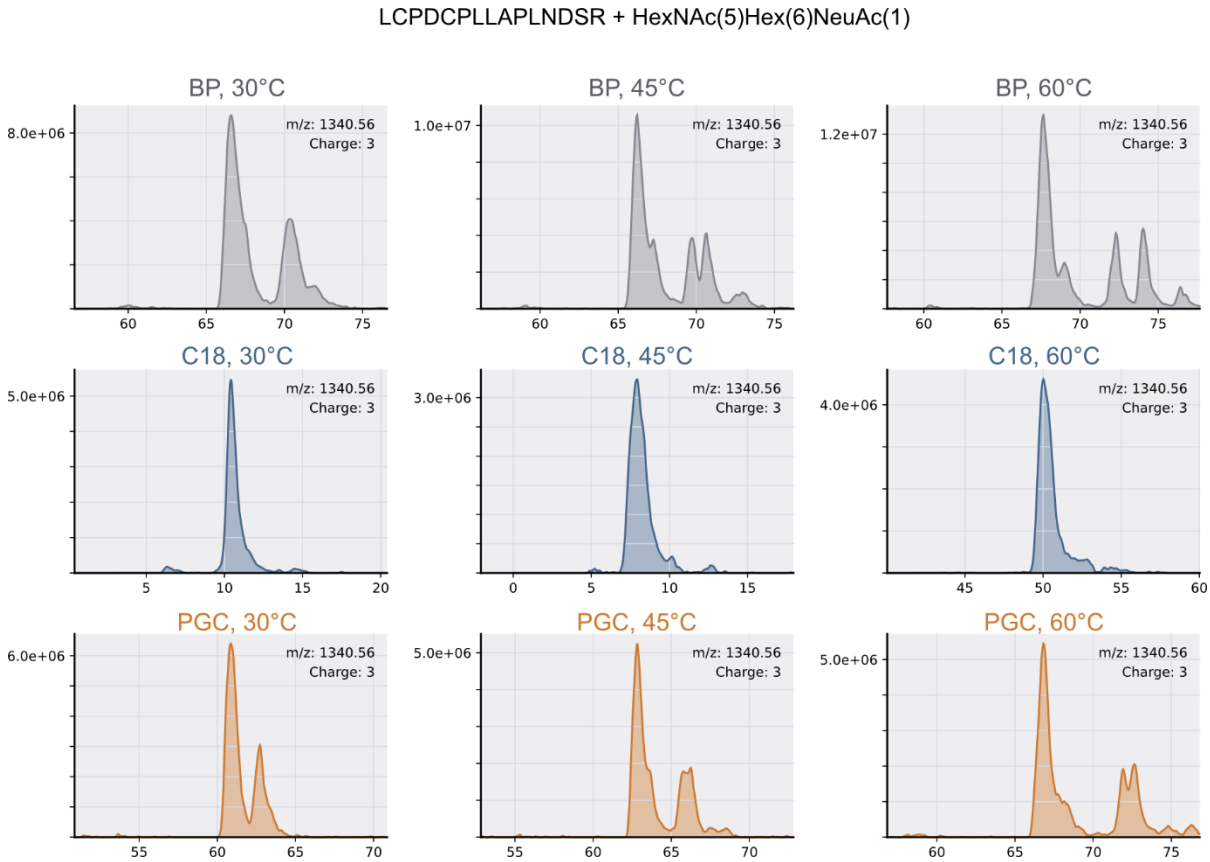


Figure S2.13 Representative glycopeptide extract ion chromatograms displaying greater resolution in BP separations than in PGC.

Representative extract ion chromatogram of a glycopeptide identified in all columns at all temperatures, displaying greater resolution in BP separations than in PGC. This observation indicates a long PGC stationary phase is not necessary to provide liquid-phase separation and a shorter (L=15cm) PGC stationary component could be coupled with a high-resolution RPLC component to provide better glycoproteome profiling. Masses are extracted to a 10ppm error.

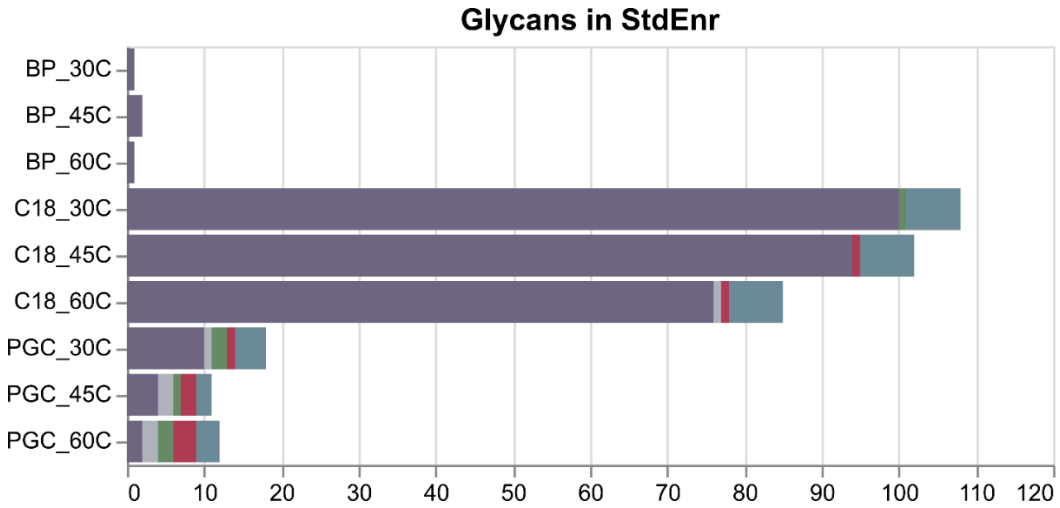


Figure S2.14 Distribution of glycan modifications identified in mixture of digested glycoprotein standards.

The prevalence of sialoglycans stems primarily from the high proportion of glycopeptides identified from bovine fetuin and alpha-1-acid glycoprotein. These identifications contrast those seen in analyses of the complex samples, indicating there is no bias towards glycans of a particular composition within any separation regime. As well, analysis of this sialoglycan-rich sample highlights the observation that higher temperatures provide lower identification rates of sialylated glycans. Though more experimentation is needed, one explanation may be the additional energy imparted on labile glycan bonds may force premature dissociation of glycan bonds. ‘StdEnr’ represents the standard glycopeptide mixture discussed in the main text.

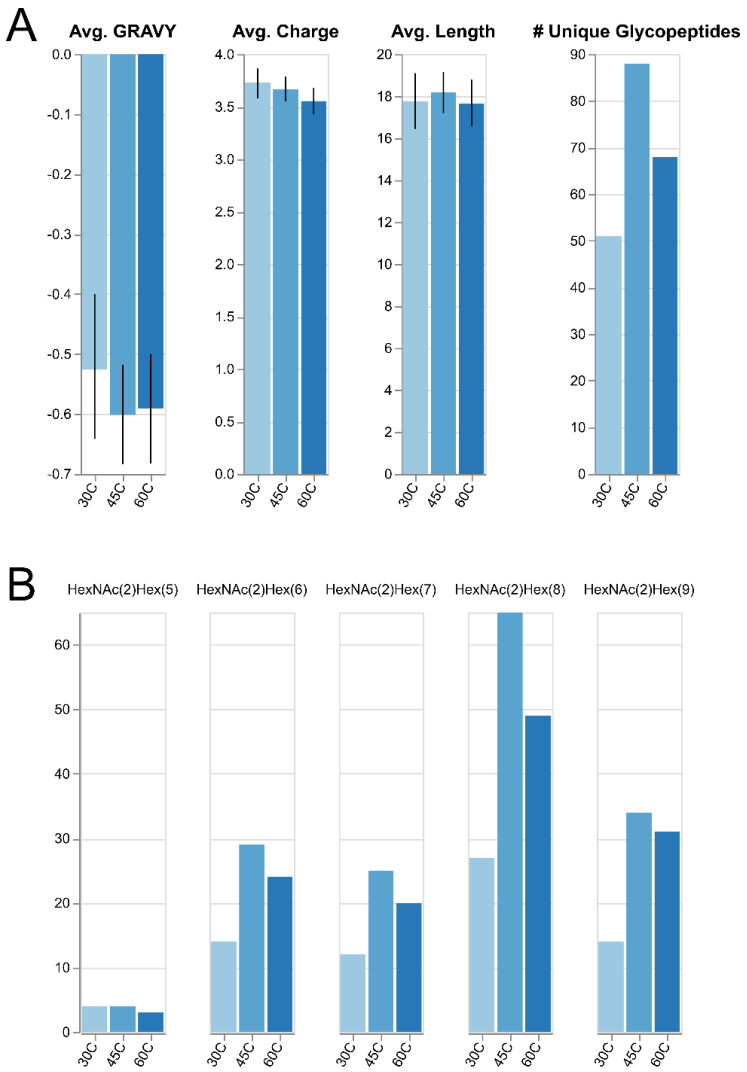


Figure S2.15 Underlying trends of all glycopeptides identified following RPLC separation.

A) As 45°C displayed higher numbers of glycopeptide identifications than any other temperature, we evaluated whether any of the peptide-level differences were a cause of the improved identification rates. As seen, there are only marginal differences between the average charge and length, while hydrophilicity of identified glycopeptides changes substantially at 45°C (error bars = 95% c.i., *n=3*, each sample). B) In order to evaluate if our results were in response to preferential retention of unique glycan compositions, we divided high mannose glycopeptides according to their glycan moiety and tallied the number of occurrences within identified glycopeptides. As shown, almost all moieties follow the general trend of having lowest identifications at 30°C,

highest at 45°C, with 60°C lying somewhere in the middle. Given that even the elongated Man_9 glycans are observed in higher quantities at 60°C, we suggest that this trend is not due to early dissociation of glycan moieties, which could create false databases matches.

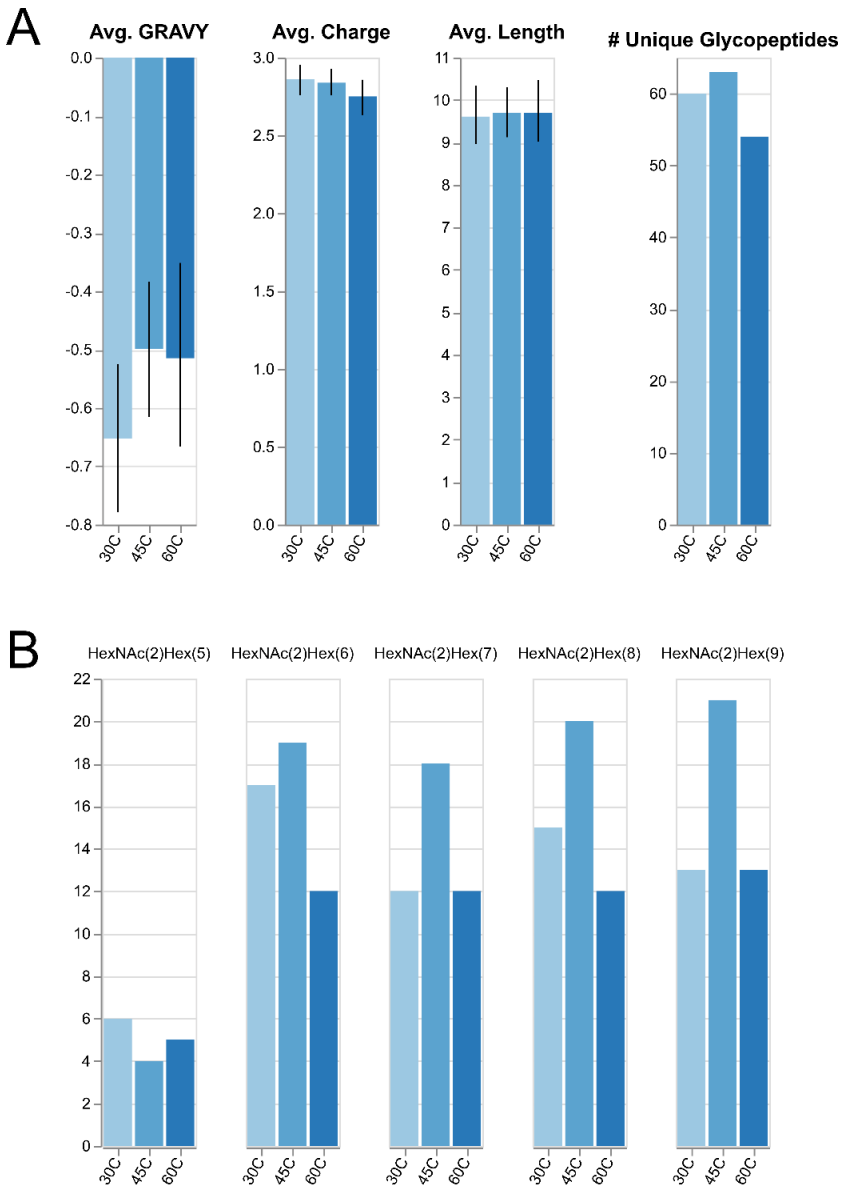


Figure S2.16 Underlying trends of all glycopeptides following PGC separation.

A) Following the same trends seen in RPLC separations (Figure S2.13), there are only slight differences presented at the peptide level. As changes fluctuate when temperatures increase, we posit this is not a cause for the differences in glycopeptide identifications. B) Similar to RPLC separations, all high mannose glycans have highest identification rates at 45°C. Taken together with the data in the main text and the information presented in Figure S2.12, column running temperatures impart a notable effect on peptide/glycopeptide identifications. While both PGC and

RPLC analyses seem to suggest more similarity between 45°C and 60°C than between any two other temperatures, more extensive experimentation is required to fully illuminate this relationship. Future studies should seek to expand the collection of proteins and peptides identified (either through highly fractionated studies or more intensive preparation steps) in order to illuminate the effect of temperature on analytical performance. As well, it would be of interest to include analysis of standards to see whether or not temperature inspires any biotransformation/alteration that could lead to incorrect identifications.

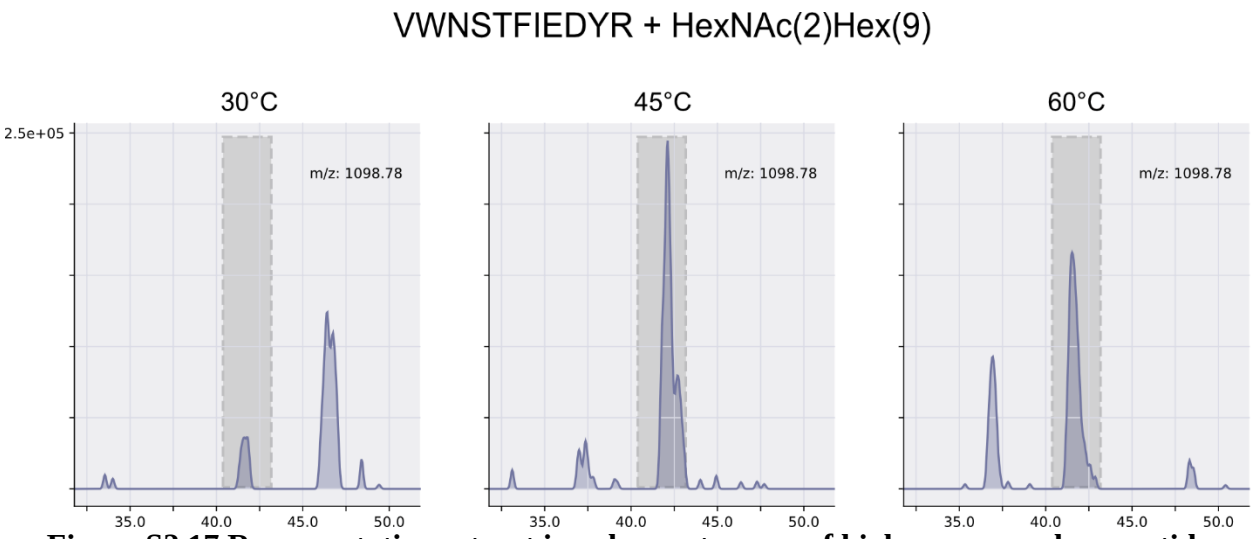


Figure S2.17 Representative extract ion chromatogram of high mannose glycopeptide identified in RPLC separations.

Mimicking the trend seen in identification rates, this species exhibits highest MS^1 reporting signal at 45°C, drastically improved from 30°C experiments. Though signal intensity and peak area cannot be directly compared from experiment to experiment, this trend is shared among numerous glycopeptide species. As such, we feel this observation is directly related to the identification success in our experiments and indicates higher temperatures may provide benefit in glycopeptide analyses.

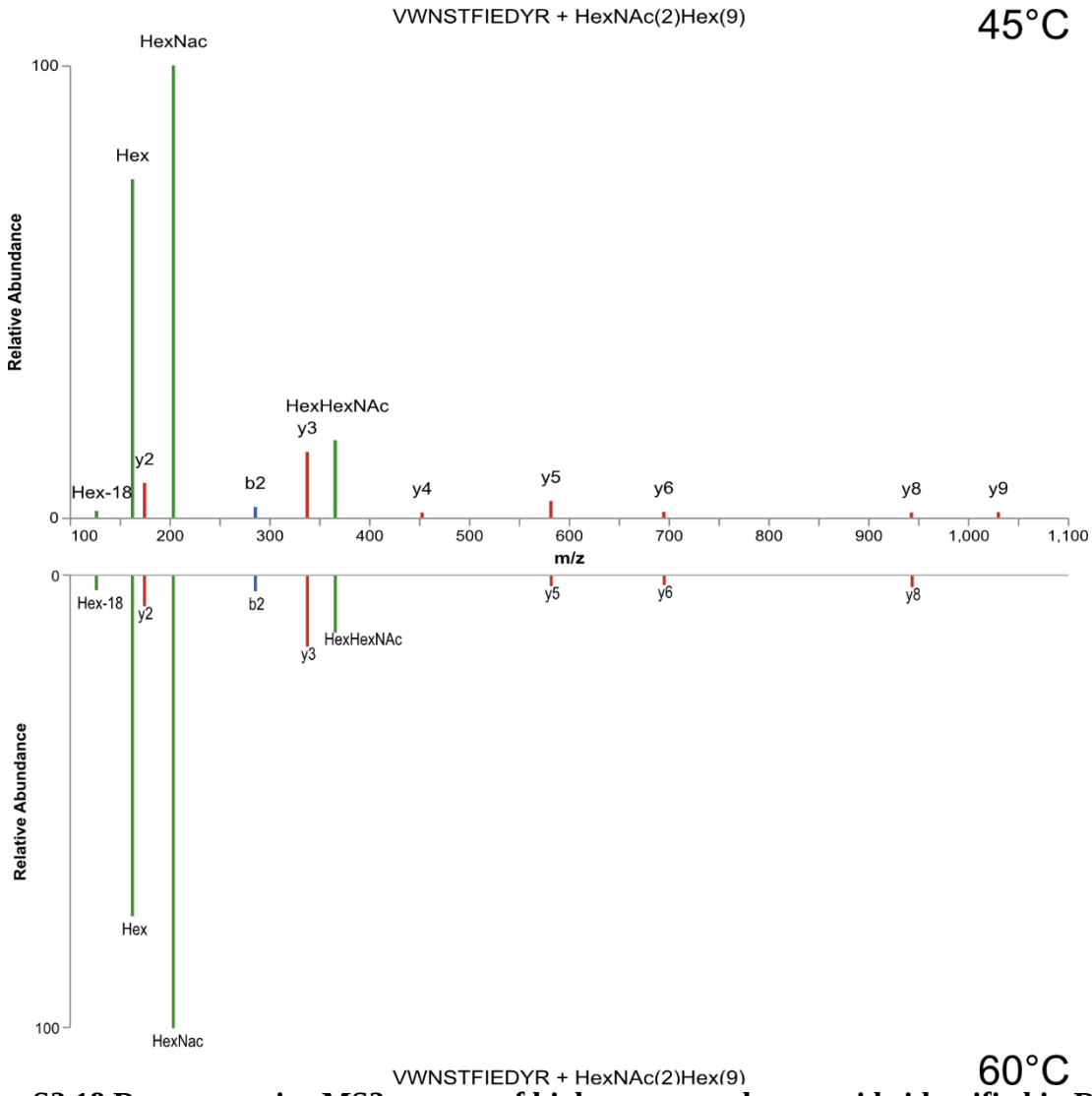


Figure S2.18 Representative MS2 spectra of high mannose glycopeptide identified in RPLC experiments.

This is also the same species shown in Figure S2.17. Note, this precursor mass was only selected for dissociation at 45°C and 60°C and no MS^2 information exists from 30°C experiments. Despite the differences in MS^1 reporting signal, we do not observe significant differences in MS^2 fragmentation, indicating that our manual evaluations of glycopeptide EICs is not obfuscated by mistaken identity and that increases in temperature are not inspiring differences in fragmentation.

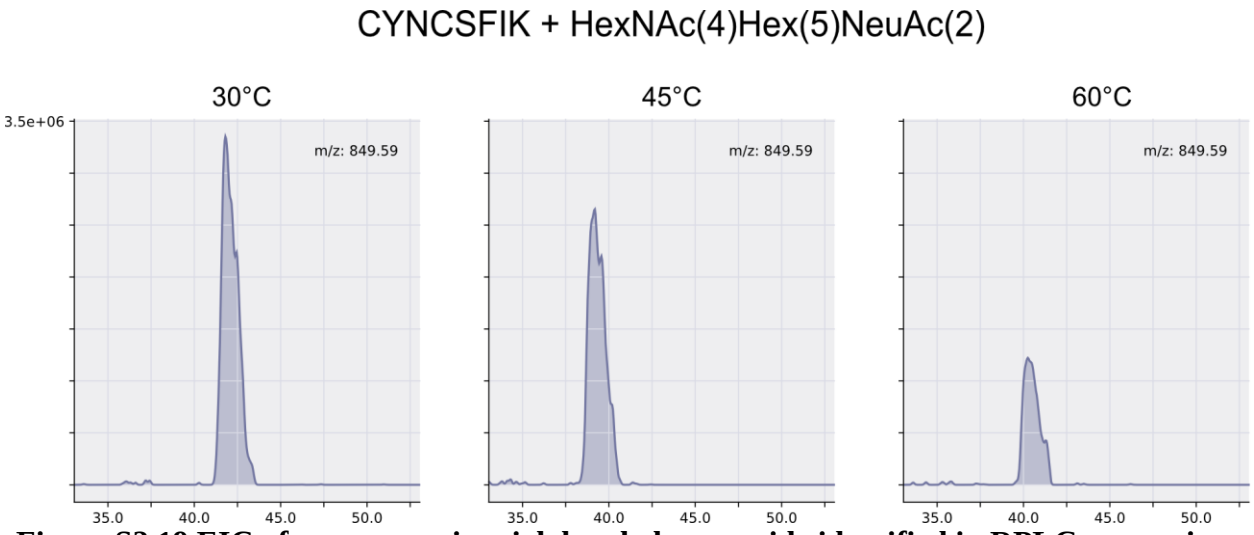


Figure S2.19 EIC of representative sialylated glycopeptide identified in RPLC separations at all temperatures.

A notable observation in our experiments is the gradual decline in MS^1 signal intensity for numerous sialylated glycopeptides within our experiments. Considering this trend arises with increases in temperature, we hypothesize the higher energy levels (via increased temperature) cause a loss of some labile glycosidic-bonded species. This reality would result in a lower overall reporting signal and lower identification rates of unique glycopeptides.

Chapter 3: Higher Temperature Porous Graphitic Carbon Separations Differentially Impact Distinct Glycopeptide Classes

Delafield, D. G., Miles, H. N., Ricke, W. A., & Li, L. (2023). Higher Temperature Porous Graphitic Carbon Separations Differentially Impact Distinct Glycopeptide Classes. *Journal of the American Society for Mass Spectrometry, 34*(1), 64-74.

Abstract

Mass spectrometry-based discovery glycoproteomics is highly dependent on the use of chromatography paradigms amenable to analyte retention and separation. When compared against established stationary phases such as reversed phase and hydrophilic interaction liquid chromatography, reports utilizing porous graphitic carbon (PGC) have detailed its numerous advantages. Recent efforts have detailed the utility in porous graphitic carbon in high throughput glycoproteomics, principally through enhanced profiling depth and liquid phase resolution at higher column temperatures. However, increasing column temperature has shown to impart disparaging effects in glycopeptide identification. Herein we further elucidate this trend, describing qualitative and semi-quantitative effects of increased column temperature on glycopeptide identification rates, signal intensity, resolution, and spectral count linear response. Through analysis of enriched bovine and human glycopeptides, species with high mannose and sialylated glycans were shown to most significantly benefit and suffer from high column temperatures, respectively. These results provide insight as to how porous graphitic carbon separations may be appropriately leveraged for glycopeptide identification while raising concerns over quantitative and semi-quantitative label free comparisons as temperature changes. RAW MS glycoproteomic data are available via ProteomeXchange with identifier PXD034354.

Introduction

Glycosylation is one of the most prevalent and heterogenous post-translational modifications (PTMs) within the human proteome[1-3]. The inherent complexity of its study is not felt just in considering the vast expanse of known and theoretical modification sites[2], but also through appreciation of the high degree of compositional and structural complexity[4]. Decades of targeted research have revealed the high degree of glycosylation diversity presents an equally diverse set of functions. Glycans and glycoproteins are known to participate in a litany of biological processes such as cell signaling[5-7], host-pathogen interaction[8-10], and protein folding[11-13], and are ever increasingly implicated in health and human disease. Despite the importance of these PTMs and the extensive effort contributed to their study, analytical strategies that can keep pace with biological complexity remain limited.

In order to offset the overwhelming difficulties in glycopeptide analysis, numerous areas of analytical development have received significant attention in recent years. Glycopeptide enrichment strategies have been developed to compensate for low analyte abundance in proteolytic mixtures[14-16], with some being tailored to unique glycan classes[17, 18]. As well, ubiquitous utilization of tandem-MS for glycopeptide identification has brought about validation and application of numerous unique dissociation modalities[19-22]. Even efforts for accurate quantitation of intact glycopeptides have seen a rise in popularity as isotopic and isobaric labeling strategies become more accessible[23]. These areas of development, however, largely ignore any potential benefits that may come through more effective chromatography.

Glycoproteomic analyses continue to utilize traditional reversed-phase liquid chromatography (RPLC) due to the effective retention mechanism and ease of implementation. This separation strategy provides meaningful access to the glycoproteome, but analyte retention is

dependent on a dominant hydrophobic character of the peptide backbone – a characteristic not observed for all glycopeptides. Furthermore, the large, hydrophilic glycan moiety often reduces the efficacy of RPLC for glycopeptide retention and separation[24, 25]. Hydrophilic interaction chromatography (HILIC) and electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) are popular separation modalities that promote greater retention of glycopeptides[26-28] but are less commonly implemented on-line due to the need for salt-containing buffers[27, 29] and sample phase changes prior to MS analysis.

Offering reprieve from these shortcomings, porous graphitic carbon (PGC) has demonstrated great utility in the retention and separation of polar analytes. PGC has been extensively used in the analysis of released glycans[30-35] and has even shown baseline resolution of glycan isomers when run at elevated temperatures[36]. Increasing column running temperatures promotes an expanded glycan morphology, increasing the partition coefficient and improving resolution. These improvements in glycan retention and separation have recently been validated for targeted[37] and discovery-based glycoproteomics analyses where higher temperature PGC separations were shown to enhance glycoproteomic coverage and profiling depth[38]. However, these improvements were shown to come at a cost.

Whereas previous studies demonstrated that modest increases in column temperature result in higher peptide and glycopeptide identification rates, these improvements often diminished at higher column temperatures[38]. With specific attention drawn to high mannose and sialylated glycopeptides, these two glycopeptide classes were shown to yield the most significant changes in identification rates and signal response. As PGC separations are increasingly employed for glycan and glycopeptide analysis – and due to the biological significance of affected glycopeptide classes – these observations present substantial roadblocks in the pursuit of successful glycoproteomic

analysis. Understanding the cause of altered identification rates and the effects higher temperatures impart on intact glycopeptides is imperative to promoting enhanced glycoproteomic coverage and providing guidance over experimental conditions that reduce analytical efficacy.

To survey the effects of higher column temperatures in PGC separations, we performed discovery-based glycoproteomic analysis on glycopeptides enriched from human prostate cancer cell lysate, supplemented with sialoglycopeptides from bovine standard proteins. Recreating previous experimental conditions[38], we reaffirm profiling depth is enhanced with modest increases in column temperature (45°C). Increasing temperature to 60°C, however, results in significant disparities in glycopeptide detection. Confining much of our focus to high mannose and sialylated glycopeptides – those species most differentially affected – we demonstrate how elevated temperatures are responsible for altered reporting signal and peak shape, affecting detection and identification. Further knowing the prevalence of label-free and reaction monitoring quantitative approaches that rely on precursor peak intensity or area, we also analyzed serial dilutions of enriched glycopeptide mixtures to evaluate impact of column temperature on glycopeptide quantitation. Mirroring the observations seen in glycopeptide identification, glycopeptide spectral matches were shown to significantly deviate as temperature climbs, indicating label-free comparisons across temperatures are not viable without special consideration. Nevertheless, for almost all glycopeptide classes, our data demonstrate greater spectral-counting-based quantitative accuracy at the highest temperature, 60°C. The disparity in observations between discovery and quantitative analyses suggest that column temperature must be individually tailored to suit biological discovery or quantitative accuracy. Overall, the findings presented within serve to highlight notable limitations and topical concerns for future PGC-based glycoproteomics analyses.

Methods

Materials

Dithiothreitol (DTT, D9779), iodoacetamide (IAA), sodium dodecyl sulfate (SDS), trifluoroacetic acid (TFA), bovine ribonuclease B (R7884), bovine fetuin (F3004), bovine asialofetuin (A4781), and bovine α1-acid glycoprotein (G3643) were purchased from Millipore Sigma (Burlington, MA). Urea (U15), tris-base (BP152), hydrochloric acid (A144SI), formic acid (A117), water, acetonitrile, and PGC guard columns (35003-014001) were purchased from Fisher Scientific (Waltham, MA). Capillary tubing (1068150019) was purchased from PolyMicro. Trypsin (V5113) was purchased from Promega (Madison, WI). PolyHYDROXYETHYL-A packing material was purchased from PolyLC. Pencil column heater was purchased from Phoenix S&T. Sources of all other materials are noted.

Protein Digestion

300mg of each standard glycoprotein was aliquoted into separate reaction microtubes and dried under vacuum. Urea was dissolved in 50mM Tris-HCl to a concentration of 8M, which was then used to resuspend standard glycoproteins at a concentration of 2.0µg/µL. Disulfide bonds were reduced with 5mM DTT at room temperature for 1 hour, followed by alkylation with 15mM IAA at room temperature for 30 minutes in the dark. Alkylation was quenched with excess DTT prior to diluting the mixture 1:10 to reduce Urea concentration to <1M. Trypsin was added 1:100 for 4 hours at 37°C, followed by 1:50 and overnight incubation at 37°C.

Cell Preparation

Prostate cancer cell lines ($BCaP^{MT10}$) were generated and described previously[39]. Cell lines were grown and maintained in phenol-free DMEM/Ham's F12 (Gibco) supplemented with 5%

fetal bovine serum (HyClone) and 1% penicillin-streptomycin solution (Gibco). T175 culture flasks were placed in an incubator at 5% CO2 and 98% humidity. Cells were grown to 90% confluency, washed with 1× phosphate-buffered saline (Cytiva), and harvested using a cell scraper. Approximately 1e6 cells were collected after culture. Cell pellets were washed twice using phosphate-buffered saline and stored at −80°C for subsequent processing.

Cell pellets were resuspended in 4 volumes of 4% SDS prior to lysis via ultrasonication. Protein concentration was estimated via bicinchoninic acid (BCA) assay (ThermoFisher Scientific). Disulfide bonds were reduced with 450nM DTT for 30 minutes at 55°C followed by alkylation with 10mM IAA at room temperature for 15 minutes. Protein was extracted through repeated additions of cold 80% acetone and incubation at -20°C. Protein was reconstituted in 8M Urea with 1x protease inhibitor cocktail. 300mg aliquots were taken, diluted, and digested similarly to the standard proteins.

Glycopeptide Enrichment

Glycopeptides were enriched using a custom spin tip method. Briefly, 200µL pipette tips were packed with 3mg sterile cotton and loaded with polyHYDROXYETHYL-A packing resin at a 30:1 bead-to-peptide ratio. Beads were hydrated in 1% TFA prior to loading. Beads were flushed with 1% TFA and conditioned with 80%ACN+1%TFA. Samples were resuspended in 80% ACN+1% TFA and applied to the beads a total of six times, which was followed by six washes in the same buffer to remove non-glycosylated peptides. Glycopeptides were eluted in 10% ACN+5% FA prior to drying under vacuum. Samples were reconstituted in 0.1% FA and serially diluted 1x to 16x prior to LC-MS analysis (Figure 3.1).

Column Preparation

Capillary tubing (o.d.=360um, i.d.=75um) was trimmed to a length >30cm with a small portion of the polyimide coating removed. A Sutter P-2000 micropipette laser puller was used to stretch the glass capillary into the final emitter geometry. The closed, tapered tip was then opened via emersion in 48% hydrofluoric acid for 2.5 minutes. 3um PGC packing material was harvested from PGC guard columns (Thermo Fisher Scientific, 35003-014001) and resuspended in optima grade methanol in a 1.5mL glass vial. A micro flea stir bar was added to provide agitation, and the slurry and etched capillary were placed in a pressure injection cell (Next Advance, PC77-MAG). The capillary was packed with PGC material using ~1000psi industrial grade nitrogen until 30cm of the capillary was full. Pressure was released and the same packing conditions were used to flush the column with methanol and air, sequentially. Columns were stored at room temperature until use.

LC-MS Analysis

Enriched glycopeptide mixtures were analyzed in technical triplicate using a Nano Ultimate 3000 chromatography stack coupled to an Orbitrap Fusion Lumos mass spectrometer. Glycopeptides were trapped on the column at 3% B for 18 minutes, followed by a 52-minute gradient from 15% B to 40% B. Column was washed at 75% B for ten minutes followed by another ten-minute wash at 95%B. The column was then equilibrated to 3% for the final ten minutes. The Fusion Lumos was set to perform DDA analysis using a 3-second MS^2 acquisition window rather than selecting top N precursors. MS^1 settings were as follows: resolution, 120,000; m/z range, 400-2,000; RF lens, 30%; AGC target, 2.0e5; maximum injection time, 50ms; microscans, 1; polarity, positive. MS^2 settings were as follows: resolution, 60,000; detector, orbitrap; isolation width, 1.6m/z; activation, HCD; collision energy mode, stepped HCD; collision energies, 20-30-40; first

mass, 120m/z; AGC target, 5.0e4; maximum injection time, 118ms. Precursors were also required to meet an intensity threshold of 2.5e4 for selection. Allowed charges states were 2+ to 7+; precursors were excluded after 1 occurrence for 15 seconds.

Standard Peptide Analysis

A standard disialylated glycopeptide (KVANK[HexNAc4Hex5NeuAc2]T) was purchased from TCI America (S0523). Listed at >95% pure with uniformity confirmed through gel and capillary electrophoresis, this product enables analysis of a highly uniform glycopeptide not subject to biological variation or abundance constraints. Preparation and analysis of this standard is detailed in the Supplemental Information.

Data and Code Availability

The mass spectrometry glycoproteomics data have been deposited to the ProteomeXchange[40] Consortium via the PRIDE[41, 42] partner repository with the dataset identifier PXD034354 and 10.6019/PXD034354. Data were searched using Proteome Discoverer 2.5 with the Protein Metrics Byonic node; searching and filtering parameters are described in the Supplemental Information. Search results are available in the PRIDE repository. All code used for analysis is available at https://github.com/lingjunli-research/pgc-glycosylation-lfq.

Results and Discussion

Changes in Glycopeptide Identification

Previous reports have signaled improved peptide and glycopeptide identification rates at elevated column temperatures[38], indicating potential differences in analyte behavior or optimal separation and desolvation efficiency. This trend is rearticulated in this study where, especially at higher concentrations (1x, 2x, and 4x dilutions), 45°C provided the highest overall rates of

glycopeptide identification (Figure 3.2**,** Table S1). Though this behavior is not conserved at lower concentrations (8x and 16x dilutions), a consistent trend across all analyses is that 60°C provided the lowest rates of glycopeptide identification. This latter observation is in-line with previous reports[38] and serves as an early marker for suboptimal chromatography conditions. Mapping glycopeptides back to their parent protein, our data demonstrate that the increase in glycopeptide identifications at 45°C stem from the detection of glycosites on previously undetected proteins, rather than new, additional glycosites or glycopeptides from proteins already identified (Figure 3.2). Interestingly, while our data show that the majority of glycosites are identified across all temperatures (Figure 3.2**)**, each unique glycopeptide dilution contained proportional quantities of unique glycosites with statistically insignificant differences across all but the most concentrated sample (Figure S3.1). As our data showed high intra-sample reproducibility in the identification of glycosites at all temperatures (Figure 3.2), we are confident this complementary detection is not due to analytical inaccuracies and serves to corroborate previous observations[38] that column temperatures provide access to different portions of the glycoproteome.

We investigated whether the peptide backbone played a significant role in the retention of glycopeptides at different temperatures. As shown in Figure 3.2, the relative hydrophilicity – presented here in the form of grand average of hydropathy (GRAVY) – and peptide length are well conserved across temperatures for each glycopeptide dilution. While the data show some slight preference for more hydrophilic analytes at higher temperatures, this observation is biased by fewer identified glycopeptides and should be further investigated in analyses of unmodified tryptic peptides where a hydrophilic glycan moiety plays no role. Interestingly, our data show a decrease in GRAVY score that mirrors glycopeptide concentration. This lower average value is mostly due

to the lower number of identified glycopeptides, but it does speak to the power of PGC to selectively trap, retain, and elute these highly polar, hydrophilic analytes.

However, the differences in glycopeptide identification begin to take shape when examining the classes of glycans identified (Figure 3.2). High mannose glycans were the most predominant modification identified across our analyses. These glycans, considered immature within the biosynthetic pathway[4], may be seen in greater quantity due to the known relationship between cancer cell proliferation and glycan nascence[43-45], though targeted investigation is needed for confirmation within this respective cell line model. Interestingly, while we anticipated complex glycans would be the second most prevalent glycan type, instead our data reveal paucimannose glycans as the next most common. Paucimannose glycans are relatively understudied in humans as it can be conjectured these glycans result from degradation or harsh preparation conditions[46]. However, recent studies have validated the occurrence of these glycan types in humans and have even been shown to play significant roles in human cancer[47-49]. Sialoglycopeptides, complex glycopeptides, and fucosylated glycopeptides follow in number of identifications. A clear disparity between mannose and the latter three glycan subtypes is the effect seen when elevating column temperature. While identifications of glycopeptides in all classes increased or were unaffected at 45°C, identifications of mannose subtypes continued to increase at 60°C (Figure 3.2), albeit marginally. Conversely, the remaining glycan subtypes demonstrated overall worse performance at the highest temperature. This trend is less severe in complex glycopeptides but is immediately noticeable in fucosylated and sialylated glycopeptides.

Increasing glycopeptide identifications at 45°C are consistent with previous reports[38]; however, these data show more substantial increase of mannose subtypes at higher temperatures, which we attribute to different cell lysis methods and scale-up. Nevertheless, the reproducibility

of declining identifications at high temperatures presented in our data indicates these observations are not due to experimental aberrations such as run-to-run variance or uncontrolled variables. Instead, we hypothesize that raising column temperatures induces some underlying chemical or chromatographic abnormalities that are responsible for affecting identification rates. Our results may be further dissected to provide qualitative and semi-quantitative investigation of this hypothesis.

Aberrant Chromatographic Behavior

To survey any potential underlying abnormalities that may be responsible for the observations mentioned above, we compiled extracted ion chromatograms (XICs) for all identified glycopeptide masses in the highest concentration sample. Given the different quantities of glycopeptides as temperature rises, it would be reasonable to expect that glycopeptides not identified were truly absent or were seen at such low intensities they could be mistaken for noise. Neither of these suspicions were confirmed. Our analyses demonstrate that the vast majority of glycopeptide masses were both present and observed at appreciable intensity in every run – an observation that holds true regardless of glycan class, as evidenced below. This further supports our hypothesis that column temperatures are responsible for affecting identification rates.

To avoid misinterpretation of data, we first manually inspected the XIC images, removing any species that were poorly extracted or contained traces of insufficient quality (i.e., indistinguishable major peaks, co-extracted masses, etc.). Inspecting the remaining data, several trends became obvious. First, and most expectedly, increasing column temperature resulted in nearly unanimous increases in resolution, with XICs displaying narrower full width at half maximum (FWHM) and resolution of some putative glycopeptide isomers. More interestingly, however, is the disparity in overall signal intensity and peak height seen between glycopeptides of

different classes. As temperature increased, the improved resolution for high mannose glycopeptides translated into higher signal intensity (i.e., narrower, taller peaks) (Figure 3.3). This trend is largely conserved across high mannose glycopeptides to varying extents. At the same time, sialoglycopeptides show a similar improvement in resolution and signal intensity at 45°C but show a significant drop off at 60°C (Figure 3.3). These observations rearticulate those seen elsewhere[38] and provide direct evidence of temperature-correlated analyte response. If these temperature effects were directly related to glycopeptide detection and identification, we anticipate other glycopeptide subtypes to display similar correlations. Indeed, examining paucimannose and fucosylated glycopeptides reveal similar evidence. Paucimannose glycopeptides generally benefited from higher column temperatures while fucosylated glycopeptides showed higher intensity at 45°C that waned at higher temperatures (Figure S3.2). It should be noted that the associated trends are less significant for these latter two glycopeptide classes, but further study and broader collections of analytes may serve to definitively characterize their response to elevated temperatures.

With evidence to support our hypothesis that temperature changes promote chromatographic behavior that impacts glycopeptide identification, we sought to provide further qualitative analysis to aid in characterizing these occurrences. For all glycopeptides identified in the highest concentration sample, we isolated the major peak and determined peak height, FWHM, and peak area via curve integration. We selected only the major peak for these analyses as the confident assignment of minor peaks must rely on exact match of isotopic envelopes to that of the major species that was selected for MS^2 fragmentation; we found this to be untenable, given the quantity of unique glycopeptide identifications, replicates, and possible minor species.

Examining the data underpinning these glycopeptide identifications, all classes demonstrated the expected decrease in FWHM as temperature increases, reflecting the known improvements in liquid phase resolution achieved as glycan morphology expands (Figure 3.4). As well, all glycopeptide classes demonstrated changes in peak intensity that directly reflect the observations made above, though these changes are most visible for sialylated glycopeptides. Similarly, when integrating area under the curve, high mannose and paucimannose glycopeptides demonstrate little discrepancy in peak area as temperature increases, contrasting that of complex and sialylated glycopeptides (Figure 3.4). Taken together, we posit those improvements in liquid-phase resolution for mannose-type glycopeptides at high temperatures result in sharper, narrower elution peaks that conserve the overall peak area observed at lower temperatures. Because these observations directly correlate with identification rates, deviating peak intensities and areas are likely directly responsible for the incremental improvement in identifications, as greater intensity will raise precursor priority when performing DDA-MS/MS.

In order to validate this conjecture, we utilized the extracted information from each glycopeptide to represent the fold-change in relative abundance with respect to the base temperature, 30°C. Averaging across technical triplicates, the vast majority – though not all – glycopeptides exhibit higher relative abundance at 45°C (Figure 3.4), mirroring the aforementioned XIC observations and identification rates. However, where these data begin to diverge is the relative abundance seen at the highest temperature, 60°C. For high mannose and paucimannose glycopeptides, relative abundances are often even higher than those observed at 45°C. While these glycopeptide classes do show some species to be lower in abundance at 60°C than at 45°C, the relative abundance at the highest temperature are often equivalent to or higher than the abundances seen at the base temperature of 30°C. Contrary to this observation, complex,

fucosylated, and sialylated glycopeptides show significant drops in relative abundance at 60°C, with the majority of glycopeptides displaying lower or substantially lower abundances compared to that observed at either of the two lower temperatures.

These observations and conjectures notwithstanding, we acknowledge the presence of glycopeptides that deviate from these noted trends. However, these cases are the minority, giving way to the prevailing observations discussed here. This diversity in results should be expected for any biological population and therefore do not significantly hamper our interpretation of the overall trends seen across glycopeptide classes. Fucosylated glycopeptides, as well, show diversity in their relative abundances as temperature climbs. Given the lower number of these species and that fucosylation and sialylation often co-occur, these data may be further reorganized and investigated in later experiments to provide a more comprehensive dataset.

Overall, these data serve to confirm our hypothesis that increases in temperature induce chromatographic behavior that impacts glycopeptide identification. As shown, mannose type glycopeptides benefit from increased temperatures as their increased resolution results in greater peak intensity and therefore greater selection in DDA-MS/MS. Complex, fucosylated, and sialylated glycopeptides do benefit from improved resolution at higher temperatures but often yield lower peak intensities, affecting downstream identification. Beyond this, given the demonstrated abnormalities in glycopeptide detection and peak shape as temperature increases, we suspect common label-free quantitative methods may be unreliable under these chromatographic conditions.

Assessing Spectral Count-Based Quantitation

Strategies for glycopeptide quantitation have experienced a surge in innovation over the past decade[23]. While metabolic labeling, isotopic chemical tags, and isobaric labeling strategies

have all been demonstrated as viable and effective, these approaches are often custom-tailored and require additional handling that can introduce sample loss. Label-free quantitation (LFQ) avoids these complications but is prone to missing values and run-to-run variance. Common LFQ approaches such as reaction monitoring rely on precursor area under the curve or intensity of transition ions[50] as the quantitative marker, while others such as spectral counting assume relative quantity is proportional to frequency of MS selection. Given each of these methods relies on precursor and fragment intensity, severe limitations may be met when employing PGC separations at elevated temperatures.

Averaging the number of peptide spectral matches (PSMs) across technical replicates, a subset of glycopeptides was shown to be quantifiable (i.e., identified in ≥3 dilutions) with good linearity (Figure 3.5). Notably, fucosylated glycopeptides demonstrated the worst linear response; this is due to the relatively low abundance of this class of glycopeptide – evidenced by the low number of identifications and PSMs. When comparing across the remaining glycopeptide subtypes, high mannose glycopeptides demonstrated the highest density of species with linear regression fit >0.9, though all classes demonstrated a meaningful distribution of high linearity species. More interestingly, quantifiable complex, fucosylated and sialylated glycopeptides showed greater PSM-based linearity as temperature climbed (Figure 3.5). This observation presents a caveat to our existing discussion of how elevated temperatures bring limitations to glycopeptide identification. Though high temperatures (i.e., ≥60°C) reveal the fewest overall identifications, those that are identified tend to show excellent linear response and may be more easily relied upon for quantitation. On the other hand, mannose type glycopeptides generally showed a decrease in linear response at 45°C before becoming bimodally distributed at 60°C (Figure 3.5). These data inform us that column temperatures may be tailored based on intended

experimental outcome; 45°C typically provides best signal response but does not provide the best quantitation for all species.

While these data suggest at which temperatures spectral counting-based quantitation may be achieved, we have not established whether individual glycopeptides may be accurately compared across temperatures. As shown in Figure 3.5, all glycopeptide classes yield different quantities of PSMs at different temperatures. This observation may be obvious, given the discussion of chromatographic behavior. However, mannose and fucosylated subtypes generally showed a more conserved spectral count as temperatures increased compared to complex and sialylated glycopeptides. As seen, sialylated glycopeptides displayed a demonstrable increase in the number of PSMs at 45°C before decreasing again at 60°C, which reflects the deviation in signal intensity at these temperatures. Complex glycopeptides, on the other hand, exhibited steady increases in the number of PSMs as temperature climbs higher, being only one of two peptide classes that display such trend.

Taken together, these data are not intended to confer any quantitative information beyond the efficacy of using a label-free strategy for glycopeptide analysis when column temperature is an experimental variable. While our data demonstrate good linear response when PSMs are averaged as a function of sample dilution, this is only true when temperature is constant. In large part, glycopeptides cannot be directly compared across different temperatures in a label-free fashion. This conclusion is evidenced first by the significant changes in PSM count as temperature changes (Figure 3.5) and when considering the noted differences in precursor peak intensity and area (Figure 3.3**,** Figure S3.2). While quantitative accuracy remains to be evaluated when using area under the curve, our data suggest this approach may not be suitable; we did attempt to perform XIC area based LFQ in these analyses, but the breadth of identifications combined with technical

replicates at three temperatures was untenable and computationally inefficient. Perhaps MS^2-based transition monitoring is a more appropriate strategy for glycopeptide LFQ when using elevated column temperatures, but this claim should be investigated on its own.

Hypotheses and Future Directions

Given the nature of our data, it is imperative we consider the underlying phenomena that induce the observed chromatographic behavior shown to impact glycopeptide analysis. Focusing on sialylated glycopeptides, those species most adversely affected, two prevailing hypotheses exist that may explain the significant decrease in signal intensity at high temperatures. The first hypothesis is that sialoglycopeptides exhibit greater structural diversity in nature (i.e., antennae linkage, α/β orientation, etc.) and this structural diversity is readily resolved at high temperatures, as shown previously[36, 37]. This increased resolution distributes the density of glycopeptides eluting per unit time, lowering overall peak heights. The alternative explanation is that liquid phase separations at high temperatures impart greater energy into the system, provoking early dissociation of labile sialic acid linkages. As well, sialylated glycopeptides are known to be labile under acidic conditions[17, 51, 52]; our 0.1% FA additive during LC-MS, though common practice, likely exacerbates their lability and may contribute to early dissociation. We do not find strong evidence to support this hypothesis in our data as examining the identified sialoglycopeptides did not provide any such correlation between high sialylation states and lower sialylation states. To state tersely, we did not observe any disalylated glycopeptides converted to monosialylated glycopeptides, and so on. We did further investigate this claim through analysis of a commercially available sialoglycopeptide standard and no early dissociation could be definitively observed (Figure S3.3). In addition, we did not observe any relationship between the five high mannose

glycans and identification rate, peak height, etc., further reinforcing the trends discussed in this report as global observations and not specific to a subset of modifications.

In all, the chemical or physical cause of altered glycopeptide identifications at elevated temperatures is still not clear. Anecdotal evidence supports the idea that the acidic conditions chosen for LC-MS analysis does cause dissociation of sialic acid. Considering the addition of higher temperatures during separation and desolvation, indiscriminate cleavage of glycosidic bonds is not outside the realm of possibilities. As our study provided some semi-quantitative investigations of this anomaly, we propose a more rigorous means of quantitation may be employed to fully elucidate this trend. Establishing specific transition ions for a broad array of glycopeptides and utilizing reaction monitoring (MRM or PRM) would eliminate any errors in precursor identification, extraction, and quantitation. Regardless of future approaches that provide a succinct connection between column temperature and glycopeptide identification, the data presented here provide heuristic guidance towards appropriate experimental design, depending on analytical objectives.

Conclusions

Porous graphitic carbon separations are a powerful addition to mass spectrometry-based glycoproteomics. Providing excellent retention of hydrophilic glycopeptides with a dominant glycan moiety, PGC provides facile access to regions of the glycoproteome that may be unobtainable through traditional separation modalities. As shown within, elevated column temperatures inspire significant improvements in liquid-phase resolution for all glycopeptide classes and yield greater reporting signal that aids in auto-MS identification. However, further increasing column temperature presents a tradeoff between liquid-phase resolution and glycopeptide identification. Our analyses show unique glycopeptide classes are differentially

impacted – mannose type glycopeptides appear to benefit from high temperatures while complex and sialylated glycopeptides do not. Beyond this, we demonstrate the feasibility of performing label-free glycopeptide quantitation when temperature is held constant. However, glycopeptide species cannot be reliably compared from one temperature to the next as their spectral count and precursor area under curve are shown to deviate substantially according to their glycan composition. In summary, PGC-based glycopeptide separation and discovery is most effective at 45°C, providing excellent reporting signal and modest resolution. However, in chasing isomeric resolution at higher temperatures, special consideration must be taken to avoid misinterpretation of glycopeptide identifications – or lack thereof – and when drawing comparisons to analyses under different experimental conditions.

Acknowledgments

Support for this research is provided in part by the National Institutes of Health (NIH) grants RF1 AG052324 (LL), R01 DK071801 (LL), and U54DK104310 (WAR, LL). The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. LL acknowledges the National Science Foundation funding support (CHE- 2108223 and IOS-2010789), NIH grant support R21AG065728, S10OD028473, and S10OD025084 as well as a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Figures

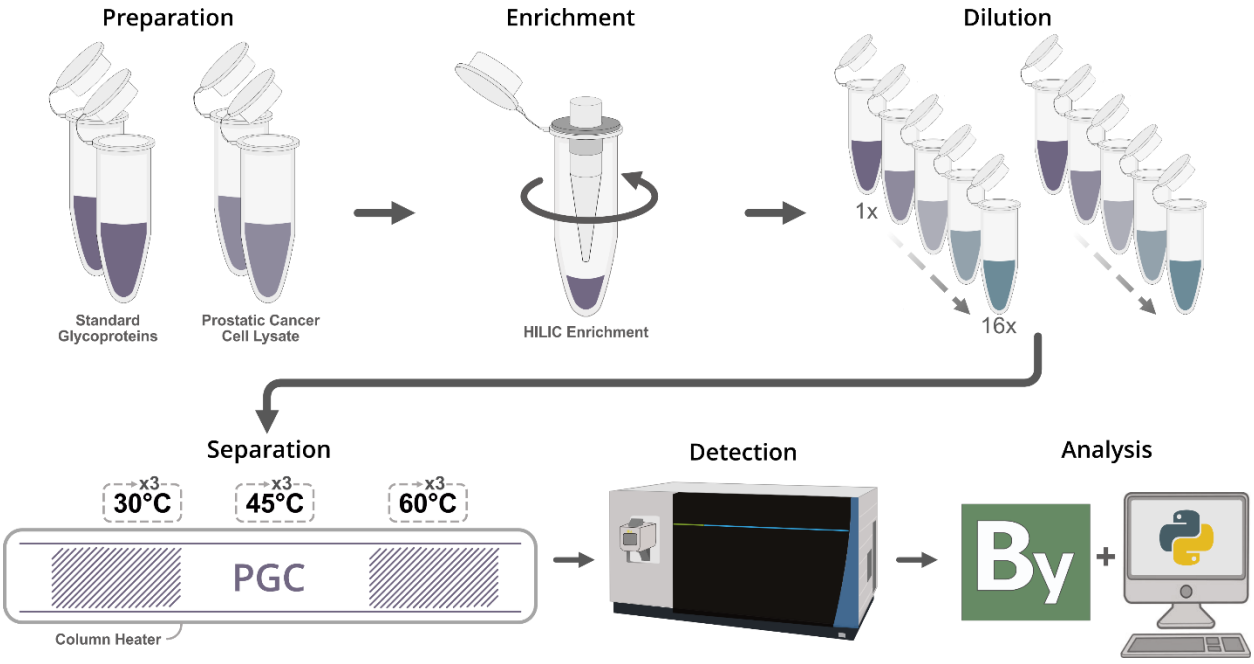


Figure 3.1 Outline of workflow.

Glycoprotein mixtures or cell lysate were enzymatically digested prior to glycopeptide enrichment. Glycopeptides were reconstituted, serially diluted, and analyzed in technical triplicate on a custom PGC-packed capillary column. Data were collected on an Orbitrap Fusion Lumos with database searching performed in Proteome Discoverer with the Protein Metrics Byonic node.

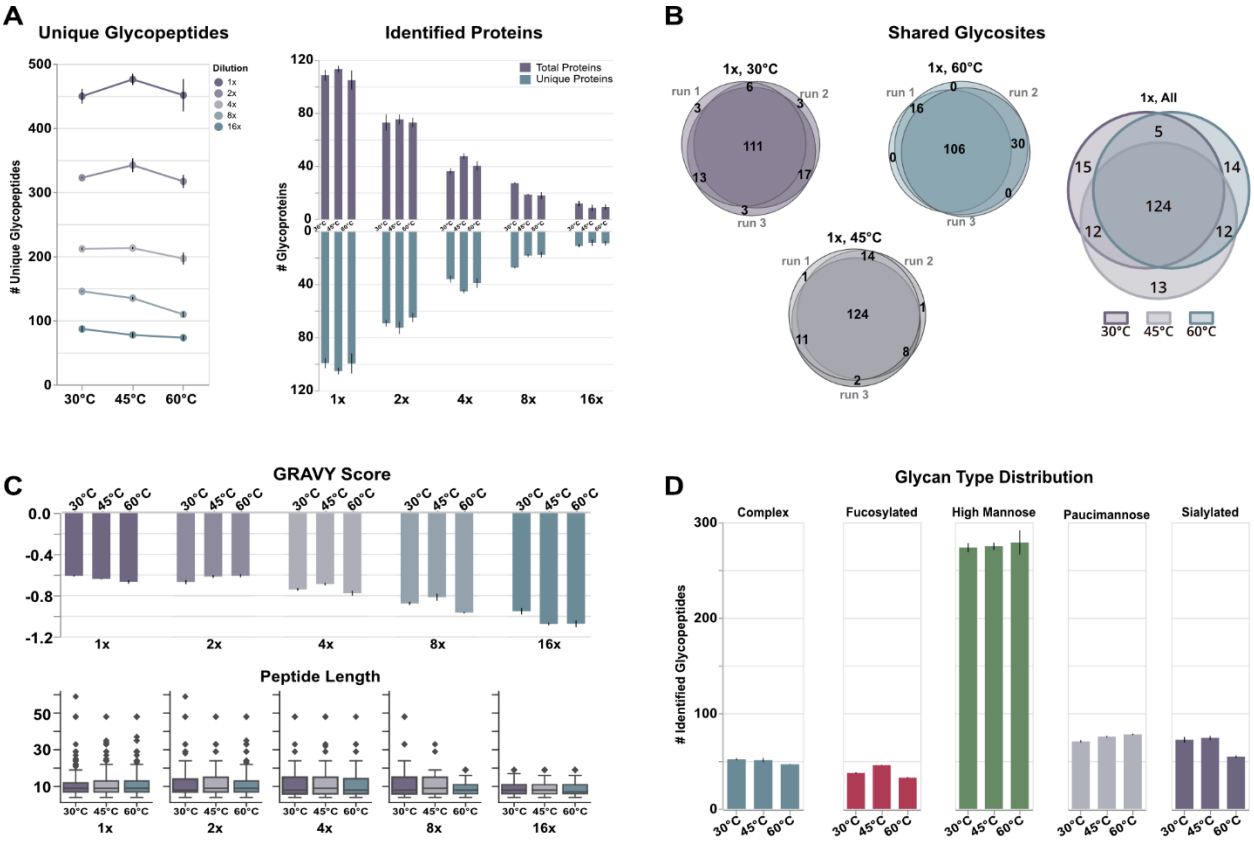


Figure 3.2 Results outlining temperature-based differences in glycopeptide identification.

A) Higher concentration glycopeptide samples (i.e., the undiluted sample (1x), and first two serial dilutions (2x and 4x)) displayed the expected increase in identifications when separated at 45°C, stemming largely from accessing new glycoprotein constituents. Concentration and temperature are independent; all concentrations (i.e., dilutions) were analyzed at all temperatures. B) Our analyses showed high technical reproducibility indicating the changes in unique glycosites between temperature is unlikely due to analytical abnormalities. C) The peptide backbone information of glycopeptides was largely conserved across temperatures, suggesting major differences are not due to the backbone itself. D) Glycopeptides demonstrated a class-dependent response to temperature with mannose type glycans benefitting from higher temperatures; complex, fucosylated and sialylated types show decreased identifications at high temperatures.

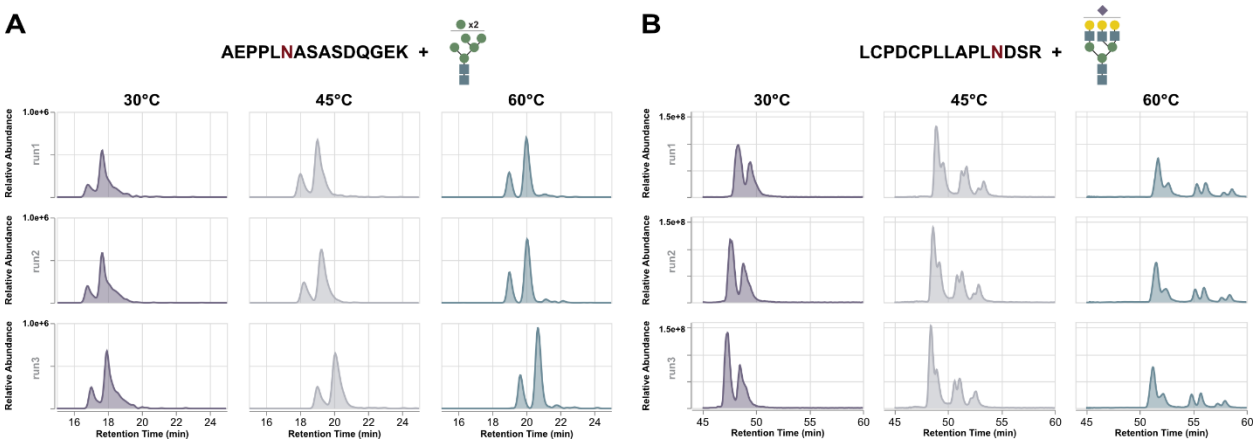


Figure 3.3 Representative extract ion chromatograms (XICs) of high mannose and sialylated glycopeptides.

Both glycopeptides demonstrate improved resolution as temperature increases but high mannose types (A) displayed further increases in signal intensity as temperature climbs. Conversely, sialylated glycopeptides (B) increase in intensity at 45°C before showing significant attenuation at 60°C.

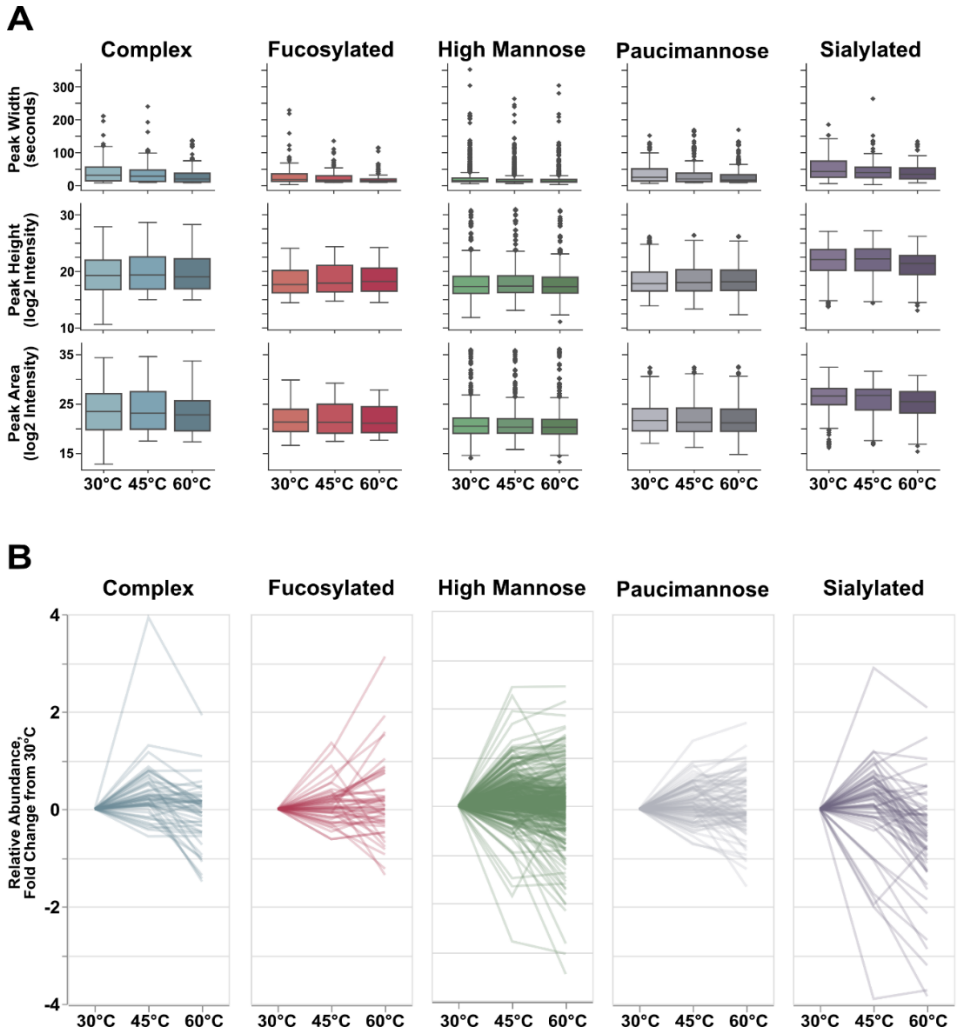


Figure 3.4 Qualitative metrics from identifiable glycopeptides.

A) All glycopeptides demonstrate improved resolution at high temperatures, but most classes demonstrate little change in median peak height or peak area. Sialylated glycopeptides, however, do show disparities in these metrics. However, this change in median value is only reflective of the whole population. B) When examining peak intensity for each individual glycopeptide, we observed that the majority of glycopeptides from each glycan class showed noticeable improvements in peak intensity at 45°C with a subset demonstrating statistical significance ($P \leq 0.05$, high mannose: 74 glycopeptides, 23.6% of high mannose population; paucimannose: 22 glycopeptides, 25.9% of population; complex: 16 glycopeptides, 30.8% of population; sialylated:

15 glycopeptides, 25.0% of population, fucosylated: 7 glycopeptides, 14.9% of population). Mannose type glycopeptide peak intensity is not largely impacted at 60°C, whereas complex and sialylated glycopeptides are – a finding that reflects their identification rates at high temperatures.

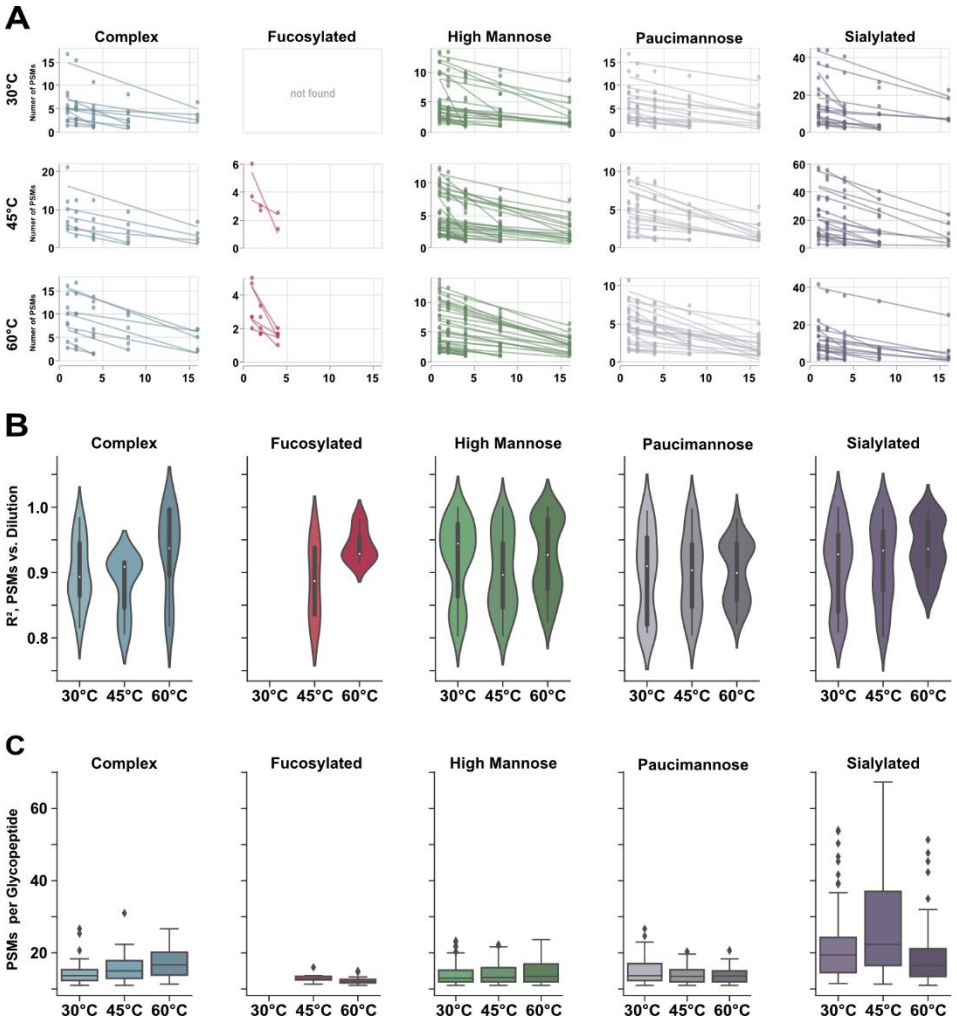


Figure 3.5 Semi-quantitative evaluations of glycopeptide dilution series across temperature. A) Plots demonstrating the linear response of glycopeptide identifications across dilutions. When temperature is held constant, spectral counting may be a viable option for LFQ. B) Distribution of linear regression for quantifiable glycopeptides as temperature increases. While 45°C is most beneficial for glycopeptide identifications, it is not the most optimal for LFQ. Conversely, identifiable glycopeptides may be best quantified at 60°C. C) Change in spectral count as temperature climbs. While PSM count for mannose type glycans is relatively conserved, complex and sialylated glycopeptides show significant differences across temperatures. This observation suggests that LFQ methods may be significantly impacted at different temperatures.

References

1. Abou-Abbass, H.; Abou-El-Hassan, H.; Bahmad, H.; Zibara, K.; Zebian, A.; Youssef, R.; Ismail, J.; Zhu, R.; Zhou, S.; Dong, X.; Nasser, M.; Bahmad, M.; Darwish, H.; Mechref, Y.; Kobeissy, F., Glycosylation and other PTMs alterations in neurodegenerative diseases: Current status and future role in neurotrauma. *Electrophoresis* **2016,** *37* (11), 1549-1561.

2. Khoury, G. A.; Baliban, R. C.; Floudas, C. A., Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific Reports* **2011,** *1* (1), 90.

3. Ramazi, S.; Zahiri, J., Post-translational modifications in proteins: resources, tools and prediction methods. *Database* **2021,** *2021*.

4. Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Mohnen, D.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., *Essentials of glycobiology*. Fourth edition. ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 2022; p pages cm.

5. Patwardhan, A.; Cheng, N.; Trejo, J., Post-Translational Modifications of G Protein–Coupled Receptors Control Cellular Signaling Dynamics in Space and Time. *Pharmacological Reviews* **2021,** *73* (1), 120-151.

6. Chandler, K. B.; Leon, D. R.; Kuang, J.; Meyer, R. D.; Rahimi, N.; Costello, C. E., N-Glycosylation regulates ligand-dependent activation and signaling of vascular endothelial growth factor receptor 2 (VEGFR2). *Journal of Biological Chemistry* **2019,** *294* (35), 13117-13130.

7. Freitas, D.; Campos, D.; Gomes, J.; Pinto, F.; Macedo, J. A.; Matos, R.; Mereiter, S.; Pinto, M. T.; Polónia, A.; Gartner, F.; Magalhães, A.; Reis, C. A., O-glycans truncation modulates gastric cancer cell signaling and transcription leading to a more aggressive phenotype. *EBioMedicine* **2019,** *40*, 349-362.

8. Lin, B.; Qing, X.; Liao, J.; Zhuo, K., Role of Protein Glycosylation in Host-Pathogen Interaction. *Cells* **2020,** *9* (4), 1022.

9. Mehaffy, C.; Belisle, J. T.; Dobos, K. M., Mycobacteria and their sweet proteins: An overview of protein glycosylation and lipoglycosylation in M. tuberculosis. *Tuberculosis* **2019,** *115*, 1-13.

10. Watanabe, Y.; Bowden, T. A.; Wilson, I. A.; Crispin, M., Exploitation of glycosylation in enveloped virus pathobiology. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2019,** *1863* (10), 1480-1497.

11. Wang, H.; Li, S.; Wang, J.; Chen, S.; Sun, X.-L.; Wu, Q., N-glycosylation in the protease domain of trypsin-like serine proteases mediates calnexin-assisted protein folding. *eLife* **2018,** *7*, e35672.

12. Nagashima, Y.; von Schaewen, A.; Koiwa, H., Function of N-glycosylation in plants. *Plant Science* **2018,** *274*, 70-79.

13. Tannous, A.; Pisoni, G. B.; Hebert, D. N.; Molinari, M., N-linked sugar-regulated protein folding and quality control in the ER. *Seminars in Cell & Developmental Biology* **2015,** *41*, 79-89.

14. Riley, N. M.; Bertozzi, C. R.; Pitteri, S. J., A Pragmatic Guide to Enrichment Strategies for Mass Spectrometry-based Glycoproteomics. *Molecular & Cellular Proteomics* **2021**, 100029.

15. Xue, Y.; Xie, J.; Fang, P.; Yao, J.; Yan, G.; Shen, H.; Yang, P., Study on behaviors and performances of universal N-glycopeptide enrichment methods. *Analyst* **2018,** *143* (8), 1870-1880.

16. Chen, Z.; Wang, D.; Yu, Q.; Johnson, J.; Shipman, R.; Zhong, X.; Huang, J.; Yu, Q.; Zetterberg, H.; Asthana, S.; Carlsson, C.; Okonkwo, O.; Li, L., In-Depth Site-Specific O-Glycosylation Analysis of Glycoproteins and Endogenous Peptides in Cerebrospinal Fluid (CSF) from Healthy Individuals, Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD) Patients. *ACS Chemical Biology* **2021**.

17. Li, M.; Huang, J.; Ma, M.; Shi, X.; Li, L., Selective Enrichment of Sialylglycopeptides Enabled by Click Chemistry and Dynamic Covalent Exchange. *Analytical Chemistry* **2022,** *94* (18), 6681-6688.

18. Huang, J.; Liu, X.; Wang, D.; Cui, Y.; Shi, X.; Dong, J.; Ye, M.; Li, L., Dual-Functional Ti(IV)-IMAC Material Enables Simultaneous Enrichment and Separation of Diverse Glycopeptides and Phosphopeptides. *Analytical Chemistry* **2021,** *93* (24), 8568-8576.

19. Riley, N. M.; Malaker, S. A.; Driessen, M. D.; Bertozzi, C. R., Optimal Dissociation Methods Differ for N- and O-Glycopeptides. *Journal of Proteome Research* **2020**.

20. Riley, N. M.; Hebert, A. S.; Westphall, M. S.; Coon, J. J., Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nature Communications* **2019,** *10* (1), 1311.

21. Escobar, E. E.; King, D. T.; Serrano-Negrón, J. E.; Alteen, M. G.; Vocadlo, D. J.; Brodbelt, J. S., Precision Mapping of O-Linked N-Acetylglucosamine Sites in Proteins Using Ultraviolet Photodissociation Mass Spectrometry. *Journal of the American Chemical Society* **2020,** *142* (26), 11569-11577.

22. Escobar, E. E.; Wang, S.; Goswami, R.; Lanzillotti, M. B.; Li, L.; McLellan, J. S.; Brodbelt, J. S., Analysis of Viral Spike Protein N-Glycosylation Using Ultraviolet Photodissociation Mass Spectrometry. *Analytical Chemistry* **2022,** *94* (15), 5776-5784.

23. Delafield, D. G.; Li, L., Recent Advances in Analytical Approaches for Glycan and Glycopeptide Quantitation. *Molecular & Cellular Proteomics* **2021,** *20*.

24. Furuki, K.; Toyo'oka, T., Retention of glycopeptides analyzed using hydrophilic interaction chromatography is influenced by charge and carbon chain length of ion-pairing reagent for mobile phase. *Biomedical Chromatography* **2017,** *31* (11), e3988.

25. Frost, D. C.; Li, L., Chapter Three - Recent Advances in Mass Spectrometry-Based Glycoproteomics. In *Advances in Protein Chemistry and Structural Biology*, Donev, R., Ed. Academic Press: 2014; Vol. 95, pp 71-123.

26. Qing, G.; Yan, J.; He, X.; Li, X.; Liang, X., Recent advances in hydrophilic interaction liquid interaction chromatography materials for glycopeptide enrichment and glycan separation. *TrAC Trends in Analytical Chemistry* **2020,** *124*, 115570.

27. Cui, Y.; Yang, K.; Tabang, D. N.; Huang, J.; Tang, W.; Li, L., Finding the Sweet Spot in ERLIC Mobile Phase for Simultaneous Enrichment of N-Glyco and Phosphopeptides. *Journal of the American Society for Mass Spectrometry* **2019,** *30* (12), 2491-2501.

28. Cui, Y.; Tabang, D. N.; Zhang, Z.; Ma, M.; Alpert, A. J.; Li, L., Counterion Optimization Dramatically Improves Selectivity for Phosphopeptides and Glycopeptides in Electrostatic Repulsion-Hydrophilic Interaction Chromatography. *Analytical Chemistry* **2021,** *93* (22), 7908-7916.

29. Alpert, A. J., Effect of salts on retention in hydrophilic interaction chromatography. *Journal of Chromatography A* **2018,** *1538*, 45-53.

30. Ashwood, C.; Lin, C.-H.; Thaysen-Andersen, M.; Packer, N. H., Discrimination of Isomers of Released N- and O-Glycans Using Diagnostic Product Ions in Negative Ion PGC-LC-ESI-MS/MS. *Journal of the American Society for Mass Spectrometry* **2018,** *29* (6), 1194-1209.

31. Ashwood, C.; Waas, M.; Weerasekera, R.; Gundry, R. L., Reference glycan structure libraries of primary human cardiomyocytes and pluripotent stem cell-derived cardiomyocytes reveal cell-type and culture stage-specific glycan phenotypes. *Journal of Molecular and Cellular Cardiology* **2020,** *139*, 33-46.

32. Ashwood, C.; Pratt, B.; MacLean, B. X.; Gundry, R. L.; Packer, N. H., Standardization of PGC-LC-MS-based glycomics for sample specific glycotyping. *The Analyst* **2019,** *144* (11), 3601-3612.

33. Wei, J.; Tang, Y.; Bai, Y.; Zaia, J.; Costello, C. E.; Hong, P.; Lin, C., Toward Automatic and Comprehensive Glycan Characterization by Online PGC-LC-EED MS/MS. *Analytical Chemistry* **2020,** *92* (1), 782-791.

34. Dong, X.; Peng, W.; Yu, C.-Y.; Zhou, S.; Donohoo, K. B.; Tang, H.; Mechref, Y., 8-plex LC–MS/MS Analysis of Permethylated N-Glycans Achieved by Using Stable Isotopic Iodomethane. *Analytical Chemistry* **2019,** *91* (18), 11794-11802.

35. Gautam, S.; Banazadeh, A.; Cho, B. G.; Goli, M.; Zhong, J.; Mechref, Y., Mesoporous Graphitized Carbon Column for Efficient Isomeric Separation of Permethylated Glycans. *Analytical Chemistry* **2021,** *93* (12), 5061-5070.

36. Zhou, S.; Huang, Y.; Dong, X.; Peng, W.; Veillon, L.; Kitagawa, D. A. S.; Aquino, A. J. A.; Mechref, Y., Isomeric Separation of Permethylated Glycans by Porous Graphitic Carbon (PGC)-LC-MS/MS at High Temperatures. *Analytical Chemistry* **2017,** *89* (12), 6590-6597.

37. Zhu, R.; Huang, Y.; Zhao, J.; Zhong, J.; Mechref, Y., Isomeric Separation of N-Glycopeptides Derived from Glycoproteins by Porous Graphitic Carbon (PGC) LC-MS/MS. *Analytical Chemistry* **2020,** *92* (14), 9556-9565.

38. Delafield, D. G.; Miles, H. N.; Liu, Y.; Ricke, W. A.; Li, L., Complementary proteome and glycoproteome access revealed through comparative analysis of reversed phase and porous graphitic carbon chromatography. *Analytical and Bioanalytical Chemistry* **2022**.

39. Liu, T. T.; Ewald, J. A.; Ricke, E. A.; Bell, R.; Collins, C.; Ricke, W. A., Modeling human prostate cancer progression in vitro. *Carcinogenesis* **2018,** *40* (7), 893-902.

40. Deutsch, E. W.; Bandeira, N.; Sharma, V.; Perez-Riverol, Y.; Carver, J. J.; Kundu, D. J.; García-Seisdedos, D.; Jarnuczak, A. F.; Hewapathirana, S.; Pullman, B. S.; Wertz, J.; Sun, Z.; Kawano, S.; Okuda, S.; Watanabe, Y.; Hermjakob, H.; MacLean, B.; MacCoss, M. J.; Zhu, Y.; Ishihama, Y.; Vizcaíno, J. A., The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Research* **2019,** *48* (D1), D1145-D1152.

41. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A., The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **2019,** *47* (D1), D442-d450.

42. Perez-Riverol, Y.; Bai, J.; Bandla, C.; García-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D. J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; Walzer, M.; Wang, S.; Brazma, A.; Vizcaíno, J. A., The PRIDE database resources in

2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* **2022,** *50* (D1), D543-d552.

43. Park, D. D.; Phoomak, C.; Xu, G.; Olney, L. P.; Tran, K. A.; Park, S. S.; Haigh, N. E.; Luxardi, G.; Lert-itthiporn, W.; Shimoda, M.; Li, Q.; Matoba, N.; Fierro, F.; Wongkham, S.; Maverakis, E.; Lebrilla, C. B., Metastasis of cholangiocarcinoma is promoted by extended high-mannose glycans. *Proceedings of the National Academy of Sciences* **2020,** *117* (14), 7633-7644.

44. Radhakrishnan, P.; Dabelsteen, S.; Madsen, F. B.; Francavilla, C.; Kopp, K. L.; Steentoft, C.; Vakhrushev, S. Y.; Olsen, J. V.; Hansen, L.; Bennett, E. P.; Woetmann, A.; Yin, G.; Chen, L.; Song, H.; Bak, M.; Hlady, R. A.; Peters, S. L.; Opavsky, R.; Thode, C.; Qvortrup, K.; Schjoldager, K. T.-B. G.; Clausen, H.; Hollingsworth, M. A.; Wandall, H. H., Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proceedings of the National Academy of Sciences* **2014,** *111* (39), E4066-E4075.

45. Boyaval, F.; Dalebout, H.; Van Zeijl, R.; Wang, W.; Fariña-Sarasqueta, A.; Lageveen-Kammeijer, G. S. M.; Boonstra, J. J.; McDonnell, L. A.; Wuhrer, M.; Morreau, H.; Heijs, B., High-Mannose N-Glycans as Malignant Progression Markers in Early-Stage Colorectal Cancer. *Cancers (Basel)* **2022,** *14* (6).

46. Wuhrer, M., Paucity of Paucimannosylation Revoked. *PROTEOMICS* **2019,** *19* (21-22), 1900244.

47. Thaysen-Andersen, M.; Venkatakrishnan, V.; Loke, I.; Laurini, C.; Diestel, S.; Parker, B. L.; Packer, N. H., Human Neutrophils Secrete Bioactive Paucimannosidic Proteins from Azurophilic Granules into Pathogen-Infected Sputum *². *Journal of Biological Chemistry* **2015,** *290* (14), 8789-8802.

48. Loke, I.; Østergaard, O.; Heegaard, N. H. H.; Packer, N. H.; Thaysen-Andersen, M., Paucimannose-Rich N-glycosylation of Spatiotemporally Regulated Human Neutrophil Elastase Modulates Its Immune Functions*. *Molecular & Cellular Proteomics* **2017,** *16* (8), 1507-1527.

49. Chatterjee, S.; Lee, L. Y.; Kawahara, R.; Abrahams, J. L.; Adamczyk, B.; Anugraham, M.; Ashwood, C.; Sumer-Bayraktar, Z.; Briggs, M. T.; Chik, J. H. L.; Everest-Dass, A.; Förster, S.; Hinneburg, H.; Leite, K. R. M.; Loke, I.; Möginger, U.; Moh, E. S. X.; Nakano, M.; Recuero, S.; Sethi, M. K.; Srougi, M.; Stavenhagen, K.; Venkatakrishnan, V.; Wongtrakul-Kish, K.; Diestel, S.; Hoffmann, P.; Karlsson, N. G.; Kolarich, D.; Molloy, M. P.; Muders, M. H.; Oehler, M. K.; Packer, N. H.; Palmisano, G.; Thaysen-Andersen, M., Protein Paucimannosylation Is an Enriched N-Glycosylation Signature of Human Cancers. *PROTEOMICS* **2019,** *19* (21-22), 1900010.

50. Miyamoto, S.; Stroble, C. D.; Taylor, S.; Hong, Q.; Lebrilla, C. B.; Leiserowitz, G. S.; Kim, K.; Ruhaak, L. R., Multiple Reaction Monitoring for the Quantitation of Serum

Protein Glycosylation Profiles: Application to Ovarian Cancer. *Journal of Proteome Research* **2018,** *17* (1), 222-233.

51. Ashwell, M.; Guo, X.; Sinnott, M. L., Pathways for the hydrolysis of glycosides of N-acetylneuraminic acid. *Journal of the American Chemical Society* **1992,** *114* (26), 10158-10166.

52. Ito, S.; Asahina, Y.; Hojo, H., Investigation of protecting group for sialic acid carboxy moiety toward sialylglycopeptide synthesis by the TFA-labile protection strategy. *Tetrahedron* **2021,** *97*, 132423.

Supplemental Methods

Standard Peptide Analysis

Sialoglycopeptide standard was purchased from TCI America (S0523). Dried peptide was reconstituted to 0.2 ug/uL in ultra-high purity water with the addition of 0.1% FA prior to analysis. These experiments were run on a fresh PGC column, prepared as stated above, in order to avoid any competing analytes from obscuring our measurements. Given the simplicity of the mixture and the propensity for PGC to retain small, hydrophilic analytes, the LC gradient was constructed as follows: analytes were trapped on the column for 15 minutes at 0% buffer B followed by a 25-minute gradient from 20% to 45%B. The column was washed with 95% B for 10 minutes and equilibrated to 0% B for ten minutes.

Samples were detected on a Q-Exactive Orbitrap mass spectrometer with the following settings. MS1 resolution, 70,000; AGC target, 1e6; maximum injection time, 250ms; m/z range, 300-2,000. MS2 resolution, 35,000; AGC target, 2e5; maximum injection time, 120ms; loop count, 10; isolation window, 2m/z; fixed first mass, 100; NCE, 30.

Database Searching

Raw MS files were processed using Proteome Discoverer 2.5 with Protein Metrics Byonic as the engine. Datafiles belonging to the bovine standard glycoprotein digests were searched against a custom fasta database containing the respective protein sequences from Uniprot. The prostatic cancer cell lines were searched against the Uniprot Human proteome (as found on October 30th, 2021). Both standard glycoproteins and complex cell digests were searched using the following parameters. Cleavage enzyme, trypsin; allowed missed cleavages, 2; precursor mass tolerance, 25ppm; fragment mass tolerance, 20ppm; fragmentation type, HCD; total common

modifications, 4; total rare modifications, 1; MS/MS diagnostic peaks, HexNAc [H^+], 204.087. Modifications were set as follows: fixed modifications - carbamidomethyl of Cysteine; variable modifications - oxidation of Methionine, deamidation of Asparagine and Glutamine, acetylation of N-terminus. The built-in 309 mammalian N-glycans were set as rare modifications. The only other notable workflow inclusion is that Proteome Discoverer was instructed to keep all peptide results rather than those determined as 'high' confidence. FDR cutoff was set to 1% and glycopeptide data were filtered to only contain those species with a score > 150, delta_mod > 10, and a log_prob > 1. During data analysis we also filtered positive results to those species identified in at least 2/3 replicates in each dilution at each temperature.

Supplemental Tables

For brevity, the tables referred to in the manuscript have been omitted. These tables may be found alongside the online version of this manuscript (10.1021/jasms.2c00249).

Supplemental Figures

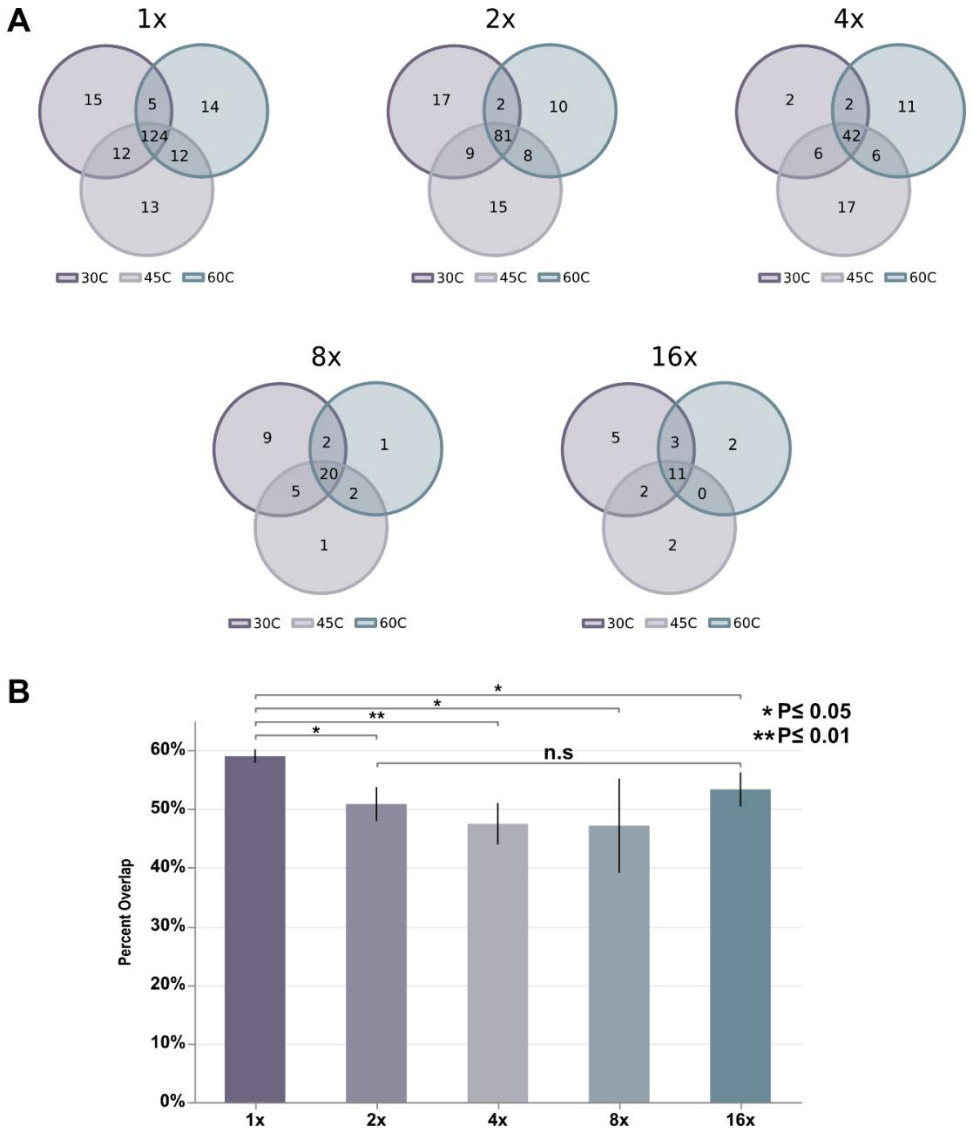


Figure S3.1 Comparisons of glycosite overlap across dilutions.

A) Venn diagrams displaying the overlap of identified glycosites for each glycopeptide dilution. For all dilutions, there are some glycosites unique to a specific temperature, highlighting the additional coverage associated with column temperatures. **B**) Relative fraction of glycosites found at all temperatures within each dilution. While the highest concentration sample (1x) showed statistically significant difference between all other dilutions, the four latter dilutions were found to not be significantly different in their relative overlap.

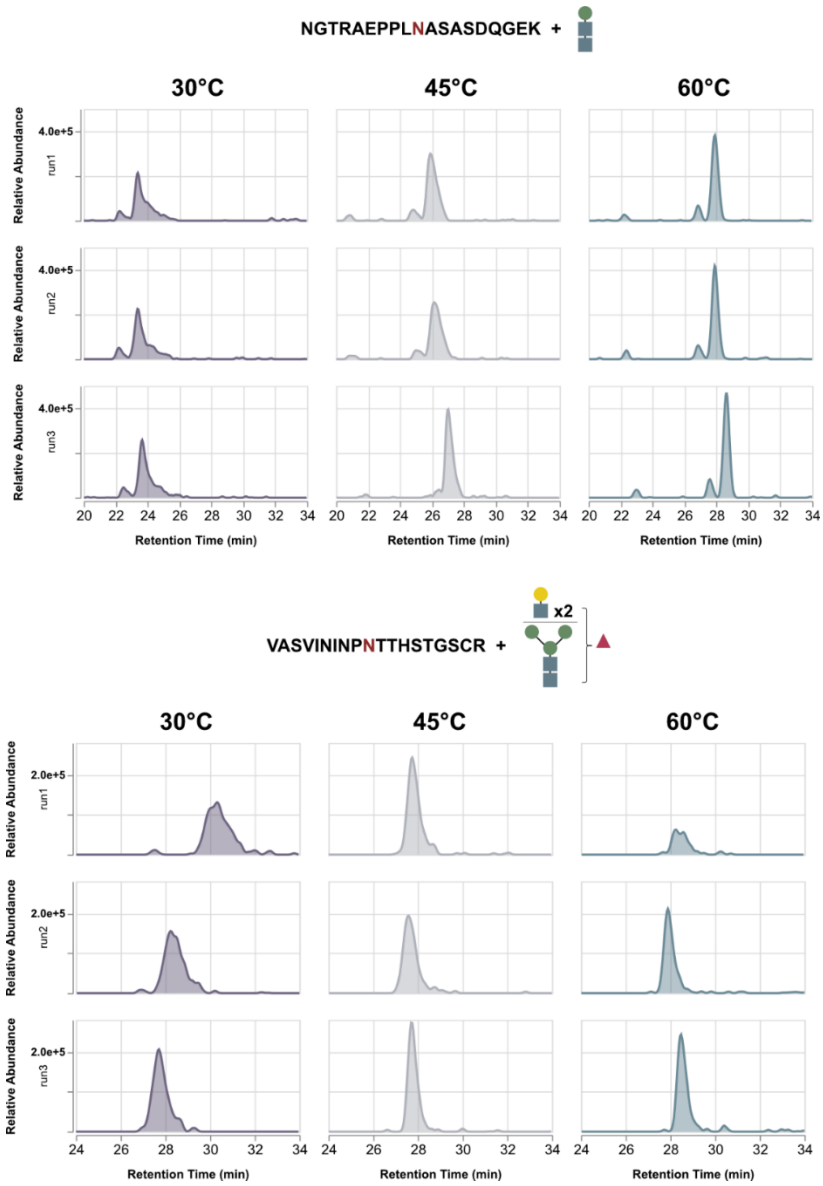


Figure S3.2 Representative XICs for paucimannose and fucosylated glycopeptides.

As shown, paucimannose glycopeptides generally benefit from increases in column temperature while fucosylated glycopeptides mirror the effects seen in sialoglycopeptides – increasing intensity at 45°C with reduced intensity at 60°C.

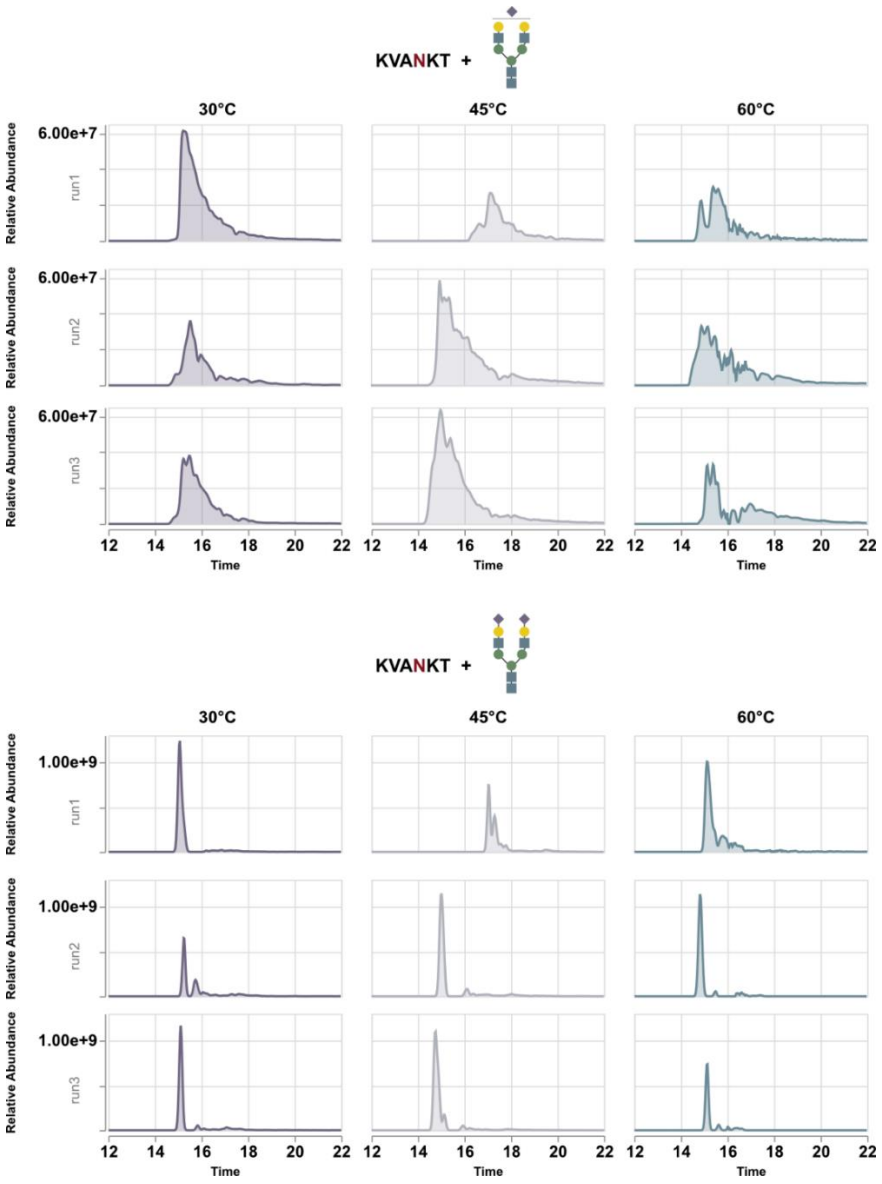


Figure S3.3 Extract ion chromatograms (XICs) of a commercial sialoglycopeptide.

These data serve to highlight that no early dissociation of sialic acid was immediately obvious. While the commercial product is listed as being biantennary with two terminal sialic acids, all analyses, regardless of temperature, were able to identify the singly sialylated form. If de-sialylation was a result of temperature increase, we would expect to observe i) no singly sialylated peptide at lower temperatures and ii) an increasing intensity of the singly sialylated form.

Neither of these observations is apparent, indicating further study is needed to determine whether early dissociation is the primary contributor to lower intensity of sialylated glycopeptides.

Chapter 4: Recent Advances in Analytical Approaches for Glycan and Glycopeptide Quantitation

Delafield, D. G., & Li, L. (2021). Recent Advances in Analytical Approaches for Glycan and Glycopeptide Quantitation. *Molecular & Cellular Proteomics*, 20.

Abstract

Growing implications of glycosylation in physiological occurrences and human disease have prompted intensive focus on revealing glycomic perturbations through absolute and relative quantification. Empowered by seminal methodologies and increasing capacity for detection, identification, and characterization, the past decade has provided a significant increase in the number of suitable strategies for glycan and glycopeptide quantification. Mass spectrometry-based strategies for glycomic quantitation have grown to include metabolic incorporation of stable isotopes, deposition of mass difference and mass defect isotopic labels, and isobaric chemical labeling, providing researchers with ample tools for accurate and robust quantitation. Beyond this, workflows have been designed to harness instrument capability for label-free quantification and numerous software packages have been developed to facilitate reliable spectrum scoring. In this review, we present and highlight the most recent advances in chemical labeling and associated techniques for glycan and glycopeptide quantification.

Introduction

Continuous developments of analytical strategies enable advancements that illuminate the roles in which post-translational modifications (PTMs) act to influence organism maturation, physiological processing, and immune response. While all members of this class of protein decorators are recognized for their alteration of protein function and contribution to proteomic diversity[1], no PTM is considered as complex or highly dynamic as that of glycosylation[2]. The downstream products of enzymatic construction and deposition of carbohydrate moieties—glycans—onto a nascent backbone[2], glycoproteins present significant challenges in analysis due to their high degree of structural and compositional complexity[2], ionization inefficiency[3], low abundance[4], and the unique phenomena of macro- and microheterogeneity[2]. Mass spectrometry (MS)-based glycomics has benefited greatly from advances in sample preparation protocols[5], enrichment strategies[6-8], and instrumental capabilities (fragmentation, data dependent and data independent acquisition, parallel reaction monitoring etc.)[9, 10], which now provide broad access to the glycoproteome.

As a result of these advances, targeted glycomic research continuously expands the implication of glycosylation in physiological processes such as cell signaling[11-14], host-pathogen interaction[15-18], and immune response[11, 19-21], with significant revelations provided in connection to human disease. Recent evaluations demonstrating the importance of glycosylation in neurodegenerative diseases[22, 23], diabetes[24, 25], and cancer[26-28] promote further interest in glycomic investigation to reveal potential biomarkers and unambiguous symptomatic protein profiles. As focus shifts from glycomic discovery and characterization to that of glycan expression levels and minute perturbations in site occupancy, the need for robust and efficient glycan and glycopeptide quantitative strategies steadily grows. In response to this demand, the last decade has seen a surge

in reports detailing novel chemical-labeling-based and label-free strategies built on both data dependent and data independent acquisition for quantitative glycomics (Figure 4.1). The previous review by Mechref *et al.*[29] provides a detailed discussion of the seminal reports paving the way for recent innovations, which may be explored in supplement to the strategies outlined herein. Discussed below are the most recent advances in metabolic incorporation, isotopic and isobaric chemical labeling, label-free approaches, and software for quantitative glycan and glycopeptide analysis.

Glycan Quantitation

As glycoconjugate function is shown to be impacted by glycan structure and composition, enzymatic or chemical release of glycans provides direct access to profiling altered glycan expression while enabling structural and compositional characterization. Considering the ever-present challenges in glycan analysis such as ionization inefficiency, highly hydrophilic character, glycosidic bond lability, and presence of negative charge, effective glycan quantitation may be achieved through strategies that offer reprieve from these ailments while providing facile labeling and reduction in spectral complexity.

Isotopic Labeling

Glycan quantification at the MS^1 level is an attractive prospect due to broad access to higher resolution instrumentation and the reduced considerations of selectivity bias in data-dependent acquisition experiments. Relative quantitation in this manner is often achieved through labeling of glycans in “heavy” and “light” channels to produce a consistent mass difference (i.e., > 1 Da). In order to avoid retention time differences between constituents of each channel and increase quantitative accuracy, heavy and light labels are engineered through the incorporation of stable isotopes, such as ^{12}C and ^{13}C. 2-aminobenzoic acid (2-AA) is a classic glycan label, often

employed for its fluorescent properties in UV-based experiments[30] and was adapted for isomer-specific quantitative glycomic evaluations[31]. The well-characterized labeling strategy, commercial availability of isotopologues and complete separation of isotopic envelopes—necessary to avoid peak overlap and inaccurate quantitation—make this strategy well-suited for facile quantitation. The importance of envelope separation was reinforced in the preliminary report of glycan reductive isotopic labeling (GRIL)[32], which employed aniline isotopologues to stabilize sialic acid linkages, eliminate negative charge, and distinguish isotopic envelopes. GRIL was later employed for glycan analysis through porous graphitic carbon (PGC) LC-MS, which enabled liquid-phase resolution of biantennary sialylated glycans[33, 34]. CID fragmentation was shown to provide antennae-specific fragmentation, further indicating the ability to quantify differential expression of isomeric glycans. Additionally, Walker, *et al*. established a method labeling glycan with isotopic hydrazide tags[35], INLIGHT[36], which echoes the importance of envelope separation to eliminate inaccurate correction or quantitation. This method was validated against glycan standards and those extracted from human plasma, demonstrating quantitative accuracy across 4 orders of magnitude.

As an alternative to carbon isotopes, glycans may be labeled with heavy oxygen (^{18}O) when enzymatic release is performed in the presence of heavy water. First reported by Tao and Orlando[37], the mechanism of glycan release with PNGase F results in a terminal amine group at the glycan reducing end, which is then replaced with a hydroxyl group after spontaneous hydrolysis. When released in heavy water, glycans will express a 2 Da mass shift over unlabeled counterparts. This method has been further applied[38] and is advantageous in that it requires no synthesis or treatment with commercial isotopologues, and that labeling efficiency is at or near 100%, depending on the purity of heavy water available. However, considering sample complexity

and the unavoidable overlap of isotopic envelopes when labeled/unlabeled pairs are separated by only 2 Da, Cao and colleagues developed a strategy for glycan reducing end dual isotopic labeling (GREDIL), which provided an additional 1 Da mass shift through $NaBH_4$/$NaBD_4$ reduction of glycans[39].

Beyond heavy carbon and oxygen, the incorporation of deuterium has been widely reported in quantitative glycomics experiments. As glycan permethylation[40] is routinely employed to reduce the high hydrophilicity of glycans and increase ionization efficiency prior to LC-MS analyses, early reports demonstrate simple workflow adaptation using iodomethane isotopologues to produce three labeling channels through light, medium, and heavy methyl labels (i.e. CH_3, CD_2H, CD_3)[41]. The same research group later expanded this workflow into an 8-plex labeling strategy that included additional heavy carbon isotopes[42]. Early reports of deuterium-based isotopic tags were provided by Zaia and colleagues, who first assessed multiple novel compounds for tetraplex labeling[43] and later applied them for glycan and glycosaminoglycan quantitation[44]. Numerous other deuterium-based isotopic labeling strategies have been employed: derivatization with phenyl-methyl-pyrazole (PMP) has been employed to produce a one-pot dual-channel labeling strategy for MALDI-based quantitation of O-glycans[45, 46], which was also adapted for in-gel labeling without significant sample loss[47]; stabilization and quantitation of sialic acid-containing glycans was promoted through a solid-phase p-toluidine labeling strategy[48]; duplex stable isotope labeling (DuSIL) was developed to discriminate neutral and sialylated glycans without the need for synthesis [49-51]; isomer-specific quantitation of sialic-acid containing glycans was achieved through Glycoqueing, which enabled sialoglycan stabilization, isomer-specific elution order, and boosted MS signal[52]; and quantitation by mutant enzyme reaction stable isotope labeling (QMERSIL) facilitated glycan release and labeling in a single step[53]. Other methods for MS^1 level

quantitation are reported by Yang and colleagues, who employed a metal chelating agent (p-NH_2-Bn-DOTA) and rare earth metals to provide a 10 Da mass shift and near 100% labeling efficiency[54], and the quantification of N-glycan types presented by Li *et al.* that couples endoglycosidase digestion with channel labeling to provide an enrichment-friendly three-plex labeling strategy composition[55].

Due to the significant sample handling necessary for glycan purification, derivatization, labeling and cleanup prior to electrospray ionization (ESI)-based MS experiments, Chen *et al.* conceived a strategy that leverages the salt-tolerant, facile nature of matrix-assisted laser desorption/ionization (MALDI)-based glycan analysis while eliminating the ion suppression that stems from sample complexity. Combining glycans after labeling with light/heavy HDEAT (2-hydrazino-4,6-bis-(diethylamino)-*s*-triazine)—which provides a 20 Da mass shift between species, HILIC separation was employed to deliver a liquid trace onto a MALDI plate. After matrix application, the liquid trace could be analyzed directly to identify N-glycans. The spatial distribution of glycans on the MALDI plate could be reconstructed into a base peak chromatogram to provide retention time of glycan species. This method reports significantly improved performance for glycan quantitation with higher sensitivity, reproducibility, and accuracy compared to MALDI alone and may be further expanded to multiplexed experiments[56].

Of particular note are strategies that reduce sample handling and associated loss by employing cellular machinery to facilitate glycan labeling, combining features of both metabolic and isotopic labeling. A pioneering study of this kind was provided by Kudelka *et al.* who introduced cellular O-glycome reporter/amplification, CORA[57]. This methodology involves the supplementation of cell culture media with paracetylated benzyl-α-N-Acetylgalactosamine (GalNAc-Bn), which is extended into a mature glycans by endogenous glycosyltransferases.

Because the reducing end of the glycan is blocked by the benzene group, these glycans are not acted on by oligotransferases, rather being excreted from the cell and escaping degradation. The benzene group also facilitates simple purification using reversed-phase cartridges for efficient MS analysis of the O-glycome constituents. This method was further developed to enable relative quantitation by employing light/heavy GalNAc-Bn in the method dubbed ICORA, isotopic labeling with cellular O-glycome reporter/amplification[58] (Figure 4.2). Highlights of this method include complete discrimination of isotopic envelopes through a 7 Da mass shift, high-levels of persistence found in Bn-protected glycans, and the ability to evaluate O-glycome perturbations in response to altered growth conditions. Though this method does not mitigate any of the challenges in glycan analyses (e.g. MS/MS of low abundance species, structural assignment, or accuracy of MALDI vs. ESI) and is not broadly useful beyond MS due to the weak absorbance of the benzene ring, this method does provide a rigorous example of how "classic" metabolic incorporation of stable isotopes and azide sugars may be employed for glycan amplification and quantitation—an idea expanded much further in quantitative glycopeptide experiments (vide infra).

Metabolic Incorporation

Though isotopic labeling is successfully employed for MS^1-level comparison of glycans, the questions of labeling efficiency, as well as reagent synthesis, cost, and availability remain. As an alternative, several researchers have turned to the classic strategy stable isotopic labeling of amino acids in cell culture (SILAC), which significantly reduces concerns over labeling efficiency while retaining the ability to perform relative quantitation and offering a means to discern glycome lifetime and stability.

IDAWG, isotopic detection of amino sugars with glutamine[59], is one of a few seminal reports on the feasibility and accuracy of metabolic incorporation for relative quantitation. Though

discussed in depth in the previous review[29], briefly, heavy nitrogen was introduced to cell culture in the form of ^{15}N-glutamine, which provided near complete labeling of glycosylation sites and aminosugars across the observed proteome. This method demonstrates the reliability of metabolic incorporation for glycosite and glycan quantification, as well as how media treatment can be used to evaluate further synthesis or degradation of aminosugar-containing glycans in response to cellular behavior. This idea was further expanded by two groups who sought to comprehensively quantify the glycome and glycoproteome through combining metabolic incorporation and isotopic labeling. Yang *et al.* accomplished characterization of bladder cancer cell lines (KK47, YTS1, J82, T24) against a normal bladder mucosa cell line (HCV29)[60]. This report employed SILAC labeling for proteomic quantification while combining lectin microarrays and sialylated glycan derivatization with heavy/light aniline to comprehensively quantify glycan expression levels. Further expansions of combinatorial methods is provided in the report of solid phase extraction of N-linked glycans and glycosite-containing peptides (NGAG) by Zhang and colleagues[61]. This method employs sequential elutions after tryptic peptides have been complexed with aldehyde-functionalized resin beads. In the first pass, lysine side chains are protected through guanidination prior to derivatization of acidic species (sialic acid and aspartic acid) with aniline, which is followed by PNGase F treatment to release N-glycans. The released glycans were then labeled with iTRAQ, isotopic tags for relative and absolute quantitation, prior to LC-MS identification and quantitative analysis. In the second pass, the newly formed aspartic acid residues that result from glycan release are then cleaved by Asp-N treatment, eluted, and quantified after combining with heavy-labeled glycosite-containing peptides from SILAC treatment. Using the NGAG method to analyze OVCAR-3 Cells, 85 unique glycan compositions and 2,044 glycosite-containing peptides were identified, offering complementary coverage to that of the previously reported SPEG

methodology[62] of the same group. These methods present an efficient strategy for quantifying the glycome and glycoproteome through metabolic incorporation of stable isotopes, providing an avenue of expansion which has since been greatly explored in quantitative glycopeptide experiments. However, given the lack of suitable stable isotopes that may be incorporated and the increasing spectral complexity when numerous isotopes are present, these mass-difference experiments are fundamentally limited by the number of channels that may be analyzed at any one time. As such, great benefit may be found in employing the strategy of mass defect-based chemical labels.

Mass Defect

While isotopic labeling and metabolic incorporation impart a mass shift of > 1 Da—a mass difference, mass defect-based strategies impart a mDa mass shift. As such, MS^1 mass spectra are significantly less complex than in mass difference experiments, redundant sampling is avoided because all labeled ions are selected for fragmentation in the same MS^2 isolation window, and quantification at the MS^1 level is retained, reducing the concerns over precursor co-isolation. Early reports of such strategies using CH_3I and CH_2DI have been reported[63, 64], but few reports exist over recent years. One example provided by Chen *et al.*, was the successful application of mass defect dimethyl pyrimidinyl ornithine (DiPyrO)[65], an amine reactive tag, for quantitative glycomics[66] (Figure 4.3). This study successfully quantified glycan expression differences between B-cells of healthy and acute lymphoblastic leukemia and demonstrated dynamic linearity across two orders of magnitude. This study provides two notable observations: i) increasing instrumental resolution will facilitate immediate expansion of DiPyrO beyond three demonstrated labeling channels and ii) employing amine reactive tags for glycan quantitation is a promising path that can be widely explored. This latter notion was explored by Feng *et al.* in the development of mass-defect isobaric

multiplex labeling reagents for carbonyl-containing compound (mdSUGAR) tags[67]. This three-channel approach is built upon the simple three-step synthesis of the original SUGAR tags[68] (see below), providing a 23.8 mDa mass shift between channels and labeling at both the reducing end and on sialic acid residues for stabilization. Beyond the significant reproducibility demonstrated when analyzing standard and complex samples, the MS^2 fragmentation spectra revealed complete *y* glycan fragment series with the mdSUGAR tag attached with additional tagged *b* ions found in sialylated glycans. This improved fragmentation series compared to unlabeled species allows for greater confidence during glycan annotation and structural assignment. These approaches represent a facile strategy for glycan labeling, with excellent accuracy and dynamic range that can be employed in scenarios where instrument resolving power is limited. Further expansion of these tags may prove useful in highly multiplexed experiments that seek to exploit rapidly evolving capabilities of novel instrumentation.

Isobaric Labeling

In order to avoid explosions in spectral complexity and the need for slower, higher-resolution MS^1 scans, numerous reports have explored the utility of glycan quantitation at the MS^2 level. By employing isobaric labels—each of which has an identical overall mass but a reporter ion region of unique mass—collision-based dissociation allows for relative quantitation to proceed through the comparison of reporter ion abundance.

At the time of last review, isobaric labeling strategies for glycan quantitation were only just emerging. iART, isobaric aldehyde reactive tags, was an early report of MS^2-based quantitation, employing a simple synthesis strategy to create two labeling channels. This method demonstrated significant improvements in glycan sensitivity post-derivatization as well as reliable quantitation when applied to quantifying the gp120 subunit of the HIV envelope[69]. The same

researchers later expanded this underlying strategy in developing a four-plex labeling strategy by developing quaternary amine containing isobaric tag for glycans, QUANTITY[70]. This method was originally validated using N-glycans released from human serum and CHO cell lines, which revealed relative quantitation of 90 and 159 N-glycans, respectively. Later, QUANTITY was employed for simultaneous quantitation of N- and O-glycans through sequential release and labeling techniques[71]. Concurrent with these studies, numerous strategies were established for glycan quantitation using commercial tandem mass tags (TMT). Though glycoTMT, a carbonyl reactive tag for N-glycan quantitation, was reported early[72], broad applicability was demonstrated through the use of the amine reactive tags, aminoxyTMT[73-77]. Notably, Zhong *et al.* demonstrated baseline resolution of TMT-labeled high-mannose glycans through capillary electrophoresis, while CE-TWIM-MS (capillary electrophoresis-traveling wave ion mobility mass spectrometry) was able to distinguish isomers of sialylated O-glycans in human milk[74]. Chen *et al.* later established the improved quantitative accuracy of N-glycans using MultiNotch MS^3 triggered by the presence of Y_1 glycan ions[76]. These recent reports indicate the utility of isobaric labeling for deep glycomic quantitation; however, the inefficiency of multi-step syntheses presented by iART and QUANTITY, as well as the high cost of commercial TMT labels, often place these workflows out of reach. In remedy to this, Li and coworkers developed Isobaric Multiplex Labeling Reagents for Carbonyl-Containing Compound (SUGAR) tags[68] (Figure 4.4). This report details a simple, three-step synthesis of SUGAR isotopologues with ~70% overall yield, and two-step labeling for near 100% labeling efficiency of all N-glycans tested. As well, the low cost of the reagents employed make this an attractive strategy that may be readily implemented in numerous research settings. Finally, in addition to the efficiency and quantitative accuracy, SUGAR tags demonstrated significantly improved glycan fragmentation in CID/HCD-based experiments for more accurate

structural and compositional assignment. Considering these numerous developments over recent years, isobaric labeling is seen as an effective strategy for glycan quantification, which is likely to be further expanded with improvements in instrument resolution and need for increased sample throughput.

Fluorescence Labeling

Fluorescence and absorbance-based labeling strategies were methods of significant interest prior to the heavy development of MS-based technology and MS-suitable sample preparations. However, fluorescence labeling is still employed due to the relative ease of glycan derivatization, the reduced need for intensive sample cleanup, and the reduction of sample loss via reduced sample handling. A notable improvement in glycan labeling efficiency was reported by Lauber *et al.*, where they demonstrated commercial RapiFluor-MS can label glycans in under 5 minutes[78] compared to the >1 hour found strategies mentioned above. RapiFluor-MS also facilitated quantitative recovery of glycans during cleanup, facilitated sensitive fluorescence, and quantitative accuracy in ESI-MS experiments. In the effort to reduce the limitations surrounding single-channel measurements of fluorescence-based strategies, Rana *et al.* developed a three-channel sensing system that employs unique fluorescent proteins to generate a multiplex output[79]. Utilizing gold nanoparticles with a glycan recognizing functional ligand, this strategy proved useful in rapidly and quantitatively comparing human cell types according to their surface glycan profiles.

Label Free

Rapidly evolving instrumental capabilities present a unique path towards quantitative glycomics. An ideal approach to quantitative experiments is the incorporation of an internal standard, but this method is not widely employed due to the complexity of glycan synthesis and the lack of commercial isotopic glycan standards. iGlycoMab, an isotope labeled monoclonal

antibody, was recently developed through ^{15}N metabolic incorporation. As heavy nitrogen will be incorporated into the aminosugars of the single glycosylation site on the Fc region, glycans released from this standard protein can serve as an internal isotopic standard in glycomics experiments. This strategy was successfully employed by Zhou and others, indicating the feasibility of isotopic glycans as internal standards[80]. An alternative strategy using the incorporation of exogenous standards was also validated for glycan quantitation[81]. As data suggests that molar responses for permethylated glycans are relatively uniform, investigators spiked in permethylated malto-series glycans at known concentrations for absolute quantification of N-glycans. These two previous reports are unique strategies for glycan quantification, but both state the need for a complete N-glycan standard series for more accurate, reliable, and broadly useful experiments. Given the unavailability of isotope-encoded glycan standards, a premium is placed on methods capable of accurate quantitation while reducing dependence on internal standards. To this end, numerous reports have validated significant increases in analytical sensitivity and quantitative accuracy when employing parallel and multiple reaction monitoring.

MS Reaction Monitoring

With rapidly expanding access to instrumentation capable deciphering highly complex mixtures, alongside the appreciation of reliable and reproducible instrument performance, a growing number of investigators have sought to exploit instrument capabilities for absolute and relative quantitation. Rapidly gaining favor in the area of glycan analysis are select, parallel, and multiple reaction monitoring (SRM, PRM and MRM). Though each has been successfully employed for glycomic quantitation MRM analyses has gained favor in glycoproteomics[82] due to more precise quantitation[83], high analytical reproducibility, better signal-to-noise, and increased dynamic range[84]. Though in-depth description of reaction monitoring concepts and considerations

may be read elsewhere[85, 86]. In brief, MRM, which is often implemented on triple quadrupole (QQQ) instrumentation, involves scanning of glycans in the first quadrupole, CID fragmentation in the second, and scanning of transitions (i.e. fragments of precursor masses) are scanned in the third. User control over valid precursor and transition masses results in a highly-selective and sensitive method for glycan identification. Noting that transition signal response is directly related to analyte concentration, iterative analyses of standard mixtures can be employed to develop calibration curves of transition abundance. After analysis of unknown sample mixtures, these curves are used to provide absolute abundance of targeted analytes. The targeted nature and considerable effort needed to establish effective MRM workflows limit their utility in high-throughput experiments, but are widely useful in glycan biomarker and protein characterization studies[85].

Of the numerous reports employing reaction monitoring, Lebrilla and colleagues have been instrumental in developing novel methods for MRM analysis of mono- and oligosaccharides. Of note, Hong *et al.* detailed the ability to perform label-free absolute quantitation of human milk oligosaccharides (HMOs) and leverage 2'-fucosylation concentration to profile samples from secretors and nonsecretors[87]. Of note, this method established quantitative accuracy across five orders of magnitude and displayed femtomole sensitivity, rearticulating the benefits of targeted MRM analyses. Later, Xu *et al.*, expanded on this approach and demonstrated that differences in retention time between monosaccharide isomers can be leveraged to create dynamic multiple reaction monitoring methods[88]—a concept discussed in detail in later sections. In addition to these fundamental reports, Xia *et al.* provided an early entry through their analysis of N- and O-glycans for diagnosis of congenital disorders[89]. Later Tao *et al.* reported a penta-HILIC-SRM-MS for the separation and identification of 2,3/2,6 sialic acid-containing N-glycan isomers[90] and Tsai *et al.*

established a protocol to for N-glycan biomarker discovery in hepatocellular carcinoma (HCC)[91]. MRM has also been used in combination with glycan permethylation to quantify 88 N-glycans from only 5 nL of human blood[92]. Finally, Orlando and coworkers have pursued absolute N-glycan quantitation from biotherapeutic antibodies[93] and Mank *et al.* expanded on the earlier reports of HMO analysis to provide structural selectivity[94]. These reports are among those that signal increasing interest in label-free, instrument-dependent methodologies for glycomic quantitation. Though the benefits and drawbacks of these strategies must be carefully weighed again those mentioned for chemical labeling, future innovations in the area of MS reaction monitoring and instrument efficiency could pave the way for a gradual shift towards confident label-free analyses.

Critical Evaluations and Considerations

Numerous strategies have been developed for glycan quantitation, presenting unique benefits and drawbacks that must be considered prior to implementation. A guiding consideration should include relative sample complexity and need for throughput. In low complexity experiments where throughput is not needed (i.e. analyzing no more than 2 samples), isotopic labeling is an effective strategy that may be customized to fit individual needs. Isotopic labeling reveals greater benefits when employing tags that increase glycan hydrophobicity and ionization efficiency or impart positive permanent charge. As sample complexity increases, mass-defect-based isotopic labeling strategies may be implemented to offer reprieve from precursor co-isolation and spectral complexity while also providing slightly higher throughput. In high-throughput investigative experiments, if samples are relatively simple and MS1 level quantification is possible, 8-plex glycan permethylation would be of use due to the significant increases in glycan hydrophobicity for LC separations and improved ionization efficiency. However, isobaric labeling is undoubtedly the method-of-choice in high-throughput, high-

complexity experiments as quantitation is pursued in tandem with identification at the MS2 level. If seeking to perform analyses at this level, channel multiplexing, synthetic capacity, and cost will be the guiding factors. A brief summary of highlighted methods may be found in **Table 1**. No matter the application, the techniques presented above provide achievable avenues to those seeking to perform quantitative glycomic analyses.

Glycopeptide Quantitation

Direct glycan analysis after enzymatic or chemical release enables understanding of the heterogeneity found within a given glycoproteome while providing the best opportunity for structural and compositional interrogation. In pursuit of comprehensive glycoprotein characterization, glycan analysis is limited by the elimination of protein localization as no glycan can be related to a modification site without intensive experimental control. To this end, analysis of intact glycopeptides retains site-specific information while enabling modest elucidation of the attached glycan. Though traditionally limited due to low abundance within proteolytic mixtures and poor ionization efficiency, glycopeptide analyses have benefited greatly from recent advances in sample preparation[95-97], enrichment strategies[98-100], and instrumental functionality[101-103]. Enabled by broad access to the glycoproteome, revealing deviations at the glycan, modification site, and protein level are of immediate interest in the effort to provide a more comprehensive view that helps to elucidate the role of glycosylation in physiological processes and human disease. As the following reports exercise analysis of glycosylated peptides and de-glycosylated peptides, clear distinction has been provided to avoid confusion. Discussion of "glycopeptides" refers strictly to glycosylated species and all references involving release of glycans prior to analysis are noted as "deglycosylated peptides."

Metabolic Incorporation

As SILAC experiments involving the incorporation of heavy amino acids—traditionally heavy lysine and arginine—during protein translation, glycopeptide quantitation through metabolic incorporation is widely accessible. This approach was taken in early reports that detailed the utility of data-independent acquisition (DIA) of sequential isolation windows (SWATH-MS) for glycopeptide quantitation[104]. DIA analyses will be discussed further in subsequent sections, but this initial report demonstrated the sensitivity and reproducibility gained during application. Further application of heavy amino acids was reported by Poljak *et al.* who used enzymatic cleavage and parallel reaction monitoring (PRM) of glycopeptides to quantitation the N-glycosylation machinery in yeast[105], though this method did not provide evaluation of glycan expression levels. While the applicability of incorporating isotopic labels is plainly seen across proteomics, significant contributions to glycopeptide analysis have come through the development of methods that combine efficient enrichment and complete labeling. Though the following methods enable quantitation through isotopic labels, they are presented here for their unique implementation of metabolic azide sugar incorporation.

Due to the facile, highly-selective nature of copper-catalyzed cycloaddition of terminal alkynes and azides[106, 107] (i.e. click chemistry), numerous groups have employed this reaction to label, enrich, and quantify glycopeptides. A benchmark study, isotope-targeted glycoproteomics (IsoTaG), demonstrated the ability to incorporate azide-containing sugars into nascent glycans[108-110]. This azide sugar was then ligated to an acid-labile, isotopically labeled biotin tag with a terminal alkyl group for glycopeptide enrichment with streptavidin beads. The biotin tag was then cleaved, leaving behind the isotopic group, which could then be used for targeted mass spectrometry due to the characteristic mass shift against isotopic partners. The combined

efficiencies of azide sugar incorporation and biotin-streptavidin enrichment presented a powerful strategy for quantitative glycomics and glycan/glycopeptide enrichment. Though this method has difficulty in complete characterization of N-glycans—due to the unpredictable composition of sialic acid-containing glycopeptides—the authors successfully elucidated 32 N-glycopeptides with an additional 156 partial assignments and complete characterization of more than 500 O-glycopeptides. The shortcomings in N-glycopeptide detection were addressed in a later study that incorporated alkyne-sugars rather than azido-sugars, which facilitated greater access to N- and O-glycopeptides alike with 156 and 578 confident identifications, respectively[111]. A key benefit of employing IsoTaG is the accompanying software, IsoStamp[112], that aids in spectral deconvolution and quantitation. Such benefits are replicated in the study from Qin, *et al.* that detailed O-glycopeptide analysis through isotope-tagged cleavable linker (isoTCL)[113] and quantitation using MaxQuant. Though quantitative accuracy was still achieved, manual confirmation of heavy/light pairs must be performed, bolstering the value of IsoTag and IsoStamp that eliminate the need for validation. Finally, in order to eliminate the harsh solution conditions associated with acid-labile chemical probes, a photocleavable biotin tag for O-GlcNAcylated glycopeptide quantification was developed by Li *et al.*[114]. This study localized 419 and 276 O-GlcNAcylation sites from sorafenib-sensitive and sorafenib-resistant HepG2 cells, respectively, 262 of which were not previously reported.

Isotopic Labeling

Following the trend seen in glycan analyses, isotopic labeling is a method-of-choice in glycopeptide quantitation due to the well-characterized nature of numerous peptide labeling strategies. As dimethyl labeling is a highly facile method for peptide derivatization and employs reagents that are not cost-preventative, numerous reports detail the utility of dimethyl labeling in

lower throughput relative glycopeptide quantitation experiments[115, 116]. Novel applications include the association of altered glycopeptide glycosylation profiles with pancreatic cancer[117], glycoproteomic profiling in triple negative breast carcinomas through analysis of deglycosylated peptides[118], quantitative comparisons of sialic-acid containing glycopeptides in human embryotic and neural stem cells[119], and employing deglycosylated peptides to determine changes in site occupancy rates between normal liver and hepatocellular carcinoma (HCC) liver tissues[120]. Further development of this strategy has been seen in the employment of diethyl labeling of glycopeptides[121-123], which reduces retention time differences and quantitation errors by replacing incorporating heavy carbon in place of deuterium.

Though chemical labeling strategies such as dimethyl labeling are facile in nature, reagent purity and labeling efficiency are persistent factors that reduce the overall efficiency and accuracy of glycomic quantitation. However, in search of avenues for isotope incorporation with high efficiency and no need for intensive synthesis, researchers have capitalized on the mechanism of proteolytic cleavage to incorporate more advantageous stable isotopes, such as ^{18}O. A novel strategy for ^{18}O stable isotope labeling (TOSIL) of deglycosylated peptides was presented by Liu *et al.*[124] and later adapted for use in complex samples[125]. By performing trypsin digestion in the presence of heavy water, the newly formed C-terminus will be labeled with two ^{18}O atoms. PNGase F treatment of the formed peptides will result in additional ^{18}O atom being incorporated during the transition of the Asn modification site to Asp. This strategy was employed for accurate quantitation of glycosylation profiles between innovator and biosimilar antibodies[126]. Though this method retains no glycan-specific information, the authors employed selective lectin enrichment prior to glycan release to generate glycopeptide subgroups to evaluate topical modification changes. Validated in comparisons of normal and HCC liver cells, this method demonstrated high

quantitative accuracy across the dynamic range and complete isotopic envelope separation. To evaluate the utility of the original TOSIL method for N-glycoproteome quantitation, Liu *et al.* employed TOSIL in tandem with lectin microarrays to reveal potential biomarkers in HCC metastasis[127].

In addition to these innovations, numerous groups have developed novel chemical tags useful for glycopeptide labeling, which are easily translated to quantitative experiments after synthesizing the deuterium isotopologue. For example, Kurogochi *et al.* employed benzoic acid N-succinimidyl ester to enhance ionization efficiency of glycopeptides in MALDI-based quantitative experiments[128], while Pabst *et al.* later determined galactosylation and sialylation patterns in Immunoglobulin G (IgG) glycopeptides in both ESI and MALDI regimes through derivatization with succinic anhydride[129]. As routine proteolytic digestion involves reduction of disulfide bonds and protection through alkylation, reports have detailed the utilization of these processes for direct peptide labeling. Kim and coworkers conceived the use of isotope-coded carbamidomethylation[130] to label deglycosylated peptide species in tandem with free thiol protection, while Zhao *et al.* employed isotopic dithiothreitol to label O-glycosylation sites after beta-elimination to produce deglycosylated peptides[131]. These are attractive methods for peptide quantification as it does not involve subsequent sample handling or cleanup beyond those used in routine digestion workflows. Validated strategies such as these provide a litany of facile labeling strategies for relative glycopeptide quantitation but are inherently limited by low channel number and the inability to facilitate absolute quantitation. In remedy, recent reports have demonstrated the utility of isotopic internal standard peptides for absolute quantitation and novel application.

First, Zhu *et al.* reported a strategy to determine absolute quantitation of glycosite occupancy in experiments deglycosylated peptide abundance to isotope-coded synthetic

peptides[132]. Noting deamidation of Asn residues is shown to occur spontaneously during sample preparation and therefore skew quantitative comparisons of deglycosylated peptides, this work synthesized isotopic deglycosylated peptide partners. This allowed site occupancy to be reliably quantified by subtracting the concentration of non-glycosylated protein from total protein concentration. Later, Roy *et al.* reported a strategy for absolute quantitation of IgG subclasses by synthesizing isotopic glycopeptides using Asn-GlcNAc residues that display no difference in retention time from glycopeptides produced during digestion[133]. This method could be readily expanded due to the flexibility of peptide synthesis and accuracy of internal standard calibration. Finally, Nilson *et al.* reported a method to quantify the recently reported amyloid-β (Aβ) glycopeptide as well as unmodified Aβ in cerebrospinal fluid[134]. Though Aβ glycopeptide contains a rare Tyrosine O-glycosylation (Tyr-GalNAc) and internal standards require intensive derivation prior to peptide synthesis, the report accurately quantifies differences in glycosylated Aβ-15 and Aβ-17 fragments across 20 Alzheimer's Disease patients and 20 non-demented controls. As synthetic peptide production evolves and the reliability and accuracy of multi and parallel-reaction monitoring strategies continue to improve, these reports are likely to serve as a basis for broad absolute glycopeptide quantitation.

Isobaric Labeling

The multiplexing capacity of isobaric peptide labeling provides a high degree of experimental accuracy and throughput in quantitative proteomic investigations. Traditional methods such as isobaric tags for relative and absolute quantitation (iTRAQ) and tandem mass tags (TMT) have been widely employed for glycoproteomic profiling in various biological samples. Among these, iTRAQ has been utilized for N-glycopeptide analyses in neurodegenerative diseases[135] and cardiac hypertrophy[136], analyzing glycopeptides to profile the glycoproteome of

human tear fluid[137], and interrogating deglycosylated peptides to reveal dynamic glycoprotein regulation in maize seedlings[138], representing the utility of iTRAQ to further glycomic experimentation. TMT has seen even greater utility in quantitative experiments as they have been applied to evaluate glycopeptide perturbations in HCC patient plasma[139], pancreatic cancer serum[140], aggressive prostate cancer cell lines[141] and urinary profiles of prostate cancer patients[142], human cell surfaces[143], cerebrospinal fluid[144] (glycopeptides and deglyscylated peptides), and PNGase F-resistant N-glycopeptides[145], as well as the evaluation of glycopeptide enrichment strategies[146] via direct analysis through ETD[147] and strategies for simultaneous phosphopeptide and glycopeptide quantitation[148]. iTRAQ and TMT are attractive strategies for those in seek of reliable relative glycopeptide quantitation, with added benefits of well-documented workflow, commercial availability and quality control. However, the steep cost of these commercial reagents makes them unsuitable for use during method development or exploratory studies and are not amenable to bespoke method alteration. Recently, a promising alternative was presented that allows for a significant reduction in cost, facile in-house synthesis, and a high degree of flexibility for method experimentation.

N,N-dimethyl leucine (DiLeu) isobaric tags were originally presented in 2010 by Xiang *et al.* as a novel 4-plex strategy for quantitative proteomics[149]. With commercial leucine as the starting material, each reporting channel is synthesized in no more than two simple reactions, each of which employs commonly available reagents—emphasizing cost-efficiency. Due to the comparable performance when evaluated against iTRAQ, DiLeu was expanded to a 5-plex platform for absolute quantitation (iDiLeu)[150] and an 8-plex relative quantitation strategy that maintained the overall ease-of-synthesis from the original report. DiLeu was further developed to facilitate 12-plex relative quantitation[151] , utilizing mass-defect principle and higher-resolution

instrumentation that is becoming more readily available, and this strategy was then coupled with dimethyl labeling, producing an effective 24-channel strategy for relative quantitation[152]. DiLeu isobaric labels have been evaluated in a number of proteomic and peptidomic experiments[150, 152-155], and has also been developed into an absolute quantitation strategy. Hybrid offset-triggered multiplex absolute quantification (HOTMAQ), combines 4-plex iDiLeu with 12-plex isobaric tags to create an internal calibration curve at the MS1 level in tandem with identification of peptides at the MS2 level[156] (Figure 4.5). This strategy provides up to a 12-fold increase in throughput during absolute quantitation experiments.

Of interest, DiLeu tags were recently applied for site-specific characterization and quantitation of N-glycopeptides in PANC1 pancreatic cancer and PKM2 knockout breast cancer cells[157]. As sialylated glycans are known to be upregulated in various cancers and show distinct expression across lifetime, this study provided an early report on the most efficient strategy for sialylated N-glycopeptide extraction and enrichment. Method validation in PANC1 experiments revealed 1067 N-glycopeptides, 311 glycosites, and 88 glycan compositions from 205 glycoproteins. Quantitative evaluations of PKM2 cells provided evidence that N-glycosylation signaling pathways are tightly regulated by cellular metabolism, with 484 N-glycopeptides quantified and 81 showing significant changes in expression. As this method offers comparable performance to the hallmark commercial methods of TMT and iTRAQ, as well as providing an avenue for mass defect-based proteomics[65], development and employment of DiLeu isobaric labels is a beneficial strategy for accurate, cost-effective proteomic and glycoproteomic quantitation with great room for further implementation.

Label Free and MS Reaction Monitoring

While a small number of reports detail the implementation of mathematical modeling to facilitate accurate, label-free quantitation of glycopeptides—such as that detailed by Mayampurath *et al.*[158]—glycopeptide quantitation has benefited greatly from the implementation of PRM and MRM. Similar to strategies implemented for glycan analyses, reaction monitoring of glycopeptides does offer high quantitative accuracy and improved sensitivity, but requires deeper consideration. MRM analysis requires effective ionization of glycopeptides and the production or reproducible, quantifiable fragments. As hydrophilic glycans reduce the overall ionization efficiency and the heterogeneity of glycosylation divides the intensity of glycopeptides across several glycoforms[86] enrichment strategies are often required to improve detectability against complex peptide backgrounds and avoid loss of minor glycoforms within the mixture[159]. However, these strategies have not prevented the successful implementation of MRM for numerous novel investigations. Of note, MRM has successfully quantified differential expression of IgG subclass glycosylation[160] , haptoglobin glycoforms[161, 162] and core fucosylation[163] in liver disease, profile changes in galactosylation and sialylation in rheumatoid arthritis (RA) patients[164], quantify glycoproteins in esophagus disease[165], reveal alterations in Murine immunoglobulin glycoforms[166], characterizing the function and importance of UDP-GlcNAc transporter[167], and quantitation of golgi-resident glycosylation enzymes from cultured human cells[168]. In addition, researchers have also detailed methods for glycopeptide quantitation in a range of human biofluids such as human serum[169, 170] and liver cancer plasma[171]. Pinpointing some standouts, Srikanth *et al.* provided a quantitative method that combines ^{18}O labeling and MRM, Jian *et al.* established the feasibility of top-down glycoprotein characterization when protein length is short[172], Hammura *et al.* detailed a method to both synthesize and quantify rare bisecting N-glycans in therapeutic antibodies[173], and van der

Burgt *et al.* implemented a strategy to quantify sialic acid linkage isomers of prostate specific antigen (PSA)[174]. The later also provides a topical comparison of various analytical methods for linkage isomer analysis on the basis of throughput, robustness, quantification ability, recognition of glycoforms, and isomer separation, which may be of interest to some readers.

In addition to these above reports, Lebrilla and colleagues have fundamental to the expansion of MRM glycopeptide analysis. Offering numerous reports of MRM analysis that identify and quantify immunoglobulin classes (i.e. IgG, IgA, IgM) and their glycosylation profiles[175, 176], as well as quantify site-specific glycosylation in recombinant antibody drugs[177], this group has also provided accurate quantitation of human milk protein glycoforms[178] and evaluated the differential expression of serum glycoproteins to serve as biomarkers in ovarian cancer[179]. On top of their numerous applications, this research group demonstrates the improvements found in implementing dynamic multiple reaction monitoring (dMRM). Though routine MRM analyses are highly-specific, minimizing the ailments surrounding coeluting peptides that may cause ion suppression and fail to identify low-abundance analytes, monitoring specific targets and transitions over the entire chromatographic timeframe severely reduces the number of analytes that may be quantified. As such, Li *et al.*[180] hypothesized retention time may be leveraged to reduce the time spent searching for selected precursor and transition masses, thereby increasing the number of novel species quantified. Employing multi-enzyme standard protein digestion to produce smaller glycopeptides and increase sample coverage, this strategy first employed orbitrap-based analysis of enriched glycopeptides that were identified by Byonic (vide infra). In addition to the identified glycopeptides, the authors imputed missing values for undetected species by generating in silico transition masses and predicting retention time according to the relative hydrophobicity of the glycopeptides. Using the retention times, precursor masses and unique transitions of all identified

and suspected analytes to build a dMRM method, the authors were able to quantify nearly 700 glycopeptides in a single 50-minute LC run, which was then validated on human serum samples. With low femtomolar limits of detection and quantification, this method illustrates the utility of MRM for complex sample quantification and the ability to accommodate higher-throughput. Taken together, the specificity, enhanced sensitivity, and uncompromised quantification accuracy of MRM is an attractive strategy for glycopeptide and glycoprotein quantitation with much room left for novel innovation and application.

As typical limitations in glycopeptide detection and identification include low concentration of glycopeptides within proteolytic mixtures and poor ionization efficiency, many glycopeptide species are overlooked and not selected for MS/MS fragmentation in data-dependent acquisition (DDA) experiments. For this reason, data-independent acquisition (DIA) has gained steady traction in broad proteomic and glycoproteomic experiments for its ability to expand profiling depth and select low-lying precursor masses, offering potential remedy to the low-throughput of MRM analyses[181]. Typical DIA experiments such as SWATH-MS (i.e. sequential window acquisition of all theoretical fragment ion spectra mass spectrometry) require user definition of *m/z* windows to be used for fragmentation. As most peptides are found within 400-1250 *m/z*, common practice is to set consistent window sizes (~25-36 *m/z*) over this range. However, due to the large mass addition of glycans, glycopeptides are not evenly distributed along this range and are concentrated between 950-1200 *m/z*. As such, Zhou *et al.* validated a more effective strategy, GP-SWATH, that narrows selection window width across the glycopeptide region to provide more accurate and robust glycopeptide detection and quantification[182]. A notable limitation in DIA analysis is the deconvolution of tandem MS spectra as DIA experiments commonly lose precursor information, making identification of post-translationally modified

peptides a challenge—especially for O-glycopeptides. Offering alleviation of this ailment, Ye *et al.* recently established Glyco-DIA, a strategy to provide enhanced O-glycopeptide identification and quantitation[183]. As illustrated in Figure 4.6, this method constructs spectral libraries from numerous DDA experiments, which can be expanded *in silico* to provide missing values. Evaluation of this methodology revealed significantly improved performance of O-glycopeptides in direct analyses with even greater benefit in runs performed after enrichment. Though the authors state limitations in this method such as biasing towards abundant O-glycopeptides in DDA experiments, Glyco-DIA may be rapidly expanded for O-glycoproteome coverage and tailored for individual, targeted analyses.

Software Advances

Accurate glycopeptide annotation is dependent on efficient glycan and peptide fragmentation, as the high compositional complexity of all glycans and the challenges in glycosite assignment of O-glycans can easily be misinterpreted and result in false identifications. Though few studies have evaluated the efficacy of decoy glycopeptide databases[184-187], numerous advances have been made in developing open source and commercial software capable of adept peptide annotation and quantitative comparisons. Premier Biosoft International provided early access into spectral deconvolution for glycan analysis. Touting a robust relational database of glycans and glycoproteins, support for MALDI and ESI file formats, glycopeptide qualitative analysis, built-in functionality to process TMT-based quantitative information, and the ability to assign glycan structure from MS^n data, SimGlycan remains a relevant and effective tool for glycomic investigation. Bern *et al.* (Protein Metrics) introduced Byonic in 2012 for peptide and protein identification, which remains a premier method for glycopeptide identification[188]. Following suit, Protein Metrics later introduced Byologic to facilitate an identification/quantitative analysis

pipeline which has been validated in a number of glycopeptide studies[189, 190]. As these licensed commercial software packages may be cost preventative and not widely employed by individual users, open source alternatives have been reported. LaCyTools[191] and GlycopeptideGraphMS[192] are python-based utilities that have reported improved glycopeptide identification and quantitation, while GPSeeker[116] facilitates structural N-glycoproteomics by integrating previously reported software from the same research group[193-195]. SugarQb[145, 196] was developed to provide glycan and glycopeptide insights within the Proteome Discoverer (Thermo) environment. An alternative to working within Proteome Discoverer is presented by Maxwell *et al*. in their development on GlycReSoft[197]. Building on their validated strategy for targeted glycan analyses, Manatee[198], GlycReSoft implements a data deconvolution algorithm to enable the rapid extraction and confidence scoring of glycan and glycopeptide identifications in both supervised and unsupervised analyses. In addition, GlycReSoft provides a user-friendly web-based application that can also leverage distributed computation to accommodate broad search space. The same research group later validated novel tools for increased glycomic profiling[199, 200], which utilized knowledge of biosynthetic pathways to improve glycan feature recognition. Finally, Integrated GlycoProteome Analyzer (I-GPA) enables global characterization of site-specific structural features and reliable, automated label-free quantitation[201].

One open-source alternative that has gained much attention is pGlyco[202] and its latest iteration, pGlyco 2.0[203]. As the initial software was a useful tool for glycopeptide spectra deconvolution, the authors state the need for expansion due to the existing limitations in high-throughput tools for peptide and glycan identification, the inability of current software to provide built-in manual interpretation and validation, and most notably, the lack of robust quality control and FDR estimation that drastically underperform in adjacent bioinformatic tools. The latter point

is echoed by Park *et al.*,[201] who provided topical comparisons of FDR approximations through GlycoFraqWork[204], GP Finder[205], Sweet-Heart[206], and GPS[207]. Further, as stepped collisional energy (SCE) dissociation was nascent at the time of publication but was shown to outperform single regime (i.e. CID, HCD, and ETD) and hybrid fragmentation modes (i.e. ETciD and EThcD), pGlyco 2.0 provided early access to using SCE for broad glycopeptide analysis. pGlyco 2.0 validated an improved FDR estimation through isotope-based and entrapment-based strategies. Complete details of these strategies may be read within[203], but performing database searches of the same data (i.e. yeast cell lysate digest) using pGlyco 2.0 provided <1% FDR while Byonic resulted in >19%, and every identification may be visually inspected in pGlyco 2.0 using the built-in gLabel software. In terms of raw performance, five mouse tissues (brain, heart, kidney, liver, and lung) were analyzed and subjected to pGlyco searching, which revealed 10,009 site-specific glycans on 1,988 glycosites from 955 glycoproteins with quantitation enabled through pQuant. pGlyco was then used to re-evaluate the previously discussed NGAG dataset[61] that used GPQuest as the search engine and revealed a 97% increase in glycopeptide identifications from the same data. Though pGlyco 2.0 was not heavy utilized for O-glycopeptide discovery, topical analyses of asialofetuin standard glycoprotein revealed reliable N- and O-glycopeptide identifications, indicating analytical potential. Taken together, pGlyco 2.0 presents a powerful, open-source option for robust glycopeptide identification.

Conclusions and Future Directions

The field of glycan and glycopeptide quantitation has experienced tremendous growth over the past decade. Widely accepted as an area of significant analytical challenge, the numerous creative strategies demonstrated above have proven successful as they directly address areas of topical concern in glycomic analyses. Ionization efficiency may be improved through glycan permethylation or by employing labels that increase hydrophobicity or impart permanent positive charge. The need for effective enrichment was addressed by methods that incorporate azide-containing sugars during cell culture for use in click chemistry labeling experiments. And instrumental functionality such as multiple reaction monitoring and data independent acquisition alleviate consequences of low glycopeptide abundance within a proteolytic sample. However, though these examples present significant advances in glycomic analysis, many improvements are still needed.

As pursuit of quantitative glycomics increases, researchers will be left searching for higher throughput methodologies and inevitably seek strategies for absolute quantitation. Methods presented above will lay the foundation for these new techniques, most likely seeing numerous strategies used in tandem, such as the workflow demonstrated in HOTMAQ[156]. Additionally, coverage of the glycome and glycoproteome will benefit from improvements in sensitivity. Lower- and nanoflow, chip-based technologies facilitate much greater signal response from glycan and glycopeptide species and are likely to be invaluable strategies moving forward. As well, capillary electrophoresis is likely to see greater implementation in glycomics investigations, owing to the extremely low sample consumption, ability to resolve isomeric mixtures and ultrahigh resolution. Alternatively, researchers may choose to boost glycan and glycopeptide abundance at the MS^1 level by using methods more amenable to the labeling strategy

employed, such as that shown in BASIL (boosting to amplify signal with isobaric labeling)[208]. Finally, computational tools and software capable of accurately deconvoluting and correctly assigning glycomic observations will be an area of continual need. Decoy database creation and implementation will see greater utilization as quantitative glycomics gains popularity, and resource bottlenecks (e.g. CPU processing speed and available cores) must be alleviated as access to the glycome increases.

Taken together, the field of quantitative glycomics is a space rich in invention, novel implementation, and discovery. Numerous labeling strategies have enabled facile, accurate investigations of disease-relevant glycoproteins and are well suited to uncover future biomarkers and discern symptomatic protein profiles. The developments in instrumental capability over the next several years are likely to provide greater expansion in chemical labeling experiments and possibly enable greater implementation of label-free quantitative strategies. But no matter the direction, quantitative glycomics and glycoproteomics will remain an area of significant active focus for years to come, with numerous challenges still to be presented and overcome.

Acknowledgements

Support for this research was provided in part by the NIH grants U01CA231081, R01 DK071801, RF1 AG052324, and w. LL acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Tables

Table 4.1 Comparisons of Labeling Strategies for Glycan Quantitation

Type	Method Name	Pros	Cons
Metabolic Incorporation/Isotopic Labeling	ICORA[58]	Improved reporting signal through increased O-glycan abundance, increased enrichment efficiency, optimal labeling efficiency	Only validated for O-glycans, time-restrictive, growth conditions must be carefully monitored
Isotopic Labeling	Dimethyl Labeling	Low cost reagents, facile labeling, slight increase in glycan hydrophobicity	Limited throughput (low multiplexing capacity)
Isotopic Labeling	Isotopic Permethylation[42]	Significant improvements in glycan hydrophobicity and ionization efficiency, 8-channel multiplexing	Toxicity of iodomethane reagents
Isotopic Labeling	Custom Tags (e.g. PMP-, Gerard's reagent P-, aniline-based etc.)	Highly customizable, effective in bespoke tagging workflows, stabilization of sialic acid residues, fixing of permanent positive charges	Concerns over labeling efficiency, need for optimization and method design
Mass Defect	DiPyrO[66]	Greatly reduced spectral complexity, elimination of redundant sampling, precursor co-isolation does not affect quantification, amine reactive tag (may be applied to glycans, peptides, and proteins)	Low multiplexing capacity (3-channels), requires higher-resolution MS^1 scans, current instrumentation outperforms multiplexing capacity
Mass Defect	mdSUGAR[67]	Labeling at glycan reducing end and on sialic acids, improved glycan fragmentation compared to commercial tags	Carbonyl-reactive tags are not as flexible in peptide and protein quantification, offers three-channel multiplexing
Isobaric Labeling	QUANTITY[70]	Improved fragmentation and reporter ion signal, high labeling efficiency. Quaternary amin imparts permanent positive charge	Requires multi-step synthesis, offers 4-channel multiplexing
Isobaric Labeling	TMT	Commercial quality control, well characterized protocols, 8-channel multiplexing, fits within Thermo "ecosystem"	Cost-preventative
Isobaric Labeling	SUGAR[68]	Improved b/y glycan fragment series for identification, synthesized in three high-yield steps, near 100% labeling efficiency, higher reporter ion signal for quantitation	Offers 4-plex multiplexing

Figures

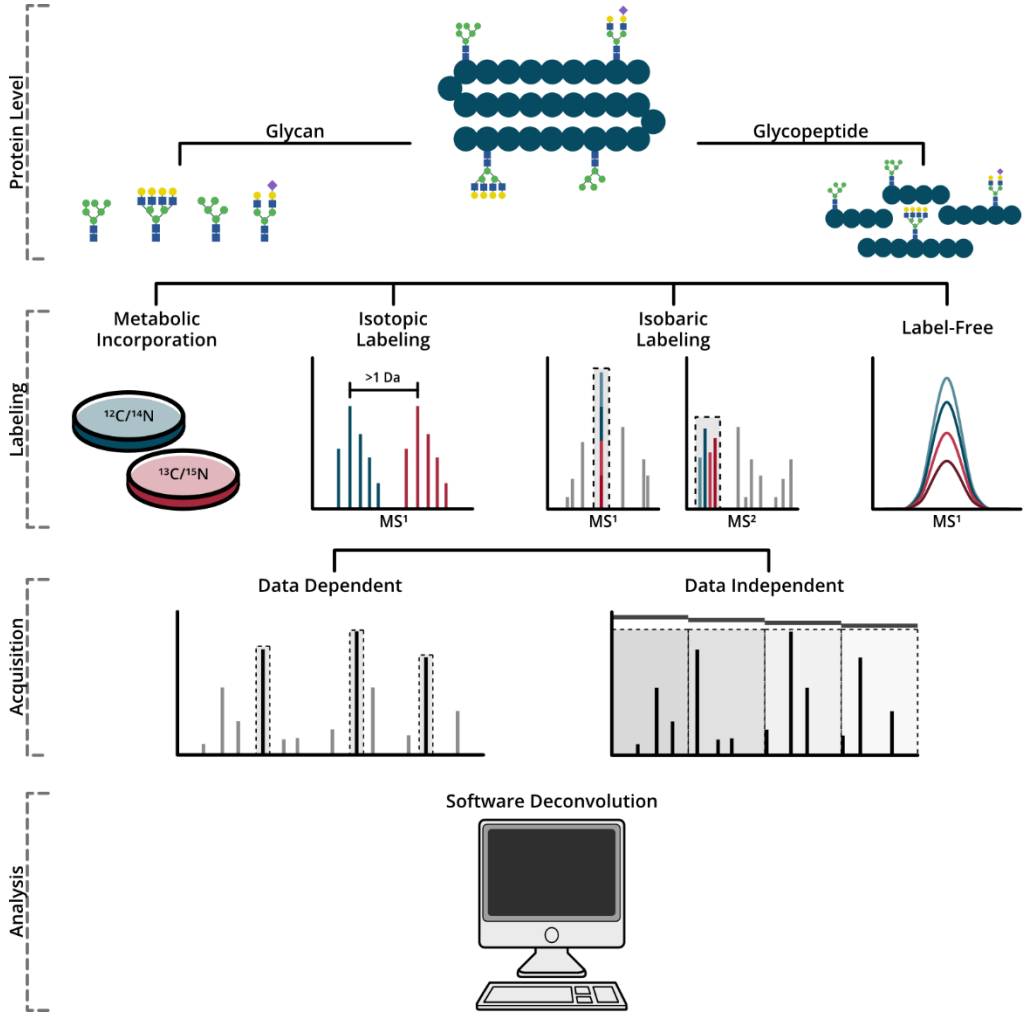


Figure 4.1 Graphical representation of quantitative glycomics and glycoproteomic analyses.

Glycomic evaluations, as discussed here, may take place at either the glycan or glycopeptide level and pursued through incorporation of stable isotopes, deposition of isotopic labels for MS^1 level quantification, isobaric labeling for MS^2 level quantitation, or label-free comparison. Both data dependent and data independent acquisition are effectively employed for glycome or glycopeptide detection with numerous software tools available to perform identification and quantitative analysis.

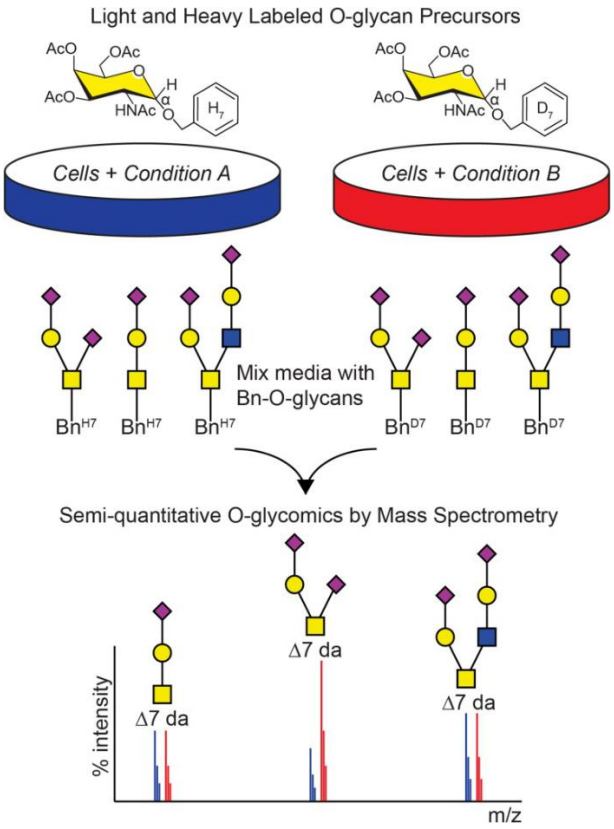


Figure 4.2 Overview of isotopic labeling with cellular O-glycome reporter/amplification (ICORA).

Cells undergoing condition A are incubated with Ac3GalNAc-BnH7 while cells undergoing condition B are incubated with Ac3GalNAc-BnD7. Ac3GalNAc-Bn crosses the plasma membrane, is de-esterified in the cytosol, taken up into the Golgi apparatus, and modified by endogenous glycosyltransferases to produce light H7 or heavy D7 labeled Bn-O-glycans before being secreted into the media. Media from the two conditions is mixed together and heavy and light Bn-O-glycans are purified, permethylated, and analyzed by mass spectrometry. A 7 Da mass shift distinguishes the light and heavy O-glycans, enabling quantification of shifts in relative abundance and comparison of O-glycans in condition A versus condition B. Reprinted from Kudelka *et al*[58] with permission from the author.

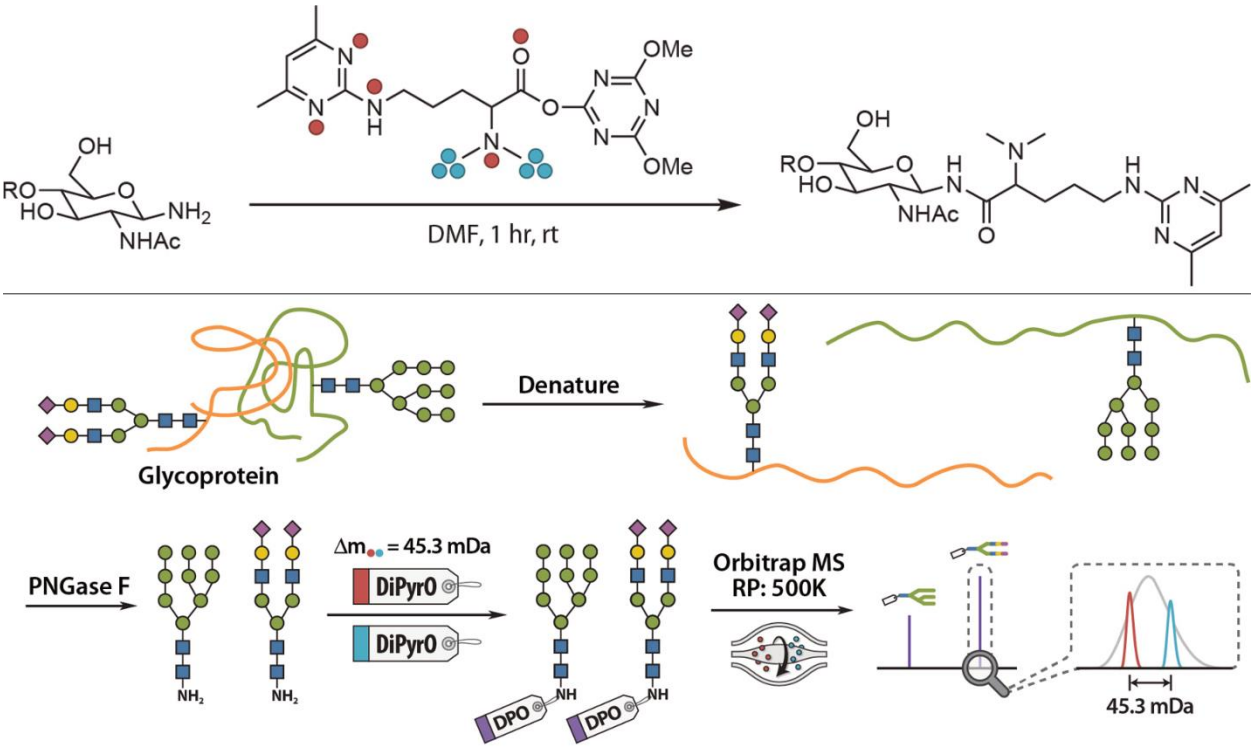


Figure 4.3 DiPyrO labeling

Top) DiPyrO Labeling of Glycosylamine; Red dots represent heavy isotopic atoms (15N418O) in the light DiPyrO tag; blue dots represent heavy isotopic atoms (2H6) in the heavy DiPyrO tag. Bottom) Workflow for the relative quantification of DiPyrO-labeled N-glycans illustrating the microenvironment. Adapted from Chen *et al.*[66] with permission.

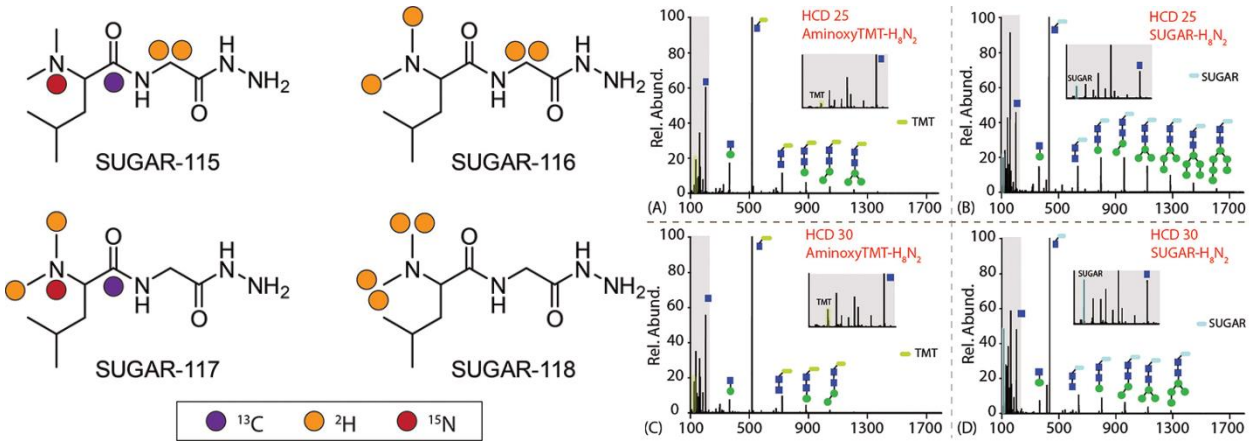


Figure 4.4 SUGAR Labeling

Left) Structure and isotope configurations of 4-plex SUGAR tags. Purple dot: ^{13}C, orange dot: ^{2}H, red dot: ^{15}N. Right) ESI-MS/MS fragmentation comparison of aminoxyTMT-labeled and SUGAR-labeled N-glycans. AminoxyTMT-labeled H_8N_2 ([aminoxyTMT – H_8N_2 + 2H]$^{2+}$) at NCE 25 (A) and 30 (C), SUGAR-labeled H_8N_2 ([SUGAR – H_8N_2 + 2H]$^{2+}$) at NCE 25 (B) and 30 (D). Adapted from Feng *et al.*[68] with permission.

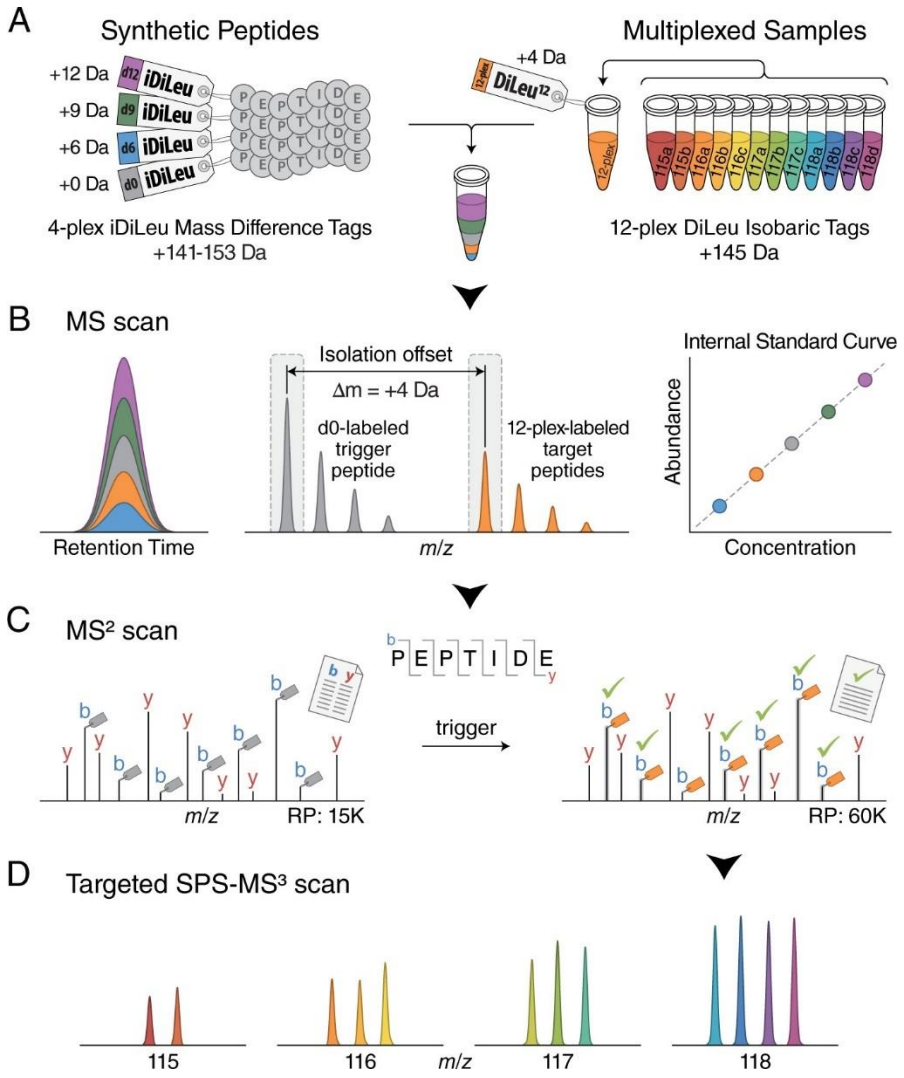


Figure 4.5 Schematic illustration for the HOTMAQ method.

A) Synthetic peptides are labeled with 4-plex iDiLeu at different concentrations and spiked into 12-plex DiLeu-labeled analytes. B) Labeled peptides are detected with identical chromatographic elution profiles as five precursor ion clusters. The iDiLeu labeled-synthetic peptides are used to generate internal calibration curves to quantify the total amount of multiplexed target peptides. iDiLeu d0-labeled synthetic trigger peptides and multiplexed DiLeu-labeled target peptides are separated in MS1 spectra by a mass offset of 4.01 Da, which enables synthetic trigger peptides to initiate quantitative analysis of target peptides via MS2 regardless of target peptide precursor abundances. C) Real-time MS2 analysis of d0-labeled synthetic peptides by matching MS2

spectrum to a product mass inclusion list unambiguously triggers fragmentation of 12-plex DiLeu-labeled target peptides in a predefined monitoring window. Acquisition parameters alternate between a low-resolution scan for monitoring d0-labeled trigger peptides and a high-resolution scan for quantifying 12-plex DiLeu-labeled target peptides. Fragment ions of 12-plex DiLeu-labeled target peptides are selected for synchronous precursor selection (SPS)-MS3 analysis. D) The relative abundance of each 12-plex DiLeu-labeled peptide is accurately determined by targeted SPS-MS3 acquisition at a resolving power of 60K (at m/z 200). The absolute amounts of target peptides are quantified by integrating the total amount obtained using the standard curve. Adapted from Zhang *et al.*[156] with permission.

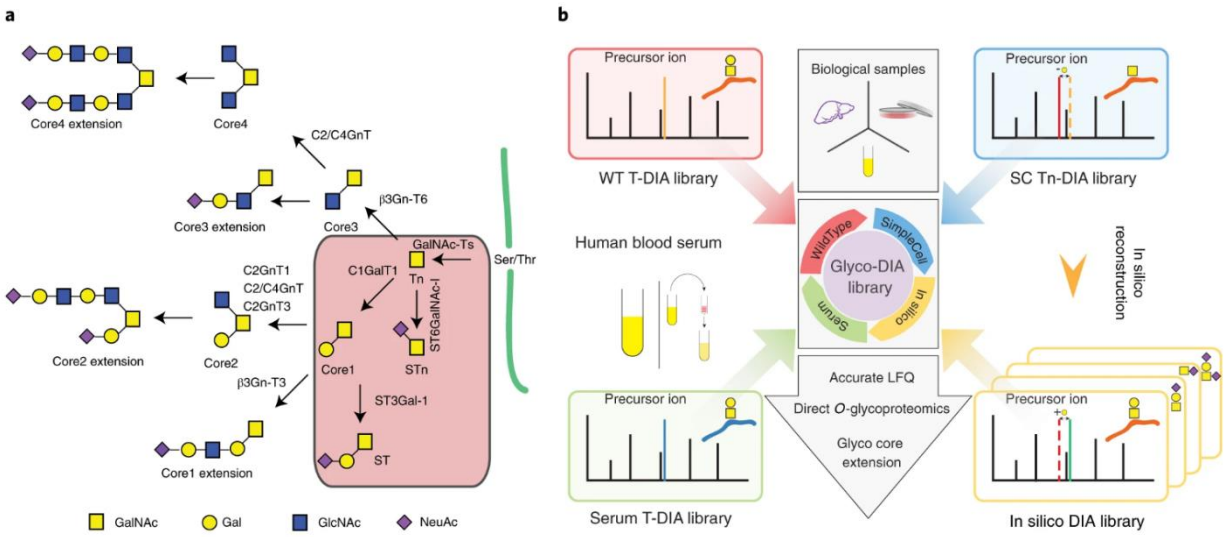


Figure 4.6 Outline of Glyco-DIA.

A) The major biosynthetic steps and enzymes involved in core1–4 *O*-glycan structures with extensions and capping by sialic acid are illustrated. B) Overview of Glyco-DIA libraries. The Glyco-DIA library consists of several sublibraries, including Tn-DIA libraries from SC cell lines, T-DIA libraries from WT cell lines, T-DIA library from human blood serum and in silico-expanded libraries. LFQ, label-free quantification. Reprinted from Ye *et al*[183] with permission from the author.

References

1. Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; MacCoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schlüter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlén, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschlager, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B., How many human proteoforms are there? *Nature Chemical Biology* **2018,** *14* (3), 206-214.

2. Varki, A., *Essentials of glycobiology*. Third edition. ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 2017; p xxix, 823 pages.

3. Nwosu, C. C.; Strum, J. S.; An, H. J.; Lebrilla, C. B., Enhanced detection and identification of glycopeptides in negative ion mode mass spectrometry. *Analytical chemistry* **2010,** *82* (23), 9654-9662.

4. Hart-Smith, G.; Raftery, M. J., Detection and Characterization of Low Abundance Glycopeptides Via Higher-Energy C-Trap Dissociation and Orbitrap Mass Analysis. *Journal of The American Society for Mass Spectrometry* **2012,** *23* (1), 124-140.

5. Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nature Methods* **2009,** *6* (5), 359-362.

6. Totten, S. M.; Feasley, C. L.; Bermudez, A.; Pitteri, S. J., Parallel Comparison of N-Linked Glycopeptide Enrichment Techniques Reveals Extensive Glycoproteomic Analysis of Plasma Enabled by SAX-ERLIC. *Journal of Proteome Research* **2017,** *16* (3), 1249-1260.

7. Selman, M. H. J.; Hemayatkar, M.; Deelder, A. M.; Wuhrer, M., Cotton HILIC SPE Microtips for Microscale Purification and Enrichment of Glycans and Glycopeptides. *Analytical Chemistry* **2011,** *83* (7), 2492-2499.

8. Cui, Y.; Yang, K.; Tabang, D. N.; Huang, J.; Tang, W.; Li, L., Finding the Sweet Spot in ERLIC Mobile Phase for Simultaneous Enrichment of N-Glyco and Phosphopeptides. *Journal of the American Society for Mass Spectrometry* **2019,** *30* (12), 2491-2501.

9. Ko, B. J.; Brodbelt, J. S., Comparison of glycopeptide fragmentation by collision induced dissociation and ultraviolet photodissociation. *International Journal of Mass Spectrometry* **2015,** *377*, 385-392.

10. Riley, N. M.; Hebert, A. S.; Westphall, M. S.; Coon, J. J., Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nature Communications* **2019,** *10* (1), 1311.

11. Chandler, K. B.; Leon, D. R.; Kuang, J.; Meyer, R. D.; Rahimi, N.; Costello, C. E., N-Glycosylation regulates ligand-dependent activation and signaling of vascular endothelial growth factor receptor 2 (VEGFR2). *Journal of Biological Chemistry* **2019,** *294* (35), 13117-13130.

12. Seyrek, K.; Richter, M.; Lavrik, I. N., Decoding the sweet regulation of apoptosis: the role of glycosylation and galectins in apoptotic signaling pathways. *Cell Death & Differentiation* **2019,** *26* (6), 981-993.

13. Perkey, E.; Maurice De Sousa, D.; Carrington, L.; Chung, J.; Dils, A.; Granadier, D.; Koch, U.; Radtke, F.; Ludewig, B.; Blazar, B. R.; Siebel, C. W.; Brennan, T. V.; Nolz, J.; Labrecque, N.; Maillard, I., GCNT1-Mediated O-Glycosylation of the Sialomucin CD43 Is a Sensitive Indicator of Notch Signaling in Activated T Cells. *The Journal of Immunology* **2020,** *204* (6), 1674-1688.

14. Hill, H. H.; Hill, C. H.; Asbury, G. R.; Wu, C.; Matz, L. M.; Ichiye, T., Charge location on gas phase peptides. *International Journal of Mass Spectrometry* **2002,** *219* (1), 23-37.

15. Leymarie, N.; Zaia, J., Effective use of mass spectrometry for glycan and glycopeptide structural analysis. *Anal Chem* **2012,** *84* (7), 3040-8.

16. Abouelhadid, S.; North, S. J.; Hitchen, P.; Vohra, P.; Chintoan-Uta, C.; Stevens, M.; Dell, A.; Cuccui, J.; Wren, B. W., Quantitative Analyses Reveal Novel Roles for N-Glycosylation in a Major Enteric Bacterial Pathogen. *mBio* **2019,** *10* (2), e00297-19.

17. Ozdilek, A.; Paschall, A. V.; Dookwah, M.; Tiemeyer, M.; Avci, F. Y., Host protein glycosylation in nucleic acid vaccines as a potential hurdle in vaccine design for nonviral pathogens. *Proceedings of the National Academy of Sciences* **2020,** *117* (3), 1280-1282.

18. Marín-Menguiano, M.; Moreno-Sánchez, I.; Barrales, R. R.; Fernández-Álvarez, A.; Ibeas, J. I., N-glycosylation of the protein disulfide isomerase Pdi1 ensures full Ustilago maydis virulence. *PLOS Pathogens* **2019,** *15* (11), e1007687.

19. Lucchetta, M.; da Piedade, I.; Mounir, M.; Vabistsevits, M.; Terkelsen, T.; Papaleo, E., Distinct signatures of lung cancer types: aberrant mucin O-glycosylation and compromised immune response. *BMC Cancer* **2019,** *19* (1), 824.

20. Läubli, H.; Borsig, L., Altered Cell Adhesion and Glycosylation Promote Cancer Immune Suppression and Metastasis. *Frontiers in Immunology* **2019,** *10* (2120).

21. Zhang, W.; Zhou, X.; Muchena, J.; Shuai, Z.; Sun, Y.; Yang, G.-X.; Liu, C. H.; Dhaliwal, S.; Invernizzi, P.; Bowlus, C.; Leung, P. S. C.; Lebrilla, C.; Gershwin, M. E., The Glycosylation and Glycoprotein Signature of Immunoglobulins in Patients with Autoimmune Liver Diseases. *The Journal of Immunology* **2019,** *202* (1 Supplement), 179.3-179.3.

22. Regan, P.; McClean, P. L.; Smyth, T.; Doherty, M., Early Stage Glycosylation Biomarkers in Alzheimer's Disease. *Medicines* **2019,** *6* (3), 92.

23. Fang, P.; Xie, J.-J.; Sang, S.-M.; Zhang, L.; Liu, M.-Q.; Yang, L.-J.; Xu, Y.-T.; Yan, G.-Q.; Yao, J.; Gao, X.; Qian, W.-J.; Wang, Z.-F.; Zhang, Y.; Yang, P.-Y.; Shen, H.-L., Multilayered N-glycoproteomics reveals impaired N-glycosylation promoting Alzheimer's disease. *bioRxiv* **2019**, 615989.

24. Rudman, N.; Gornik, O.; Lauc, G., Altered N-glycosylation profiles as potential biomarkers and drug targets in diabetes. *FEBS Letters* **2019,** *593* (13), 1598-1615.

25. HIGASHIOKA, M.; HIRAKAWA, Y.; YOSHINARI, M.; HONDA, T.; SAKATA, S.; SHIBATA, M.; YOSHIDA, D.; HATA, J.; KITAZONO, T.; OSAWA, H.; NINOMIYA, T., 1522-P: Serum Mac-2 Binding Protein Glycosylation Isomer Level and the Risk of Type 2 Diabetes in a Japanese Community: The Hisayama Study. *Diabetes* **2019,** *68* (Supplement 1), 1522-P.

26. Legler, K.; Rosprim, R.; Karius, T.; Eylmann, K.; Rossberg, M.; Wirtz, R. M.; Müller, V.; Witzel, I.; Schmalfeldt, B.; Milde-Langosch, K.; Oliveira-Ferrer, L., Reduced mannosidase MAN1A1 expression leads to aberrant N-glycosylation and impaired survival in breast cancer. *British Journal of Cancer* **2018,** *118* (6), 847-856.

27. Burch, T. C.; Mackay, S. S.; Oduor, I. O.; Otto, J. J.; Lance, R. S.; Troyer, D. A.; Semmes, O. J.; Nyalwidhe, J. O., Abstract 3008: PSMA glycosylation and aggressive prostate cancer progression. *Cancer Research* **2019,** *79* (13 Supplement), 3008-3008.

28. Tang, L.; Chen, X.; Zhang, X.; Guo, Y.; Su, J.; Zhang, J.; Peng, C.; Chen, X., N-Glycosylation in progression of skin cancer. *Medical Oncology* **2019,** *36* (6), 50.

29. Mechref, Y.; Hu, Y.; Desantos-Garcia, J. L.; Hussein, A.; Tang, H., Quantitative Glycomics Strategies. *Molecular & Cellular Proteomics : MCP* **2013,** *12* (4), 874-884.

30. Ruhaak, L. R.; Zauner, G.; Huhn, C.; Bruggink, C.; Deelder, A. M.; Wuhrer, M., Glycan labeling strategies and their use in identification and quantification. *Analytical and Bioanalytical Chemistry* **2010,** *397* (8), 3457-3481.

31. Váradi, C.; Mittermayr, S.; Millán-Martín, S.; Bones, J., Quantitative twoplex glycan analysis using 12C6 and 13C6 stable isotope 2-aminobenzoic acid labelling and capillary electrophoresis mass spectrometry. *Analytical and Bioanalytical Chemistry* **2016,** *408* (30), 8691-8700.

32. Giménez, E.; Sanz-Nebot, V.; Rizzi, A., Relative quantitation of glycosylation variants by stable isotope labeling of enzymatically released N-glycans using [12C]/[13C] aniline and ZIC-HILIC-ESI-TOF-MS. *Analytical and Bioanalytical Chemistry* **2013,** *405* (23), 7307-7319.

33. Michael, C.; Rizzi, A. M., Quantitative isomer-specific N-glycan fingerprinting using isotope coded labeling and high performance liquid chromatography–electrospray ionization-mass spectrometry with graphitic carbon stationary phase. *Journal of Chromatography A* **2015,** *1383*, 88-95.

34. Michael, C.; Rizzi, A. M., Tandem mass spectrometry of isomeric aniline-labeled N-glycans separated on porous graphitic carbon: Revealing the attachment position of terminal sialic acids and structures of neutral glycans. *Rapid Communications in Mass Spectrometry* **2015,** *29* (13), 1268-1278.

35. Walker, S. H.; Budhathoki-Uprety, J.; Novak, B. M.; Muddiman, D. C., Stable-isotope labeled hydrophobic hydrazide reagents for the relative quantification of N-linked glycans by electrospray ionization mass spectrometry. *Analytical chemistry* **2011,** *83* (17), 6738-6745.

36. Walker, S. H.; Taylor, A. D.; Muddiman, D. C., Individuality Normalization when Labeling with Isotopic Glycan Hydrazide Tags (INLIGHT): a novel glycan-relative quantification strategy. *Journal of the American Society for Mass Spectrometry* **2013,** *24* (9), 1376-1384.

37. Tao, S.; Orlando, R., A novel method for relative quantitation of N-glycans by isotopic labeling using ^{18}O-water. *J Biomol Tech* **2014,** *25* (4), 111-117.

38. Zhang, W.; Cao, W.; Huang, J.; Wang, H.; Wang, J.; Xie, C.; Yang, P., PNGase F-mediated incorporation of 18O into glycans for relative glycan quantitation. *Analyst* **2015,** *140* (4), 1082-1089.

39. Cao, W.; Zhang, W.; Huang, J.; Jiang, B.; Zhang, L.; Yang, P., Glycan reducing end dual isotopic labeling (GREDIL) for mass spectrometry-based quantitative N-glycomics. *Chemical Communications* **2015,** *51* (71), 13603-13606.

40. Ciucanu, I.; Kerek, F., A simple and rapid method for the permethylation of carbohydrates. *Carbohydrate Research* **1984,** *131* (2), 209-217.

41. Hu, Y.; Desantos-Garcia, J. L.; Mechref, Y., Comparative glycomic profiling of isotopically permethylated N-glycans by liquid chromatography/electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **2013,** *27* (8), 865-877.

42. Dong, X.; Peng, W.; Yu, C.-Y.; Zhou, S.; Donohoo, K. B.; Tang, H.; Mechref, Y., 8-plex LC–MS/MS Analysis of Permethylated N-Glycans Achieved by Using Stable Isotopic Iodomethane. *Analytical Chemistry* **2019,** *91* (18), 11794-11802.

43. Bowman, M. J.; Zaia, J., Tags for the Stable Isotopic Labeling of Carbohydrates and Quantitative Analysis by Mass Spectrometry. *Analytical Chemistry* **2007,** *79* (15), 5777-5784.

44. Bowman, M. J.; Zaia, J., Comparative Glycomics Using a Tetraplex Stable-Isotope Coded Tag. *Analytical Chemistry* **2010,** *82* (7), 3023-3031.

45. Sić, S.; Maier, N. M.; Rizzi, A. M., Quantitative fingerprinting of O-linked glycans released from proteins using isotopic coded labeling with deuterated 1-phenyl-3-methyl-5-pyrazolone. *Journal of Chromatography A* **2015,** *1408*, 93-100.

46. Wang, C.; Zhang, P.; Jin, W.; Li, L.; Qiang, S.; Zhang, Y.; Huang, L.; Wang, Z., Quantitative O-glycomics based on improvement of the one-pot method for nonreductive O-glycan release and simultaneous stable isotope labeling with 1-(d0/d5)phenyl-3-methyl-5-pyrazolone followed by mass spectrometric analysis. *Journal of Proteomics* **2017,** *150*, 18-30.

47. Sić, S.; Maier, N. M.; Rizzi, A. M., Quantitative profiling of O-glycans by electrospray ionization- and matrix-assisted laser desorption ionization-time-of-flight-mass spectrometry after in-gel derivatization with isotope-coded 1-phenyl-3-methyl-5-pyrazolone. *Analytica Chimica Acta* **2016,** *935*, 187-196.

48. Shah, P.; Yang, S.; Sun, S.; Aiyetan, P.; Yarema, K. J.; Zhang, H., Mass Spectrometric Analysis of Sialylated Glycans with Use of Solid-Phase Labeling of Sialic Acids. *Analytical Chemistry* **2013,** *85* (7), 3606-3613.

49. Wei, L.; Cai, Y.; Yang, L.; Zhang, Y.; Lu, H., Duplex Stable Isotope Labeling (DuSIL) for Simultaneous Quantitation and Distinction of Sialylated and Neutral N-Glycans by MALDI-MS. *Analytical Chemistry* **2018,** *90* (17), 10442-10449.

50. Yang, L.; Du, X.; Peng, Y.; Cai, Y.; Wei, L.; Zhang, Y.; Lu, H., Integrated Pipeline of Isotopic Labeling and Selective Enriching for Quantitative Analysis of N-Glycome by Mass Spectrometry. *Analytical Chemistry* **2019,** *91* (2), 1486-1493.

51. Wang, L.; Yang, L.; Zhang, Y.; Lu, H., Dual isotopic labeling combined with fluorous solid-phase extraction for simultaneous discovery of neutral/sialylated N-glycans as biomarkers for gastric cancer. *Analytica Chimica Acta* **2020**.

52. Jin, W.; Wang, C.; Yang, M.; Wei, M.; Huang, L.; Wang, Z., Glycoqueuing: Isomer-Specific Quantification for Sialylation-Focused Glycomics. *Analytical Chemistry* **2019,** *91* (16), 10492-10500.

53. Shi, Q.; Hashimoto, R.; Otsubo, T.; Ikeda, K.; Todoroki, K.; Mizuno, H.; Jin, D.; Toyo'oka, T.; Jiang, Z.; Min, J. Z., A novel, simplified strategy of relative quantification N-glycan: Quantitative glycomics using electrospray ionization mass spectrometry through the stable isotopic labeling by transglycosylation reaction of mutant enzyme Endo-M-N175Q. *Journal of Pharmaceutical and Biomedical Analysis* **2018,** *149*, 365-373.

54. Yang, L.; Peng, Y.; Jiao, J.; Tao, T.; Yao, J.; Zhang, Y.; Lu, H., Metallic Element Chelated Tag Labeling (MeCTL) for Quantitation of N-Glycans in MALDI-MS. *Analytical Chemistry* **2017,** *89* (14), 7470-7476.

55. Li, H.; Li, L.; Cheng, K.; Ning, Z.; Mayne, J.; Zhang, X.; Walker, K.; Chen, R.; Twine, S.; Li, J.; Figeys, D., Chemoenzymatic Method for Glycoproteomic N-Glycan Type Quantitation. *Analytical Chemistry* **2020,** *92* (1), 1618-1627.

56. Chen, Z.; Zhong, X.; Tie, C.; Chen, B.; Zhang, X.; Li, L., Development of a hydrophilic interaction liquid chromatography coupled with matrix-assisted laser desorption/ionization-mass spectrometric imaging platform for N-glycan relative quantitation using stable-isotope labeled hydrazide reagents. *Analytical and Bioanalytical Chemistry* **2017,** *409* (18), 4437-4447.

57. Kudelka, M. R.; Antonopoulos, A.; Wang, Y.; Duong, D. M.; Song, X.; Seyfried, N. T.; Dell, A.; Haslam, S. M.; Cummings, R. D.; Ju, T., Cellular O-Glycome Reporter/Amplification to explore O-glycans of living cells. *Nature Methods* **2016,** *13* (1), 81-86.

58. Kudelka, M. R.; Nairn, A. V.; Sardar, M. Y.; Sun, X.; Chaikof, E. L.; Ju, T.; Moremen, K. W.; Cummings, R. D., Isotopic labeling with cellular O-glycome reporter/amplification (ICORA) for comparative O-glycomics of cultured cells. *Glycobiology* **2018,** *28* (4), 214-222.

59. Orlando, R.; Lim, J.-M.; Atwood, J. A.; Angel, P. M.; Fang, M.; Aoki, K.; Alvarez-Manilla, G.; Moremen, K. W.; York, W. S.; Tiemeyer, M.; Pierce, M.; Dalton, S.; Wells, L., IDAWG: Metabolic Incorporation of Stable Isotope Labels for Quantitative Glycomics of Cultured Cells. *Journal of Proteome Research* **2009,** *8* (8), 3816-3823.

60. Yang, G.; Tan, Z.; Lu, W.; Guo, J.; Yu, H.; Yu, J.; Sun, C.; Qi, X.; Li, Z.; Guan, F., Quantitative Glycome Analysis of N-Glycan Patterns in Bladder Cancer vs Normal Bladder Cells Using an Integrated Strategy. *Journal of Proteome Research* **2015,** *14* (2), 639-653.

61. Sun, S.; Shah, P.; Eshghi, S. T.; Yang, W.; Trikannad, N.; Yang, S.; Chen, L.; Aiyetan, P.; Höti, N.; Zhang, Z.; Chan, D. W.; Zhang, H., Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nature Biotechnology* **2016,** *34* (1), 84-88.

62. Zhang, H.; Li, X.-j.; Martin, D. B.; Aebersold, R., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nature Biotechnology* **2003,** *21* (6), 660-666.

63. Atwood, J. A.; Cheng, L.; Alvarez-Manilla, G.; Warren, N. L.; York, W. S.; Orlando, R., Quantitation by Isobaric Labeling: Applications to Glycomics. *Journal of Proteome Research* **2008,** *7* (1), 367-374.

64. Botelho, J. C.; Atwood, J. A.; Cheng, L.; Alvarez-Manilla, G.; York, W. S.; Orlando, R., Quantification by isobaric labeling (QUIBL) for the comparative glycomic study of O-linked glycans. *International Journal of Mass Spectrometry* **2008,** *278* (2), 137-142.

65. Hao, L.; Johnson, J.; Lietz, C. B.; Buchberger, A.; Frost, D.; Kao, W. J.; Li, L., Mass Defect-Based N,N-Dimethyl Leucine Labels for Quantitative Proteomics and Amine Metabolomics of Pancreatic Cancer Cells. *Analytical Chemistry* **2017,** *89* (2), 1138-1146.

66. Chen, B.; Feng, Y.; Frost, D. C.; Zhong, X.; Buchberger, A. R.; Johnson, J.; Xu, M.; Kim, M.; Puccetti, D.; Diamond, C.; Ikonomidou, C.; Li, L., Quantitative Glycomic Analysis by Mass-Defect-Based Dimethyl Pyrimidinyl Ornithine (DiPyrO) Tags and High-Resolution Mass Spectrometry. *Analytical Chemistry* **2018,** *90* (13), 7817-7823.

67. Feng, Y.; Li, M.; Lin, Y.; Chen, B.; Li, L., Multiplex Quantitative Glycomics Enabled by Periodate Oxidation and Triplex Mass Defect Isobaric Multiplex Reagents for Carbonyl-Containing Compound Tags. *Analytical Chemistry* **2019,** *91* (18), 11932-11937.

68. Feng, Y.; Chen, B.; Yu, Q.; Zhong, X.; Frost, D. C.; Ikonomidou, C.; Li, L., Isobaric Multiplex Labeling Reagents for Carbonyl-Containing Compound (SUGAR) Tags: A Probe for Quantitative Glycomic Analysis. *Analytical Chemistry* **2019,** *91* (4), 3141-3146.

69. Yang, S.; Yuan, W.; Yang, W.; Zhou, J.; Harlan, R.; Edwards, J.; Li, S.; Zhang, H., Glycan analysis by isobaric aldehyde reactive tags and mass spectrometry. *Analytical chemistry* **2013,** *85* (17), 8188-8195.

70. Yang, S.; Wang, M.; Chen, L.; Yin, B.; Song, G.; Turko, I. V.; Phinney, K. W.; Betenbaugh, M. J.; Zhang, H.; Li, S., QUANTITY: An Isobaric Tag for Quantitative Glycomics. *Scientific Reports* **2015,** *5* (1), 17585.

71. Yang, S.; Hu, Y.; Sokoll, L.; Zhang, H., Simultaneous quantification of N- and O-glycans using a solid-phase method. *Nature Protocols* **2017,** *12* (6), 1229-1244.

72. Hahne, H.; Neubert, P.; Kuhn, K.; Etienne, C.; Bomgarden, R.; Rogers, J. C.; Kuster, B., Carbonyl-Reactive Tandem Mass Tags for the Proteome-Wide Quantification of N-Linked Glycans. *Analytical Chemistry* **2012,** *84* (8), 3716-3724.

73. Zhou, S.; Hu, Y.; Veillon, L.; Snovida, S. I.; Rogers, J. C.; Saba, J.; Mechref, Y., Quantitative LC–MS/MS Glycomic Analysis of Biological Samples Using AminoxyTMT. *Analytical Chemistry* **2016,** *88* (15), 7515-7522.

74. Zhong, X.; Chen, Z.; Snovida, S.; Liu, Y.; Rogers, J. C.; Li, L., Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry for Quantitative Analysis of Glycans Labeled with Multiplex Carbonyl-Reactive Tandem Mass Tags. *Analytical Chemistry* **2015,** *87* (13), 6527-6534.

75. Khatri, K.; Klein, J. A.; Haserick, J. R.; Leon, D. R.; Costello, C. E.; McComb, M. E.; Zaia, J., Microfluidic Capillary Electrophoresis–Mass Spectrometry for Analysis of Monosaccharides, Oligosaccharides, and Glycopeptides. *Analytical Chemistry* **2017,** *89* (12), 6645-6655.

76. Chen, B.; Zhong, X.; Feng, Y.; Snovida, S.; Xu, M.; Rogers, J.; Li, L., Targeted MultiNotch MS3 Approach for Relative Quantification of N-Glycans Using Multiplexed Carbonyl-Reactive Isobaric Tags. *Analytical Chemistry* **2018,** *90* (2), 1129-1135.

77. Barrientos, R. C.; Zhang, Q., Isobaric Labeling of Intact Gangliosides toward Multiplexed LC–MS/MS-Based Quantitative Analysis. *Analytical Chemistry* **2018,** *90* (4), 2578-2586.

78. Lauber, M. A.; Yu, Y.-Q.; Brousmiche, D. W.; Hua, Z.; Koza, S. M.; Magnelli, P.; Guthrie, E.; Taron, C. H.; Fountain, K. J., Rapid Preparation of Released N-Glycans for HILIC Analysis Using a Labeling Reagent that Facilitates Sensitive Fluorescence and ESI-MS Detection. *Analytical Chemistry* **2015,** *87* (10), 5401-5409.

79. Rana, S.; Le, N. D. B.; Mout, R.; Duncan, B.; Elci, S. G.; Saha, K.; Rotello, V. M., A Multichannel Biosensor for Rapid Determination of Cell Surface Glycomic Signatures. *ACS Central Science* **2015,** *1* (4), 191-197.

80. Zhou, S.; Tello, N.; Harvey, A.; Boyes, B.; Orlando, R.; Mechref, Y., Reliable LC-MS quantitative glycomics using iGlycoMab stable isotope labeled glycans as internal standards. *ELECTROPHORESIS* **2016,** *37* (11), 1489-1497.

81. Mehta, N.; Porterfield, M.; Struwe, W. B.; Heiss, C.; Azadi, P.; Rudd, P. M.; Tiemeyer, M.; Aoki, K., Mass Spectrometric Quantification of N-Linked Glycans by Reference to Exogenous Standards. *Journal of Proteome Research* **2016,** *15* (9), 2969-2980.

82. Kailemia, M. J.; Park, D.; Lebrilla, C. B., Glycans and glycoproteins as specific biomarkers for cancer. *Analytical and Bioanalytical Chemistry* **2017,** *409* (2), 395-410.

83. Veillon, L.; Huang, Y.; Peng, W.; Dong, X.; Cho Byeong, G.; Mechref, Y., Characterization of isomeric glycan structures by LC-MS/MS. *ELECTROPHORESIS* **2017,** *38* (17), 2100-2114.

84. James, A.; Jorgensen, C., Basic Design of MRM Assays for Peptide Quantification. In *LC-MS/MS in Proteomics: Methods and Applications*, Cutillas, P. R.; Timms, J. F., Eds. Humana Press: Totowa, NJ, 2010; pp 167-185.

85. Shubhakar, A.; Reiding, K. R.; Gardner, R. A.; Spencer, D. I. R.; Fernandes, D. L.; Wuhrer, M., High-Throughput Analysis and Automation for Glycomics Studies. *Chromatographia* **2015,** *78* (5), 321-333.

86. Ruhaak, L. R.; Lebrilla, C. B., Applications of Multiple Reaction Monitoring to Clinical Glycomics. *Chromatographia* **2015,** *78* (5), 335-342.

87. Hong, Q.; Ruhaak, L. R.; Totten, S. M.; Smilowitz, J. T.; German, J. B.; Lebrilla, C. B., Label-Free Absolute Quantitation of Oligosaccharides Using Multiple Reaction Monitoring. *Analytical Chemistry* **2014,** *86* (5), 2640-2647.

88. Xu, G.; Amicucci, M. J.; Cheng, Z.; Galermo, A. G.; Lebrilla, C. B., Revisiting monosaccharide analysis – quantitation of a comprehensive set of monosaccharides using dynamic multiple reaction monitoring. *Analyst* **2018,** *143* (1), 200-207.

89. Xia, B.; Zhang, W.; Li, X.; Jiang, R.; Harper, T.; Liu, R.; Cummings, R. D.; He, M., Serum N-glycan and O-glycan analysis by mass spectrometry for diagnosis of congenital disorders of glycosylation. *Analytical Biochemistry* **2013,** *442* (2), 178-185.

90. Tao, S.; Huang, Y.; Boyes, B. E.; Orlando, R., Liquid Chromatography-Selected Reaction Monitoring (LC-SRM) Approach for the Separation and Quantitation of Sialylated N-Glycans Linkage Isomers. *Analytical Chemistry* **2014,** *86* (21), 10584-10590.

91. Tsai, T.-H.; Wang, M.; Di Poto, C.; Hu, Y.; Zhou, S.; Zhao, Y.; Varghese, R. S.; Luo, Y.; Tadesse, M. G.; Ziada, D. H.; Desai, C. S.; Shetty, K.; Mechref, Y.; Ressom, H. W., LC–MS Profiling of N-Glycans Derived from Human Serum Samples for Biomarker Discovery in Hepatocellular Carcinoma. *Journal of Proteome Research* **2014,** *13* (11), 4859-4868.

92. Zhou, S.; Hu, Y.; DeSantos-Garcia, J. L.; Mechref, Y., Quantitation of Permethylated N-Glycans through Multiple-Reaction Monitoring (MRM) LC-MS/MS. *Journal of the American Society for Mass Spectrometry* **2015,** *26* (4), 596-603.

93. Orlando, R.; Popov, M.; Libert, B.; Boyes, B., Absolute Quantitation of the N-Linked Glycans from Biotheraputic IgGs. *J Biomol Tech* **2019,** *30* (Suppl), S3-S3.

94. Mank, M.; Welsch, P.; Heck, A. J. R.; Stahl, B., Label-free targeted LC-ESI-MS2 analysis of human milk oligosaccharides (HMOS) and related human milk groups with enhanced structural selectivity. *Analytical and Bioanalytical Chemistry* **2019,** *411* (1), 231-250.

95. Bollineni, R. C.; Koehler, C. J.; Gislefoss, R. E.; Anonsen, J. H.; Thiede, B., Large-scale intact glycopeptide identification by Mascot database search. *Scientific Reports* **2018,** *8* (1), 2117.

96. Malaker, S. A.; Pedram, K.; Ferracane, M. J.; Bensing, B. A.; Krishnan, V.; Pett, C.; Yu, J.; Woods, E. C.; Kramer, J. R.; Westerlind, U.; Dorigo, O.; Bertozzi, C. R., The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. *Proceedings of the National Academy of Sciences* **2019,** *116* (15), 7278-7287.

97. Zhang, T.; Madunić, K.; Holst, S.; Zhang, J.; Jin, C.; ten Dijke, P.; Karlsson, N. G.; Stavenhagen, K.; Wuhrer, M., Development of a 96-well plate sample preparation method for integrated N- and O-glycomics using porous graphitized carbon liquid chromatography-mass spectrometry. *Molecular Omics* **2020**.

98. Yang, G.; Höti, N.; Chen, S.-Y.; Zhou, Y.; Wang, Q.; Betenbaugh, M.; Zhang, H., One-Step Enrichment of Intact Glycopeptides From Glycoengineered Chinese Hamster Ovary Cells. *Front Chem* **2020,** *8*, 240-240.

99. Qing, G.; Yan, J.; He, X.; Li, X.; Liang, X., Recent advances in hydrophilic interaction liquid interaction chromatography materials for glycopeptide enrichment and glycan separation. *TrAC Trends in Analytical Chemistry* **2020,** *124*, 115570.

100. Zhang, C.; Ye, Z.; Xue, P.; Shu, Q.; Zhou, Y.; Ji, Y.; Fu, Y.; Wang, J.; Yang, F., Evaluation of Different N-Glycopeptide Enrichment Methods for N-Glycosylation Sites Mapping in Mouse Brain. *Journal of Proteome Research* **2016,** *15* (9), 2960-2968.

101. Vreeker, G. C. M.; Nicolardi, S.; Madunic, K.; Kotsias, M.; van der Burgt, Y. E. M.; Wuhrer, M., O- and N-glycosylation analysis of cell lines by ultrahigh resolution MALDI-FTICR-MS. *International Journal of Mass Spectrometry* **2020,** *448*, 116267.

102. Dang, L.; Jia, L.; Zhi, Y.; Li, P.; Zhao, T.; Zhu, B.; Lan, R.; Hu, Y.; Zhang, H.; Sun, S., Mapping human N-linked glycoproteins and glycosylation sites using mass spectrometry. *TrAC Trends in Analytical Chemistry* **2019,** *114*, 143-150.

103. Reiding, K. R.; Bondt, A.; Franc, V.; Heck, A. J. R., The benefits of hybrid fragmentation methods for glycoproteomics. *TrAC Trends in Analytical Chemistry* **2018,** *108*, 260-268.

104. Liu, Y.; Hüttenhain, R.; Surinova, S.; Gillet, L. C.; Mouritsen, J.; Brunner, R.; Navarro, P.; Aebersold, R., Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. *PROTEOMICS* **2013,** *13* (8), 1247-1256.

105. Poljak, K.; Selevsek, N.; Ngwa, E.; Grossmann, J.; Losfeld, M. E.; Aebi, M., Quantitative Profiling of N-linked Glycosylation Machinery in Yeast Saccharomyces cerevisiae. *Molecular & Cellular Proteomics* **2018,** *17* (1), 18-30.

106. Tornøe, C. W.; Christensen, C.; Meldal, M., Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *The Journal of Organic Chemistry* **2002,** *67* (9), 3057-3064.

107. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angewandte Chemie International Edition* **2002,** *41* (14), 2596-2599.

108. Woo, C. M.; Iavarone, A. T.; Spiciarich, D. R.; Palaniappan, K. K.; Bertozzi, C. R., Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nature methods* **2015,** *12* (6), 561-567.

109. Woo, C. M.; Felix, A.; Byrd, W. E.; Zuegel, D. K.; Ishihara, M.; Azadi, P.; Iavarone, A. T.; Pitteri, S. J.; Bertozzi, C. R., Development of IsoTaG, a Chemical Glycoproteomics Technique for Profiling Intact N- and O-Glycopeptides from Whole Cell Proteomes. *Journal of Proteome Research* **2017,** *16* (4), 1706-1718.

110. Taga, Y.; Kusubata, M.; Ogawa-Goto, K.; Hattori, S., Site-specific Quantitative Analysis of Overglycosylation of Collagen in Osteogenesis Imperfecta Using Hydrazide Chemistry and SILAC. *Journal of Proteome Research* **2013,** *12* (5), 2225-2232.

111. Woo, C. M.; Felix, A.; Zhang, L.; Elias, J. E.; Bertozzi, C. R., Isotope-targeted glycoproteomics (IsoTaG) analysis of sialylated N- and O-glycopeptides on an Orbitrap Fusion Tribrid using azido and alkynyl sugars. *Analytical and Bioanalytical Chemistry* **2017,** *409* (2), 579-588.

112. Palaniappan, K. K.; Pitcher, A. A.; Smart, B. P.; Spiciarich, D. R.; Iavarone, A. T.; Bertozzi, C. R., Isotopic Signature Transfer and Mass Pattern Prediction (IsoStamp): An Enabling Technique for Chemically-Directed Proteomics. *ACS Chemical Biology* **2011,** *6* (8), 829-836.

113. Qin, K.; Zhu, Y.; Qin, W.; Gao, J.; Shao, X.; Wang, Y.-l.; Zhou, W.; Wang, C.; Chen, X., Quantitative Profiling of Protein O-GlcNAcylation Sites by an Isotope-Tagged Cleavable Linker. *ACS Chemical Biology* **2018,** *13* (8), 1983-1989.

114. Li, J.; Li, Z.; Duan, X.; Qin, K.; Dang, L.; Sun, S.; Cai, L.; Hsieh-Wilson, L. C.; Wu, L.; Yi, W., An Isotope-Coded Photocleavable Probe for Quantitative Profiling of Protein O-GlcNAcylation. *ACS Chemical Biology* **2019,** *14* (1), 4-10.

115. Weng, Y.; Qu, Y.; Jiang, H.; Wu, Q.; Zhang, L.; Yuan, H.; Zhou, Y.; Zhang, X.; Zhang, Y., An integrated sample pretreatment platform for quantitative N-glycoproteome analysis with combination of on-line glycopeptide enrichment, deglycosylation and dimethyl labeling. *Analytica Chimica Acta* **2014,** *833*, 1-8.

116. Xiao, K.; Tian, Z., GPSeeker Enables Quantitative Structural N-Glycoproteomics for Site- and Structure-Specific Characterization of Differentially Expressed N-Glycosylation in Hepatocellular Carcinoma. *Journal of Proteome Research* **2019,** *18* (7), 2885-2895.

117. Pan, S.; Chen, R.; Tamura, Y.; Crispin, D. A.; Lai, L. A.; May, D. H.; McIntosh, M. W.; Goodlett, D. R.; Brentnall, T. A., Quantitative Glycoproteomics Analysis Reveals Changes in N-Glycosylation Level Associated with Pancreatic Ductal Adenocarcinoma. *Journal of Proteome Research* **2014,** *13* (3), 1293-1306.

118. Chen, X.; Wu, J.; Huang, H.; Ding, Q.; Liu, X.; Chen, L.; Zha, X.; Liang, M.; He, J.; Zhu, Q.; Wang, S.; Xia, T., Comparative Profiling of Triple-Negative Breast Carcinomas Tissue Glycoproteome by Sequential Purification of Glycoproteins and Stable Isotope Labeling. *Cellular Physiology and Biochemistry* **2016,** *38* (1), 110-121.

119. Melo-Braga, M. N.; Schulz, M.; Liu, Q.; Swistowski, A.; Palmisano, G.; Engholm-Keller, K.; Jakobsen, L.; Zeng, X.; Larsen, M. R., Comprehensive Quantitative Comparison of the Membrane Proteome, Phosphoproteome, and Sialiome of Human Embryonic and Neural Stem Cells. *Molecular & Cellular Proteomics* **2014,** *13* (1), 311-328.

120. Zhang, Z.; Sun, Z.; Zhu, J.; Liu, J.; Huang, G.; Ye, M.; Zou, H., High-Throughput Determination of the Site-Specific N-Sialoglycan Occupancy Rates by Differential Oxidation of Glycoproteins Followed with Quantitative Glycoproteomics Analysis. *Analytical Chemistry* **2014,** *86* (19), 9830-9837.

121. Xue, B.; Xiao, K.; Wang, Y.; Tian, Z., Site- and structure-specific quantitative N-glycoproteomics study of differential N-glycosylation in MCF-7 cancer cells. *Journal of Proteomics* **2020,** *212*, 103594.

122. Xu, F.; Wang, Y.; Xiao, K.; Hu, Y.; Tian, Z.; Chen, Y., Quantitative site- and structure-specific N-glycoproteomics characterization of differential N-glycosylation in MCF-7/ADR cancer stem cells. *Clinical Proteomics* **2020,** *17* (1), 3.

123. Wang, Y.; Xu, F.; Chen, Y.; Tian, Z., A quantitative N-glycoproteomics study of cell-surface N-glycoprotein markers of MCF-7/ADR cancer stem cells. *Analytical and Bioanalytical Chemistry* **2020**.

124. Liu, Z.; Cao, J.; He, Y.; Qiao, L.; Xu, C.; Lu, H.; Yang, P., Tandem 18O Stable Isotope Labeling for Quantification of N-Glycoproteome. *Journal of Proteome Research* **2010,** *9* (1), 227-236.

125. Wang, J.; Zhou, C.; Zhang, W.; Yao, J.; Lu, H.; Dong, Q.; Zhou, H.; Qin, L., An integrative strategy for quantitative analysis of the N-glycoproteome in complex biological samples. *Proteome Science* **2014,** *12* (1), 4.

126. Srikanth, J.; Agalyadevi, R.; Babu, P., Targeted, Site-specific quantitation of N- and O-glycopeptides using 18O–labeling and product ion based mass spectrometry. *Glycoconjugate Journal* **2017,** *34* (1), 95-105.

127. Liu, T.; Shang, S.; Li, W.; Qin, X.; Sun, L.; Zhang, S.; Liu, Y., Assessment of Hepatocellular Carcinoma Metastasis Glycobiomarkers Using Advanced Quantitative N-glycoproteome Analysis. *Frontiers in Physiology* **2017,** *8* (472).

128. Kurogochi, M.; Amano, J., Relative Quantitation of Glycopeptides Based on Stable Isotope Labeling Using MALDI-TOF MS. *Molecules* **2014,** *19* (7), 9944-9961.

129. Pabst, M.; Benešová, I.; Fagerer, S. R.; Jacobsen, M.; Eyer, K.; Schmidt, G.; Steinhoff, R.; Krismer, J.; Wahl, F.; Preisler, J.; Zenobi, R., Differential Isotope Labeling of Glycopeptides for Accurate Determination of Differences in Site-Specific Glycosylation. *Journal of Proteome Research* **2016,** *15* (1), 326-331.

130. Kim, J. Y.; Oh, D.; Kim, S.-K.; Kang, D.; Moon, M. H., Isotope-Coded Carbamidomethylation for Quantification of N-Glycoproteins with Online Microbore Hollow Fiber Enzyme Reactor-Nanoflow Liquid Chromatography-Tandem Mass Spectrometry. *Analytical Chemistry* **2014,** *86* (15), 7650-7657.

131. Zhao, P.; Stalnaker, S. H.; Wells, L., Approaches for Site Mapping and Quantification of O-Linked Glycopeptides. In *Mass Spectrometry of Glycoproteins: Methods and Protocols*, Kohler, J. J.; Patrie, S. M., Eds. Humana Press: Totowa, NJ, 2013; pp 229-244.

132. Zhu, Z.; Go, E. P.; Desaire, H., Absolute Quantitation of Glycosylation Site Occupancy Using Isotopically Labeled Standards and LC-MS. *Journal of the American Society for Mass Spectrometry* **2014,** *25* (6), 1012-1017.

133. Roy, R.; Ang, E.; Komatsu, E.; Domalaon, R.; Bosseboeuf, A.; Harb, J.; Hermouet, S.; Krokhin, O.; Schweizer, F.; Perreault, H., Absolute Quantitation of Glycoforms of Two Human IgG Subclasses Using Synthetic Fc Peptides and Glycopeptides. *Journal of The American Society for Mass Spectrometry* **2018,** *29* (6), 1086-1098.

134. Nilsson, J.; Brinkmalm, G.; Ramadan, S.; Gilborne, L.; Noborn, F.; Blennow, K.; Wallin, A.; Svensson, J.; Abo-Riya, M. A.; Huang, X.; Larson, G., Synthetic standard aided quantification and structural characterization of amyloid-beta glycopeptides enriched from cerebrospinal fluid of Alzheimer's disease patients. *Scientific Reports* **2019,** *9* (1), 5522.

135. Shi, M.; Hwang, H.; Zhang, J., Quantitative Characterization of Glycoproteins in Neurodegenerative Disorders Using iTRAQ. In *Mass Spectrometry of Glycoproteins: Methods and Protocols*, Kohler, J. J.; Patrie, S. M., Eds. Humana Press: Totowa, NJ, 2013; pp 279-296.

136. Yang, S.; Mishra, S.; Chen, L.; Zhou, J.-y.; Chan, D. W.; Chatterjee, S.; Zhang, H., Integrated Glycoprotein Immobilization Method for Glycopeptide and Glycan Analysis of Cardiac Hypertrophy. *Analytical Chemistry* **2015,** *87* (19), 9671-9678.

137. Zhou, L.; Beuerman, R. W., Quantitative Proteomic Analysis of N-linked Glycoproteins in Human Tear Fluid. In *Mass Spectrometry of Glycoproteins: Methods and Protocols*, Kohler, J. J.; Patrie, S. M., Eds. Humana Press: Totowa, NJ, 2013; pp 297-306.

138. Bu, T.-t.; Shen, J.; Chao, Q.; Shen, Z.; Yan, Z.; Zheng, H.-y.; Wang, B.-c., Dynamic N-glycoproteome analysis of maize seedling leaves during de-etiolation using Concanavalin A lectin affinity chromatography and a nano-LC–MS/MS-based iTRAQ approach. *Plant Cell Reports* **2017,** *36* (12), 1943-1958.

139. Lee, H.-J.; Cha, H.-J.; Lim, J.-S.; Lee, S. H.; Song, S. Y.; Kim, H.; Hancock, W. S.; Yoo, J. S.; Paik, Y.-K., Abundance-Ratio-Based Semiquantitative Analysis of Site-Specific N-Linked Glycopeptides Present in the Plasma of Hepatocellular Carcinoma Patients. *Journal of Proteome Research* **2014,** *13* (5), 2328-2338.

140. Tan, Z.; Yin, H.; Nie, S.; Lin, Z.; Zhu, J.; Ruffin, M. T.; Anderson, M. A.; Simeone, D. M.; Lubman, D. M., Large-Scale Identification of Core-Fucosylated Glycopeptide Sites in Pancreatic Cancer Serum Using Mass Spectrometry. *Journal of Proteome Research* **2015,** *14* (4), 1968-1978.

141. Zhou, J.; Yang, W.; Hu, Y.; Höti, N.; Liu, Y.; Shah, P.; Sun, S.; Clark, D.; Thomas, S.; Zhang, H., Site-Specific Fucosylation Analysis Identifying Glycoproteins Associated with Aggressive Prostate Cancer Cell Lines Using Tandem Affinity Enrichments of Intact Glycopeptides Followed by Mass Spectrometry. *Analytical Chemistry* **2017,** *89* (14), 7623-7630.

142. Kawahara, R.; Ortega, F.; Rosa-Fernandes, L.; Guimarães, V.; Quina, D.; Nahas, W.; Schwämmle, V.; Srougi, M.; Leite, K. R. M.; Thaysen-Andersen, M.; Larsen, M. R.; Palmisano, G., Distinct urinary glycoprotein signatures in prostate cancer patients. *Oncotarget* **2018,** *9* (69), 33077-33097.

143. Xiao, H.; Wu, R., Quantitative investigation of human cell surface N-glycoprotein dynamics. *Chemical Science* **2017,** *8* (1), 268-277.

144. Kroksveen, A. C.; Guldbrandsen, A.; Vaudel, M.; Lereim, R. R.; Barsnes, H.; Myhr, K.-M.; Torkildsen, Ø.; Berven, F. S., In-Depth Cerebrospinal Fluid Quantitative Proteome and Deglycoproteome Analysis: Presenting a Comprehensive Picture of Pathways and Processes Affected by Multiple Sclerosis. *Journal of Proteome Research* **2017,** *16* (1), 179-194.

145. Stadlmann, J.; Hoi, D. M.; Taubenschmid, J.; Mechtler, K.; Penninger, J. M., Analysis of PNGase F-Resistant N-Glycopeptides Using SugarQb for Proteome Discoverer 2.1 Reveals Cryptic Substrate Specificities. *PROTEOMICS* **2018,** *18* (13), 1700436.

146. Yang, W.; Shah, P.; Hu, Y.; Toghi Eshghi, S.; Sun, S.; Liu, Y.; Zhang, H., Comparison of Enrichment Methods for Intact N- and O-Linked Glycopeptides Using Strong Anion Exchange and Hydrophilic Interaction Liquid Chromatography. *Analytical Chemistry* **2017,** *89* (21), 11193-11197.

147. Ye, H.; Boyne, M. T.; Buhse, L. F.; Hill, J., Direct Approach for Qualitative and Quantitative Characterization of Glycoproteins Using Tandem Mass Tags and an LTQ Orbitrap XL Electron Transfer Dissociation Hybrid Mass Spectrometer. *Analytical Chemistry* **2013,** *85* (3), 1531-1539.

148. Cho, K.-C.; Chen, L.; Hu, Y.; Schnaubelt, M.; Zhang, H., Developing Workflow for Simultaneous Analyses of Phosphopeptides and Glycopeptides. *ACS Chemical Biology* **2019,** *14* (1), 58-66.

149. Xiang, F.; Ye, H.; Chen, R.; Fu, Q.; Li, L., N,N-Dimethyl Leucines as Novel Isobaric Tandem Mass Tags for Quantitative Proteomics and Peptidomics. *Analytical Chemistry* **2010,** *82* (7), 2817-2825.

150. Greer, T.; Lietz, C. B.; Xiang, F.; Li, L., Novel isotopic N,N-dimethyl leucine (iDiLeu) reagents enable absolute quantification of peptides and proteins using a standard curve approach. *Journal of the American Society for Mass Spectrometry* **2015,** *26* (1), 107-119.

151. Frost, D. C.; Greer, T.; Xiang, F.; Liang, Z.; Li, L., Development and characterization of novel 8-plex DiLeu isobaric labels for quantitative proteomics and peptidomics. *Rapid communications in mass spectrometry : RCM* **2015,** *29* (12), 1115-1124.

152. Frost, D. C.; Rust, C. J.; Robinson, R. A. S.; Li, L., Increased N,N-Dimethyl Leucine Isobaric Tag Multiplexing by a Combined Precursor Isotopic Labeling and Isobaric Tagging Approach. *Analytical Chemistry* **2018,** *90* (18), 10664-10669.

153. Zhong, X.; Wang, J.; Carlsson, C.; Okonkwo, O.; Zetterberg, H.; Li, L., A Strategy for Discovery and Verification of Candidate Biomarkers in Cerebrospinal Fluid of Preclinical Alzheimer's Disease. *Front Mol Neurosci* **2019,** *11* (483).

154. Hao, L.; Zhong, X.; Greer, T.; Ye, H.; Li, L., Relative quantification of amine-containing metabolites using isobaric N,N-dimethyl leucine (DiLeu) reagents via LC-ESI-MS/MS and CE-ESI-MS/MS. *Analyst* **2015,** *140* (2), 467-475.

155. Zhong, X.; Frost, D. C.; Li, L., High-Resolution Enabled 5-plex Mass Defect-Based N,N-Dimethyl Leucine Tags for Quantitative Proteomics. *Analytical Chemistry* **2019,** *91* (13), 7991-7995.

156. Zhong, X.; Yu, Q.; Ma, F.; Frost, D. C.; Lu, L.; Chen, Z.; Zetterberg, H.; Carlsson, C. M.; Okonkwo, O.; Li, L., HOTMAQ: a multiplexed absolute quantification method for targeted proteomics. *Analytical Chemistry* **2019**.

157. Chen, Z.; Yu, Q.; Hao, L.; Liu, F.; Johnson, J.; Tian, Z.; Kao, W. J.; Xu, W.; Li, L., Site-specific characterization and quantitation of N-glycopeptides in PKM2 knockout breast cancer cells using DiLeu isobaric tags enabled by electron-transfer/higher-energy collision dissociation (EThcD). *Analyst* **2018,** *143* (11), 2508-2519.

158. Mayampurath, A.; Song, E.; Mathur, A.; Yu, C.-y.; Hammoud, Z.; Mechref, Y.; Tang, H., Label-Free Glycopeptide Quantification for Biomarker Discovery in Human Sera. *Journal of Proteome Research* **2014,** *13* (11), 4821-4832.

159. Goldman, R.; Sanda, M., Targeted methods for quantitative analysis of protein glycosylation. *PROTEOMICS – Clinical Applications* **2015,** *9* (1-2), 17-32.

160. Yuan, W.; Sanda, M.; Wu, J.; Koomen, J.; Goldman, R., Quantitative analysis of immunoglobulin subclasses and subclass specific glycosylation by LC–MS–MRM in liver disease. *Journal of Proteomics* **2015,** *116*, 24-33.

161. Sanda, M.; Pompach, P.; Brnakova, Z.; Wu, J.; Makambi, K.; Goldman, R., Quantitative Liquid Chromatography-Mass Spectrometry-Multiple Reaction Monitoring (LC-MS-MRM) Analysis of Site-specific Glycoforms of Haptoglobin in Liver Disease. *Molecular & Cellular Proteomics* **2013,** *12* (5), 1294-1305.

162. Sanda, M.; Pompach, P.; Brnakova, Z.; Wu, J.; Makambi, K.; Goldman, R., Quantitative LC-MS-MRM analysis of site-specific glycoforms of haptoglobin in liver disease. *Molecular & Cellular Proteomics* **2013**, mcp.M112.023325.

163. Ma, J.; Sanda, M.; Wei, R.; Zhang, L.; Goldman, R., Quantitative analysis of core fucosylation of serum proteins in liver diseases by LC-MS-MRM. *Journal of Proteomics* **2018,** *189*, 67-74.

164. Yau, L.-F.; Liu, J.; Jiang, M.; Bai, G.; Wang, J.-R.; Jiang, Z.-H., An integrated approach for comprehensive profiling and quantitation of IgG-Fc glycopeptides with application to rheumatoid arthritis. *Journal of Chromatography B* **2019,** *1122-1123*, 64-72.

165. Song, E.; Zhu, R.; Hammoud, Z. T.; Mechref, Y., LC–MS/MS Quantitation of Esophagus Disease Blood Serum Glycoproteins by Enrichment with Hydrazide Chemistry and Lectin Affinity Chromatography. *Journal of Proteome Research* **2014,** *13* (11), 4808-4820.

166. Han, J.; Liu, Q.; Xu, X.; Qin, W.; Pan, Y.; Qin, R.; Zhao, R.; Gu, Y.; Gu, J.; Ren, S., Relative Quantitation of Subclass-Specific Murine IgG Fc N-Glycoforms by Multiple Reaction Monitoring. *ACS Omega* **2020,** *5* (15), 8564-8571.

167. Ebert, B.; Rautengarten, C.; McFarlane, H. E.; Rupasinghe, T.; Zeng, W.; Ford, K.; Scheller, H. V.; Bacic, A.; Roessner, U.; Persson, S.; Heazlewood, J. L., A Golgi UDP-GlcNAc transporter delivers substrates for N-linked glycans and sphingolipids. *Nature Plants* **2018,** *4* (10), 792-801.

168. Lin, C.-H.; Chik, J. H. L.; Packer, N. H.; Molloy, M. P., Multidimensional Fractionation Is a Requirement for Quantitation of Golgi-Resident Glycosylation Enzymes from Cultured Human Cells. *Journal of Proteome Research* **2015,** *14* (2), 747-755.

169. Sanda, M.; Pompach, P.; Benicky, J.; Goldman, R., LC-MS3 quantification of O-glycopeptides in human serum. *ELECTROPHORESIS* **2013,** *34* (16), 2342-4349.

170. Kim, K. H.; Lee, S.-Y.; Hwang, H.; Lee, J. Y.; Ji, E. S.; An, H. J.; Kim, J. Y.; Yoo, J. S., Direct Monitoring of Fucosylated Glycopeptides of Alpha-Fetoprotein in Human Serum for Early Hepatocellular Carcinoma by Liquid Chromatography–Tandem Mass Spectrometry with Immunoprecipitation. *PROTEOMICS – Clinical Applications* **2018,** *12* (6), 1800062.

171. Lee, J. Y.; Kim, J. Y.; Cheon, M. H.; Park, G. W.; Ahn, Y. H.; Moon, M. H.; Yoo, J. S., MRM validation of targeted nonglycosylated peptides from N-glycoprotein biomarkers using direct trypsin digestion of undepleted human plasma. *Journal of Proteomics* **2014,** *98*, 206-217.

172. Jian, W.; Edom, R. W.; Wang, D.; Weng, N.; Zhang, S., Relative Quantitation of Glycoisoforms of Intact Apolipoprotein C3 in Human Plasma by Liquid Chromatography–High-Resolution Mass Spectrometry. *Analytical Chemistry* **2013,** *85* (5), 2867-2874.

173. Hammura, K.; Ishikawa, A.; H. V, R. K.; Miyoshi, R.; Yokoi, Y.; Tanaka, M.; Hinou, H.; Nishimura, S.-I., Synthetic Glycopeptides Allow for the Quantitation of Scarce Nonfucosylated IgG Fc N-Glycans of Therapeutic Antibody. *ACS Medicinal Chemistry Letters* **2018,** *9* (9), 889-894.

174. van der Burgt, Y. E. M.; Siliakus, K. M.; Cobbaert, C. M.; Ruhaak, L. R., HILIC–MRM–MS for Linkage-Specific Separation of Sialylated Glycopeptides to Quantify Prostate-Specific Antigen Proteoforms. *Journal of Proteome Research* **2020**.

175. Hong, Q.; Lebrilla, C. B.; Miyamoto, S.; Ruhaak, L. R., Absolute Quantitation of Immunoglobulin G and Its Glycoforms Using Multiple Reaction Monitoring. *Analytical Chemistry* **2013,** *85* (18), 8585-8593.

176. Hong, Q.; Ruhaak, L. R.; Stroble, C.; Parker, E.; Huang, J.; Maverakis, E.; Lebrilla, C. B., A Method for Comprehensive Glycosite-Mapping and Direct Quantitation of Serum Glycoproteins. *Journal of Proteome Research* **2015,** *14* (12), 5179-5192.

177. Yang, N.; Goonatilleke, E.; Park, D.; Song, T.; Fan, G.; Lebrilla, C. B., Quantitation of Site-Specific Glycosylation in Manufactured Recombinant Monoclonal Antibody Drugs. *Analytical Chemistry* **2016,** *88* (14), 7091-7100.

178. Huang, J.; Kailemia, M. J.; Goonatilleke, E.; Parker, E. A.; Hong, Q.; Sabia, R.; Smilowitz, J. T.; German, J. B.; Lebrilla, C. B., Quantitation of human milk proteins and

their glycoforms using multiple reaction monitoring (MRM). *Analytical and Bioanalytical Chemistry* **2017,** *409* (2), 589-606.

179. Miyamoto, S.; Stroble, C. D.; Taylor, S.; Hong, Q.; Lebrilla, C. B.; Leiserowitz, G. S.; Kim, K.; Ruhaak, L. R., Multiple Reaction Monitoring for the Quantitation of Serum Protein Glycosylation Profiles: Application to Ovarian Cancer. *Journal of Proteome Research* **2018,** *17* (1), 222-233.

180. Li, Q.; Kailemia, M. J.; Merleev, A. A.; Xu, G.; Serie, D.; Danan, L. M.; Haj, F. G.; Maverakis, E.; Lebrilla, C. B., Site-Specific Glycosylation Quantitation of 50 Serum Glycoproteins Enhanced by Predictive Glycopeptidomics for Improved Disease Biomarker Discovery. *Analytical Chemistry* **2019,** *91* (8), 5433-5445.

181. Lin, C.-H.; Krisp, C.; Packer, N. H.; Molloy, M. P., Development of a data independent acquisition mass spectrometry workflow to enable glycopeptide analysis without predefined glycan compositional knowledge. *Journal of Proteomics* **2018,** *172*, 68-75.

182. Zhou, C.; Schulz, B. L., Glycopeptide variable window SWATH for improved Data Independent Acquisition glycoprotein analysis. *Analytical Biochemistry* **2020**, 113667.

183. Ye, Z.; Mao, Y.; Clausen, H.; Vakhrushev, S. Y., Glyco-DIA: a method for quantitative O-glycoproteomics with in silico-boosted glycopeptide libraries. *Nature Methods* **2019**.

184. Zhu, Z.; Su, X.; Go, E. P.; Desaire, H., New Glycoproteomics Software, GlycoPep Evaluator, Generates Decoy Glycopeptides de Novo and Enables Accurate False Discovery Rate Analysis for Small Data Sets. *Analytical Chemistry* **2014,** *86* (18), 9212-9219.

185. Liu, G.; Cheng, K.; Lo, C. Y.; Li, J.; Qu, J.; Neelamegham, S., A Comprehensive, Open-source Platform for Mass Spectrometry-based Glycoproteomics Data Analysis. *Molecular & cellular proteomics : MCP* **2017,** *16* (11), 2032-2047.

186. Shipman, J. T.; Su, X.; Hua, D.; Desaire, H., DecoyDeveloper: An On-Demand, De Novo Decoy Glycopeptide Generator. *Journal of Proteome Research* **2019,** *18* (7), 2896-2902.

187. Woodin, C. L.; Hua, D.; Maxon, M.; Rebecchi, K. R.; Go, E. P.; Desaire, H., GlycoPep grader: a web-based utility for assigning the composition of N-linked glycopeptides. *Anal Chem* **2012,** *84* (11), 4821-9.

188. Bern, M.; Kil, Y. J.; Becker, C., Byonic: advanced peptide and protein identification software. *Curr Protoc Bioinformatics* **2012,** *Chapter 13*, Unit13.20-Unit13.20.

189. Yang, S.; Wu, W. W.; Shen, R.-F.; Bern, M.; Cipollo, J., Identification of Sialic Acid Linkages on Intact Glycopeptides via Differential Chemical Modification Using IntactGIG-HILIC. *Journal of The American Society for Mass Spectrometry* **2018,** *29* (6), 1273-1283.

190. Zhu, J.; Chen, Z.; Zhang, J.; An, M.; Wu, J.; Yu, Q.; Skilton, S. J.; Bern, M.; Ilker Sen, K.; Li, L.; Lubman, D. M., Differential Quantitative Determination of Site-Specific Intact N-Glycopeptides in Serum Haptoglobin between Hepatocellular Carcinoma and Cirrhosis Using LC-EThcD-MS/MS. *Journal of Proteome Research* **2019,** *18* (1), 359-371.

191. Jansen, B. C.; Falck, D.; de Haan, N.; Hipgrave Ederveen, A. L.; Razdorov, G.; Lauc, G.; Wuhrer, M., LaCyTools: A Targeted Liquid Chromatography–Mass Spectrometry Data Processing Package for Relative Quantitation of Glycopeptides. *Journal of Proteome Research* **2016,** *15* (7), 2198-2210.

192. Choo, M. S.; Wan, C.; Rudd, P. M.; Nguyen-Khuong, T., GlycopeptideGraphMS: Improved Glycopeptide Detection and Identification by Exploiting Graph Theoretical Patterns in Mass and Retention Time. *Analytical Chemistry* **2019,** *91* (11), 7236-7244.

193. Xiao, K.; Wang, Y.; Shen, Y.; Han, Y.; Tian, Z., Large-scale identification and visualization of N-glycans with primary structures using GlySeeker. *Rapid Communications in Mass Spectrometry* **2018,** *32* (2), 142-148.

194. Wang, Y.; Xu, F.; Xiao, K.; Chen, Y.; Tian, Z., Site- and structure-specific characterization of N-glycoprotein markers of MCF-7 cancer stem cells using isotopic-labelling quantitative N-glycoproteomics. *Chemical Communications* **2019,** *55* (55), 7934-7937.

195. Xiao, K.; Yu, F.; Tian, Z., Top-down protein identification using isotopic envelope fingerprinting. *Journal of Proteomics* **2017,** *152*, 41-47.

196. Stadlmann, J.; Taubenschmid, J.; Wenzel, D.; Gattinger, A.; Dürnberger, G.; Dusberger, F.; Elling, U.; Mach, L.; Mechtler, K.; Penninger, J. M., Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. *Nature* **2017,** *549* (7673), 538-542.

197. Maxwell, E.; Tan, Y.; Tan, Y.; Hu, H.; Benson, G.; Aizikov, K.; Conley, S.; Staples, G. O.; Slysz, G. W.; Smith, R. D.; Zaia, J., GlycReSoft: A Software Package for Automated Recognition of Glycans from LC/MS Data. *PLOS ONE* **2012,** *7* (9), e45474.

198. Dreyfuss, J. M.; Jacobs, C.; Gindin, Y.; Benson, G.; Staples, G. O.; Zaia, J., Targeted analysis of glycomics liquid chromatography/mass spectrometry data. *Analytical and Bioanalytical Chemistry* **2011,** *399* (2), 727-735.

199. Kronewitter, S. R.; Slysz, G. W.; Marginean, I.; Hagler, C. D.; LaMarche, B. L.; Zhao, R.; Harris, M. Y.; Monroe, M. E.; Polyukh, C. A.; Crowell, K. L.; Fillmore, T. L.; Carlson, T. S.; Camp, D. G.; Moore, R. J.; Payne, S. H.; Anderson, G. A.; Smith, R. D., GlyQ-IQ: Glycomics Quintavariate-Informed Quantification with High-Performance Computing and GlycoGrid 4D Visualization. *Analytical Chemistry* **2014,** *86* (13), 6268-6276.

200. Klein, J.; Carvalho, L.; Zaia, J., Application of network smoothing to glycan LC-MS profiling. *Bioinformatics* **2018,** *34* (20), 3511-3518.

201. Park, G. W.; Kim, J. Y.; Hwang, H.; Lee, J. Y.; Ahn, Y. H.; Lee, H. K.; Ji, E. S.; Kim, K. H.; Jeong, H. K.; Yun, K. N.; Kim, Y.-S.; Ko, J.-H.; An, H. J.; Kim, J. H.; Paik, Y.-K.; Yoo, J. S., Integrated GlycoProteome Analyzer (I-GPA) for Automated Identification and Quantitation of Site-Specific N-Glycosylation. *Scientific Reports* **2016,** *6* (1), 21175.

202. Zeng, W.-F.; Liu, M.-Q.; Zhang, Y.; Wu, J.-Q.; Fang, P.; Peng, C.; Nie, A.; Yan, G.; Cao, W.; Liu, C.; Chi, H.; Sun, R.-X.; Wong, C. C. L.; He, S.-M.; Yang, P., pGlyco: a pipeline for the identification of intact N-glycopeptides by using HCD- and CID-MS/MS and MS3. *Scientific Reports* **2016,** *6* (1), 25102.

203. Liu, M.-Q.; Zeng, W.-F.; Fang, P.; Cao, W.-Q.; Liu, C.; Yan, G.-Q.; Zhang, Y.; Peng, C.; Wu, J.-Q.; Zhang, X.-J.; Tu, H.-J.; Chi, H.; Sun, R.-X.; Cao, Y.; Dong, M.-Q.; Jiang, B.-Y.; Huang, J.-M.; Shen, H.-L.; Wong, C. C. L.; He, S.-M.; Yang, P.-Y., pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nature Communications* **2017,** *8* (1), 438.

204. Mayampurath, A.; Yu, C.-Y.; Song, E.; Balan, J.; Mechref, Y.; Tang, H., Computational Framework for Identification of Intact Glycopeptides in Complex Samples. *Analytical Chemistry* **2014,** *86* (1), 453-463.

205. Strum, J. S.; Nwosu, C. C.; Hua, S.; Kronewitter, S. R.; Seipert, R. R.; Bachelor, R. J.; An, H. J.; Lebrilla, C. B., Automated Assignments of N- and O-Site Specific Glycosylation with Extensive Glycan Heterogeneity of Glycoprotein Mixtures. *Analytical Chemistry* **2013,** *85* (12), 5666-5675.

206. Wu, S.-W.; Liang, S.-Y.; Pu, T.-H.; Chang, F.-Y.; Khoo, K.-H., Sweet-Heart — An integrated suite of enabling computational tools for automated MS2/MS3 sequencing and identification of glycopeptides. *Journal of Proteomics* **2013,** *84*, 1-16.

207. Chandler, K. B.; Pompach, P.; Goldman, R.; Edwards, N., Exploring Site-Specific N-Glycosylation Microheterogeneity of Haptoglobin Using Glycopeptide CID Tandem Mass Spectra and Glycan Database Search. *Journal of Proteome Research* **2013,** *12* (8), 3652-3666.

208. Tsai, C.-F.; Zhao, R.; Williams, S. M.; Moore, R. J.; Schultz, K.; Chrisler, W.; Pasa-Tolic, L.; Rodland, K.; Smith, R. D.; Shi, T.; Zhu, Y.; Liu, T., An improved Boosting to Amplify Signal with Isobaric Labeling (iBASIL) strategy for precise quantitative single-cell proteomics. *Molecular & Cellular Proteomics* **2020**, mcp.RA119.001857.

Chapter 5: Inclusion of Porous Graphitic Carbon Chromatography Yields Greater Protein Identification, Compartment and Process Coverage, and Enables More Reflective Protein-Level Label-Free Quantitation

Delafield, D. G., Miles, H. N., Ricke, W. A., & Li, L. Inclusion of Porous Graphitic Carbon Chromatography Yields Greater Protein Identification, Compartment and Process Coverage, and Enables More Reflective Protein-Level Label-Free Quantitation. *In submission*. 2023.

Abstract

The ubiquity of mass spectrometry-based bottom-up proteomic analyses as a component of biological investigation mandates the validation of methodologies that increase acquisition efficiency, improve sample coverage, and enhance profiling depth. Chromatographic separation is often ignored as an area of potential improvement with most analyses relying on traditional reversed-phase liquid chromatography (RPLC); this consistent reliance on a single chromatographic paradigm fundamentally limits our view of the observable proteome. Within, we build upon early reports and validate porous graphitic carbon chromatography (PGC) as a facile means to substantially enhance proteomic coverage without changes to sample preparation, instrument configuration, or acquisition method. Analysis of offline fractionated cell line digests using both separations revealed increase peptide and protein identifications by 43% and 24%, respectively. Increased identifications provided more comprehensive coverage of cellular components and biological processes independent of protein abundance, highlighting the substantial quantity of proteomic information that may go undetected in standard analyses. We further utilize these data to reveal that label-free quantitative analyses using RPLC separations alone may not be reflective of actual protein constituency. Together, these data highlight the value and comprehension offered through PGC-MS proteomic analyses. RAW proteomic data have been uploaded to the MassIVE repository with primary accession code MSV000091495.

Introduction

The long-standing need for human health- and disease-related biomolecular investigation has promoted the widespread development of numerous analytical disciplines. Among others, proteomic analyses remain a vital component of biological investigations, as these studies provide a more robust representation of functioning cells and living systems. Holistic proteomic

investigations require analysis of protein expression[1], modification[2], structure[3], and function[4], each presenting unique instrumental, preparatory and bioinformatic requirements. Mass spectrometry (MS) is now the tool-of-choice in contemporary proteomics, as this modality facilitates the breadth of measurements required and remains the only high-throughput strategy for peptide sequencing and high-resolution mass measurements[5]. The current acceptance, ubiquity, and ever-increasing expertise of MS-based proteomic analyses continues to expand the conduit towards rapid investigation of biomolecular alteration in response to external stress, disease, and treatment. However, this analytical pursuit demands continual method development and optimization. While the improvements desired in MS-based proteomics are diverse and may be discussed separately[6-9], the most fundamental need is for methodologies that enhance acquisition efficiency[10], increase sample coverage[11], and enhance profiling depth[12]. Efforts seeking to provide such improvements target either the sample preparation or instrumentation levels while chromatographic separation is relatively constant and potential improvements are underexplored[12].

By in large, high-throughput bottom-up proteomic investigations utilize reversed-phase liquid chromatography (RPLC) due to its reliability, availability, relative low cost, and extensive innovation[13-15]. This modality is preferred in bottom-up experiments as the hydrophobicity-based retention mechanism retains and separates a large portion of the average proteolytic mixture. RPLC does not, however, effectively retain hydrophilic peptides, a shortcoming exacerbated in various analyses such as post-translational modification (PTM) discovery[16, 17]. Additionally, any hydrophilic peptides that do not elute in the void volume may be poorly resolved and suffer from significant ionization suppression in the presence of more hydrophobic, basic peptides[18]. Within RPLC-MS analyses, the willful disposal of these peptides and their potential insight into protein structure and function is considered an acceptable loss in favor of high identification rates and

simpler experimental setups. For this reason, there is a critical need to implement facile, flexible experimental components that allow for these often-discarded analytes to be effectively retained, separated, and identified.

Traditionally, there have been few options when in search of chromatographic paradigms capable of retaining hydrophilic peptides. Hydrophilic interaction chromatography (HILIC) is the most common and widely reported[19-21] modality but may be considered disadvantageous as it requires mixtures to undergo phase change into organic buffers prior to analysis – an obvious limitation for hydrophilic analytes. Electrostatic repulsion-hydrophilic interaction chromatography (ERLIC)[22-25] is a recent addition to the chromatographic toolbox, reporting greater retention of hydrophilic peptides[26] and polar analytes[27-29]. A limitation of both HILIC and ERLIC is the requirement of salt-containing buffers to mitigate charge effects[23, 30] or to maintain and improve separation capacity[31], which can hinder mass spectrometry detection. Porous graphitic carbon (PGC) chromatography is an emerging chromatographic regime that has gained popularity for its ability to retain polar, hydrophilic analytes[32-34] with particular favor in the analysis of released glycans[35-45]. This separation strategy was shown to be suitable for the analysis of tryptic glycopeptides[46, 47], suggesting the utility of PGC may extend beyond metabolomic and glycomic analyses. With growing understanding of the retention mechanism, it was recently hypothesized that chromatography of this nature may be a suitable complement to traditional RPLC in untargeted, high-throughput analyses. Early reports validated this hypothesis as PGC revealed a substantial improvement in peptide and protein identification, with additional benefits seen when column temperature is optimized[48]. Stating broadly the advantages that may be seen when PGC separations are included, these initial studies did not acutely detail to what extent this additional

information serves to increase sample coverage, improve profiling depth, and affect our understanding of sample constituency.

Herein, we expand on the benefits of PGC chromatography within discovery proteomics experiments. Utilizing offline fractionation to partition prostate cancer cell line lysate, sequential analyses revealed a 43% increase in peptide identification when PGC separations are included, with almost all fractions revealing competitive identification rates between RPLC and PGC. Confident protein identifications were also increased by 23% when including PGC separations, providing greater coverage of numerous cellular compartments and biological process pathways. Interestingly, there was no significant difference in the known abundances of proteins identified through each separation, suggesting proteomic profiling can be significantly improved without the need to explore deeper into a given mixture. Finally, we compiled these data into spectral libraries that were deployed in data-independent label-free quantitative analyses. These evaluations reveal highly reproducible quantitation between PGC and RPLC separations when using the same collection of peptide precursors for quantitation. However, including the additional, complementary peptide identifications provided through PGC during quantitation produces significantly different protein expression levels than those found through RPLC alone. Overall, our work demonstrates the level of information that may go undiscovered in traditional proteomic analyses and how a narrow view of the observable proteome can impact qualitative and quantitative measurements. Despite the incalculable number of experimental optimizations intended to improve analytical throughput, each will be fundamentally hindered by a limited, chromatography-specific view of the proteome. For this reason, future development of PGC that increases retention capacity and reduces time needed to perform sequential RPLC and PGC will play a pivotal role in comprehensive proteome profiling.

Experimental

Materials

Water (H_2O, 223623) acetonitrile (ACN, A955), methanol (MeOH, A456), chloroform (C298), formic acid (FA, A117), tris base (BP152), urea (U15), and hydrochloric acid (HCl, A144SI) were purchased from Thermo Scientific. Acetone (179124), sodium dodecyl sulfate (SDS, 7173C), dithiothreitol (DTT, D9779), and iodoacetamide (IAA, I6125) were purchased from Millipore Sigma. Trypisn (V5113) was purchased from Promega (Madison, WI). RPLC packing materials (4451IP, 4472IP) were purchased from Osaka Soda Co (Osaka, Japan). PGC packing material was harvested from ThermoFisher PGC guard columns (35003-014001). Capillary tubing (1068150019) was purchased from PolyMicro. All other sources are listed.

Cell Growth

Benign prostate hypertrophy to prostate cancer (BCaP) cell lines were generated and described previously[49]. A tumorigenic cell line (BCaP-T10) and an aggressive, metastatic, tumorigenic cell line (BCaP-MT10) are used throughout these analyses. Growth conditions are listed in the supplementary information.

Protein Extraction and Digestion

Cell pellets were resuspended in 4 volumes 50mM Tris-HCl, 4% SDS prior to lysis via ultrasonication. Lysates were centrifuged to remove cell debris and protein concentration was estimated via bicinchoninic acid (BCA) assay (ThermoFisher Scientific, 23225). Disulfide bonds were reduced with 450mM DTT for 30 minutes at 55°C followed by alkylation with 10mM IAA at room temperature for 15 minutes. Protein was extracted through repeated additions of cold 80% acetone and incubation at -20°C. Protein was reconstituted in 8M urea with 1x protease inhibitor

cocktail (Roche, 05892791001 and 04906837001). Aliquots of crude protein were diluted 1:10 with 50mM Tris-HCl to reduce urea concentration to <1 M followed by two additions of trypsin for digestion. Trypsin was added 1:100 w:w and incubated for four hours at 37°C followed by a second addition of trypsin 1:50 that was left to incubate overnight at 37°C. Proteolytic mixtures were desalted, dried under vacuum, and peptide concentration was estimated via peptide assay (ThermoFisher Scientific, 23275).

HPLC Fractionation

Samples were fractionated using a Waters e2695 separation module equipped with a Waters 2489 UV-Vis detector operating in acquiring at 214 and 280nm. A Phenomenex Kinetex 2.6um PS C18 100Å column (150mm x 4.6mm) was used for separation. Buffers A and B were H_2O+0.1% FA and ACN+0.1%FA, respectively. 100ug each of T10 and MT10 lysate digest were combined, dried, and reconstituted in buffer A prior to separation. Samples were separated using a 94 min gradient of the following composition: 1% buffer B from minute 0-5, 40% buffer B at minute 50, 60% buffer B at minute 54, 70% buffer B at minute 58, 100% buffer B from minutes 59-74, 1% buffer B from minutes 74.5-94. Flow rate was set to 0.2 mL/min. Fractions were collected in 1.5-minute intervals between minutes 10-70 and were combined as described below (see Results and Discussion).

LC-MS/MS

Samples were analyzed using a Dionex nanoUltimate 3000 chromatography stack coupled to a ThermoFisher Scientific Orbitrap Fusion Lumos. Separation was performed on 15cm custom-packed capillary columns, which were prepared as described in the Supporting Information. Buffers A and B were H_2O+0.1% FA and ACN+0.1%FA, respectively. A flow rate of 350 nL/min and the following 110 gradient were used for all analyses: 3% buffer B from minutes 0-18.3, 35%

buffer B at minute 90, 95% buffer B from minutes 90.5-100, 3% buffer B from minutes 101-110. The following MS1 parameters were used for DDA analyses: resolution, 120,000; scan range, 400-1250m/z; AGC target, 2e5; maximum injection time, 50ms; intensity threshold, 2e4; charge state, 2-6; dynamic exclusion, after 1 occurrence for 45s. The MS2 parameters were as follows: resolution, 60,000; isolation width 1 m/z; activation, HCD 30; AGC target, 1e4; fixed first mass, 100m/z. For DIA analyses, MS1 settings were resolution, 120,000; scan range, 400-1250m/z; AGC target, 1e6; maximum injection time, 50ms. DIA MS2 parameters were resolution, 60,000; scan range, 200-2000m/z; isolation window 24m/z; activation, HCD 30; AGC target, 1e5; maximum injection time, 45ms; loop control N=9. All fractions and all samples were analyzed in technical duplicate.

Database Searching

DDA data were processed using FragPipe 18.0 with MSFragger[50] 3.5. An open-source Python library, easypqp, was used to generate spectral libraries from processed DDA runs; RPLC and PGC libraries were generated separately. These spectral libraries were imported to DIA-NN[51] for analysis of data-independent analyses. All parameters used within MSFragger and DIA-NN may be found in the Supporting Information.

Data and Code Availability

RAW proteomic data have been uploaded to the MassIVE repository with primary accession code MSV000091495. All code and files required to reproduce the analyses and figures presented within can be found at https://github.com/lingjunli-research/pgc-rplc-frac-profiling.

Results and Discussion

Profiling Fractionated Prostate Cancer Cell Lysate

Within mass spectrometry-based proteomic analyses, the profiling depth that may be achieved is directly tied to the efficiency with which biological mixtures are simplified during separation. Often, a single chromatographic modality does not provide the requisite simplicity for deep profiling, leading many to employ offline fractionation. As previous analyses have directly compared RPLC and PGC in shotgun analyses[48], we chose to employ offline fractionation both to further profile the level of information gained through the addition of PGC separations, as well as mimic common decomplexation techniques within bottom-up proteomics. Pooling tumorigenic and metastatic prostate cancer cell lysate, reversed phase offline fractionation was performed to partition the complex mixture and each fraction was analyzed sequentially via RPLC- and PGC-MS analysis (Figure 5.1a). After fractionation (see methods), the 280nm absorbance trace was integrated across the fractionation window (minutes 10-70). The integrated area was divided into 8 approximately equal segments, all fractions within a segment were pooled for LC-MS analysis (Figure 5.1b). As previous studies have reiterated the capacity of PGC to separate polar, hydrophilic analytes, we elected to combine fractions sequentially, keeping peptides of similar hydrophobicity together.

We hypothesized the sequential combination would result in the greatest contrast between RPLC and PGC analyses. Theoretically, the early fractions containing predominantly hydrophilic analytes should be poorly retained and elute early in subsequent RPLC-MS, whereas PGC should retain these analytes far longer and have elution profiles inversely correlated with fraction number (i.e., peptides in early fractions elute late and vice versa). Examining the time points of all confidently identified peptides, we see this theoretical expectation largely holds true in RPLC

analyses but not for PGC separations (Figure 5.2a). Rather, PGC separations demonstrate a progressive trend in peptide retention times, similar to that of RPLC analyses, suggesting that peptide hydrophobicity is not solely responsible for PGC retention. As well, we had anticipated PGC elution profiles to be more broadly distributed than those in RPLC separations, which was only marginally observed. This observation indicates the LC gradient used within our analyses – one modeled from typical RPLC experiments – is not the most appropriate for PGC separations and later optimizations will result in more effective PGC peptide separation (Figure S5.1).

Examining the overall peptide identifications within each fraction, initial comparisons show RPLC outpaces PGC across all fractions (Figure 5.2b), mirroring those observations seen elsewhere[48]. However, given the anticipated redundancy in identifications, removing peptides detected through both separation modalities reveals PGC separations are competitive, especially for those early, predominantly hydrophilic fractions (Figure 5.2**b**). Furthermore, the number of peptides specific to a single separation paradigm serves to highlight how much proteomic information may be lost during typical RPLC-MS analyses. Overall, RPLC analysis of offline fractions revealed 34,261 peptides with 21,266 unique to this separation. The inclusion of PGC separations revealed an additional 14,783 peptides, a 43% increase compared to RPLC alone, culminating in 49,044 total peptide sequences (Figure 5.2**c**). As anticipated, PGC provided greater access to those hydrophilic peptides across all fractions (Figure 5.2d) in addition to selectively retaining shorter analytes compared to RPLC (Figure 5.2e). While these high-level results are encouraging at face value, they should be further contextualized within this experiment. Here, we utilized offline fractionation, which empirically improves profiling capability of RPLC analyses. Given that we are still able to extract such an extensive quantity of additional information through

PGC analyses, it is even more clear the level of information that is lost in single-separation, RPLC shotgun proteomics.

One consideration in PGC analyses, however, is the software and parameters used during peptide identification. Within this study we utilized MSFragger, a well-recognized suite of tools with demonstrated merit[50]. A highly beneficial component of this software is the ability generate in silico tandem MS spectra and theoretical retention times that may be used as a scoring mechanism for identified peptides. Within our analyses, peptides identified in RPLC separations regularly scored higher and may be considered more confident matches than those in PGC separations (Figure 5.2f). Certainly, it is possible that all RPLC-retained peptides produced better spectra; however, peptides identified in PGC analyses also fell behind in score of the next-best peptide sequence identification, match to theoretical retention time, and PeptideProphet expectation[52] (Figure S5.2). Given the consistency with which PGC peptides score below RLPC peptides, this is most likely a reflection of how database searching tools, statistical models, and predicted expectations are largely trained upon datasets that utilize RPLC separations. We do not argue that the retention and separation capacity of RPLC is superior to that of PGC, as demonstrated here and previously[48], but given these observations and further discussion provided below, we posit that the heavy emphasis on RPLC separations in the construction and utilization of bioinformatic tools presents a fundamental limitation in the ability to correctly and confidently identify peptides in PGC-MS experiments.

Enhanced Protein Identification, Compartment Coverage and Pathway Completeness

Encouraged by the improved peptide recognition provided when PGC separations are used in addition to RPLC, we anticipated the quantity of peptides identified would directly correlate to the number of proteins identified through both separations. Considering all proteotypic peptides

identified in a given fraction, PGC and RPLC yielded virtually identical quantities of proteins except for those later fractions where RPLC excelled (Figure 5.3a). However, knowing the redundancy in peptide identifications between the two separations (Figure 5.2**c**), removing these redundant identifications reveals notable improvements in protein recognition enabled through PGC separations. Though both separations provided access to different collections of proteins, PGC outpaces RPLC in the number of unique protein identifications, especially within the inner-most fractions (Figure 5.3a). This observation is particularly valuable when considered alongside the differences in peptide identifications shown in Figure 5.2. RPLC identified more peptides overall and in all fractions except for one; however, those peptides do not map to a larger collection of proteins. This likely speaks to the known limitations in typical RPLC-MS analyses where data-dependent acquisition experiments bias towards identification of those highly abundant, hydrophobic peptides that ionize well and can cause signal suppression of unique, low-abundance analytes. PGC, which provided a greater quantity of unique protein species, likely benefits from the smaller number of peptides within each protein that are compatible with the separation modality, reducing overall number of peptides but increasing number of proteins.

In total, 3,868 proteins were identified through both separations with PGC and RPLC revealing 1,130 and 752 separation-specific proteins, respectively (Figure 5.3b). Increased identification rates alone are notable, though we hypothesized the significant increase in protein recognition likely signaled greater profiling depth across the experiment. Organizing all identified proteins into their known subcellular compartments (as listed in the UniProt knowledgebase) reveals the improved compartment coverage when PGC separations are used (Figure 5.3c). While we anticipated PGC would enable more comprehensive coverage of the cytoplasm, nucleus and other compartments with predominantly cytosolic proteins, our data also revealed PGC was able

to improve the detection of membrane and membrane-associated proteins. These observations are encouraging as it demonstrates PGC broadly provides more effective protein recognition and is not biased towards compartments dominated by hydrophilic species.

As PGC separations noticeably augmented the proteome coverage achieved through traditional RPLC-MS analyses, we allowed ourselves to consider the possibility that PGC was sampling deeper into the biological matrix, identifying species lower in abundance. When the identified proteins were sorted according to their expected abundances (see Supplemental Information), the proteins identified through both separations were those known to be highest in abundance (Figure 5.3d-e). However, proteins unique to a single separation were not significantly different in abundance, with PGC separations showing only a slightly greater density of lower abundance species. This observation alone would lead us to believe PGC does not significantly enhance profiling depth, rather it provides greater breadth. However, comparing the global protein populations provides an obtuse conclusion, as there is no information as to protein relatedness or activity. As such, we further categorized proteins according to their biological processes to determine whether PGC separations provide any better coverage of physiological pathways or protein communities. Of those biological processes represented by at least 50 members, many were enhanced through the inclusion of PGC separations, providing detection of lower abundance proteins (Figure 5.3f). For example, PGC provided greater coverage of mRNA splicing, translation, lipid metabolic process, and protein localization by identifying species lower in abundance than those seen in RPLC analyses. Noting there are other pathways where RPLC provides identification of lower abundance species (Figure 5.3f), PGC does still provide benefit in amplifying pathway coverage and revealing information that may be otherwise lost.

Overall, the peptide- and protein-level results presented here serve to illustrate the breadth and depth of information reclaimed when utilizing PGC analyses in addition to RPLC. Within this experiment, we utilized offline fractionation to reduce matrix complexity and enable greater sample coverage, anticipating PGC separations would benefit analyses of those predominantly hydrophilic fractions but would provide negligible enhancement of others. However, these expectations were largely subverted as PGC separations provided substantial increases to peptide and protein recognition in almost all cases. More interestingly, the additional proteins identified in PGC experiments showed virtually no difference in known abundance as those seen in RPLC analyses. These observations indicate current proteomics should not only focus on exploring deeper into the proteome but should also consider exploring broader. Our data shows within all analyses – not just shotgun experiments – using a single separation will always provide a limited, biased view of the proteome. By utilizing and optimizing facile, complementary separation strategies, these limitations may be systematically addressed and overcome.

PGC Analyses Enable More Representative Label-Free Protein Quantitation

Data-independent acquisition (DIA) MS is rapidly gaining favor in analysis of biological mixtures as it provides higher throughput and greater profiling depth[53]. Critically, the comprehensive and reproducible nature of DIA-MS has promoted its widespread use in label-free protein quantitation[53-55]. After confident precursor assignment, protein quantitation in DIA analyses is enabled through summating peptide or transition ion abundances or peak area, though variations to this workflow have been described[56]. As we have established the vast, complementary proteomic information provided when PGC separations are used to augment RPLC-MS analyses, our ability to quantify proteins is similarly enhanced. However, while PGC enables identification, and therefore quantitation, of proteins previously unseen in RPLC-MS, PGC also enabled the

retention and identification of additional peptides from protein sequences already identified. Knowing common label-free protein quantitation in DIA analyses utilizes summated precursor abundances, the additional peptides identified through PGC are likely to significantly impact quantitative estimations.

To investigate this claim, we compiled the data-dependent analyses of the offline fractions into two spectral libraries, one for each separation regime. After DIA analyses of tumorigenic (T10) and metastatic (MT10) prostate cancer cell line digests, these libraries were deployed for precursor assignment. Peptide identification rates resembled the trend observed in DDA analyses of fractions, though fewer were identified overall (Figure 5.4a). The number of identified proteins, however, was comparable between the two separations. This observation, which does not coincide with our DDA analyses, is likely a result of the compressed elution profile observed through PGC separations (Figure 5.4b), rearticulating the need to investigate the optimal gradient composition for this paradigm. During manual interrogation of these identifications, we noted an additional aspect of chromatographic behavior that may present limitation. Focusing on peptides identified in both separations, these peptides were almost unanimously retained longer in PGC separations (Figure 5.4c), 11.68 minutes longer on average. Expectedly, those with greatest retention time difference were generally those with highest hydrophilicity and polarity (Figure S5.3). These differences in retention time do not impact our DIA analyses as we are using empirical spectral libraries where the experimental MS spectra and retention time are known and used for identification. However, library-free analyses are gaining popularity as they are efficient and can expand profiling capacity while eliminating the need to generate extensive libraries. We posit library-free analysis built into current software is not suitable for PGC-DIA-MS analyses.

Library-free analyses, such as that offered through DIA-NN, work by using machine learning approaches to generate theoretical tandem-MS spectra and peptide retention times. These tools are constructed on the extensive body of RPLC-MS proteomics data, making them accurate, reliable tools when RPLC is the separation regime. However, the significant difference in peptide retention time observed in our PGC analyses means that library-free tools such as DIA-NN would struggle to make accurate retention time predictions and would provide limited peptide and protein identifications. We briefly investigated this claim by performing library-free analyses of our DIA datasets (see Supplemental Information). Agreeing with previous literature, library-free data processing resulted in significant improvements in the number of precursors and proteins identified in RPLC analyses (Figure S5.4). These improvements are largely due to the rigorous, well-aligned *in silico* spectra and retention times predicted for our RPLC separations. Nevertheless, library-free results for our PGC datasets were rather poor, as expected, identifying only marginally more peptides but far fewer proteins compared to our chosen spectral-library approach (Figure S5.4). As we are confident these deficiencies stem from the inability to correctly predict precursor retention time, users must either rely on empirical spectral libraries or develop custom machine learning approaches that provide rigorous, accurate retention time predictions for PGC-DIA-MS analyses.

These limitations notwithstanding, we turned our attention to ensuring technical reproducibility and quantitative accuracy. DIA-NN implements strict requirements for precursor assignment, offers matching between runs, and has a built-in FDR estimation. These features, alongside the implementation of the MaxLFQ algorithm[57], allow for highly reproducible protein-level estimations. Within our analyses, both separation paradigms provided excellent intra-sample reproducibility (Pearson R^2>0.99, Figure 5.3d) and low variance (Figure 5.3e) in protein

abundance estimations, indicating both separations are suitable for high-throughput quantitative DIA-MS analyses. To directly compare protein abundance estimations between PGC and RPLC experiments, all proteotypic peptides identified in both separations were compiled, grouped by protein precursor, and peptide MS1 areas were summed and then averaged across technical duplicates. For each of the two prostate cancer samples analyzed, approximately 2,100 proteins could be directly compared between each separation paradigm, demonstrating excellent correlation ($R^2 \approx 0.95$, Figure 5.4f). This observation indicates that extracted precursor area is conserved regardless of the separation modality employed and that protein-level estimations made through one separation modality will largely hold true in the other. Knowing this, we may reliably combine extracted precursor areas of separation-specific peptides to achieve more representative protein abundances.

To evaluate how protein quantity estimations change when PGC separations are used in tandem with RPLC, we compiled all proteotypic peptides regardless of their identifying separation, summated peptide areas and averaged technical replicates as above. We observed poor correlation (Pearson $R^2 \leq 0.5$, Figure 5.4g) of these new, adjusted protein abundances to those calculated using peptides from RPLC experiments alone. As well, protein quantities were significantly different between the two calculations, 963 proteins having notably greater calculated abundance (≥0.25 fold increase) with 465 and 133 proteins shown to be at least 1- and 2-fold greater, respectively (Figure 5.4h). These substantial differences in protein level estimations further evidence the swath of information lost or left undiscovered in routine RPLC analyses. Even if our quantitative approach was altered to utilize averaged peptide abundances or only the N-most abundant peptides from each protein, our data suggest the resulting protein abundance estimations could be significantly impacted.

Within this single experiment, PGC separations enable the retention and identification of topical peptide precursors that not only enhance proteomic coverage but also present the possibility of significantly impacting our perception of protein regulation, pathway activity, and sample constituency. As such, we hold the position that RPLC separations may never be replaced or supplanted but we can, and should, turn our attention to developing facile, high-throughput strategies that enable complementary proteomic investigations. Our data validate PGC not as the singular solution to these endeavors but as one suitable strategy that enables more comprehensive, representative analyses. We are confident PGC can gain purchase within the ever-changing analytical landscape and that engineering developments, targeted optimization, and increased utilization will help drive the proteomic investigations of tomorrow.

Conclusion

Validated methodologies that increase efficiency and enable more comprehensive sample coverage are an ever-present need in mass spectrometry-based proteomics. Whereas high-throughput measurements continue to rely on RPLC as the principal separation strategy, this report details to what extent the incorporation of PGC chromatography may enhance routine analyses. Without changes to sample preparation, gradient composition, or acquisition parameters, the inclusion of porous graphitic carbon provided a significant increase in peptide and protein identification and resulted in greater coverage of cellular compartments and biological pathways. Our report also demonstrates how these additional peptide identifications significantly impact downstream protein quantitation when compared against RPLC-MS based measurements. Ultimately, this study serves to highlight the extensive proteomic information that may be reclaimed through the simple inclusion of this nascent chromatographic paradigm and to validate this methodology as suitable for in-depth proteomic investigations. As our data indicate the

development of tailored separation and acquisition parameters will substantially improve PGC-based discovery analyses, it is clear the future development of PGC separations presents a valid, worthwhile avenue towards comprehensive proteomic analyses and label-free quantitation.

Acknowledgements

Support for this research is provided in part by the National Institutes of Health (NIH) grants RF1 AG052324 (LL), R01 DK071801 (LL), and U54DK104310 (WAR, LL). The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. LL acknowledges the National Science Foundation funding support (CHE- 2108223 and IOS-2010789), NIH grant support R21AG065728, S10OD028473, and S10OD025084 as well as a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Figures

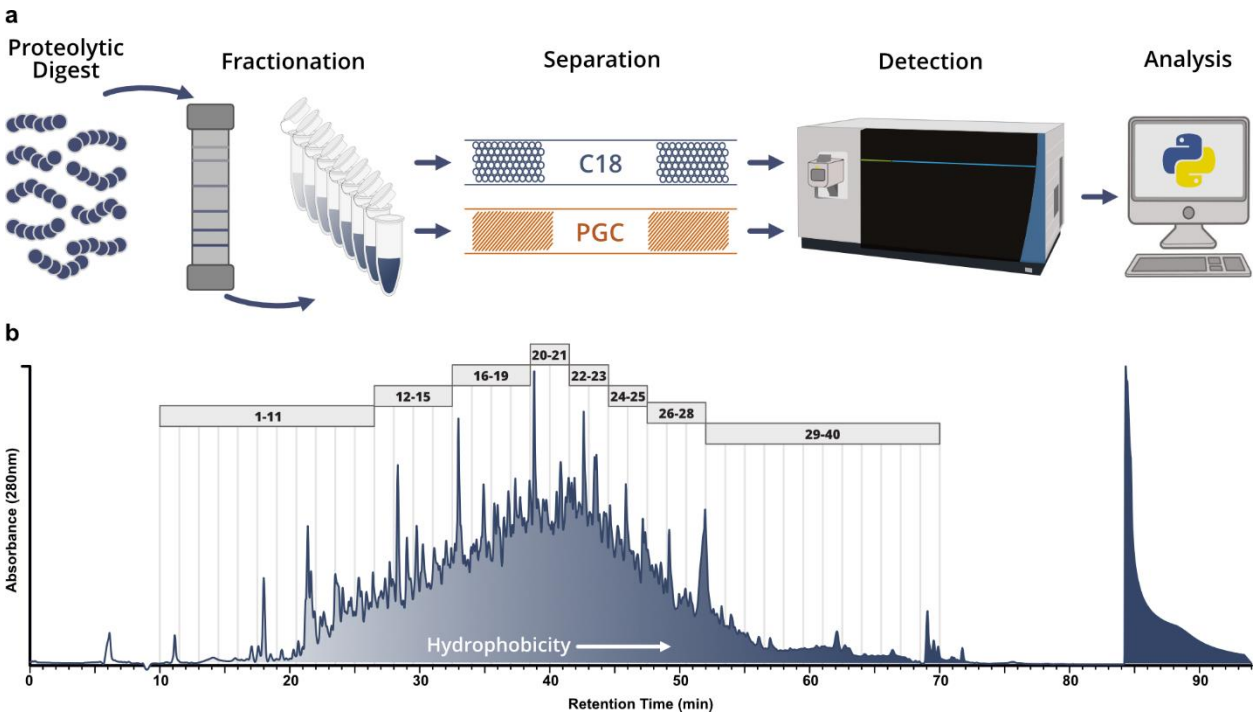


Figure 5.1 Analytical workflow and offline fractionation.

a) Graphical representation of proposed analytical workflow whereby offline-fractionated samples are analyzed through both RPLC- and PGC-MS. **b**) Absorbance (280nm) trace collected during fractionation; vertical lines represent the 1.5minute divisions of each fraction collected. The trace was integrated between minutes 10-70 and divided into 8 approximately equal components. All fractions within these 8 components (depicted by the gray boxes) were combined to make 8 final fractions used for analysis.

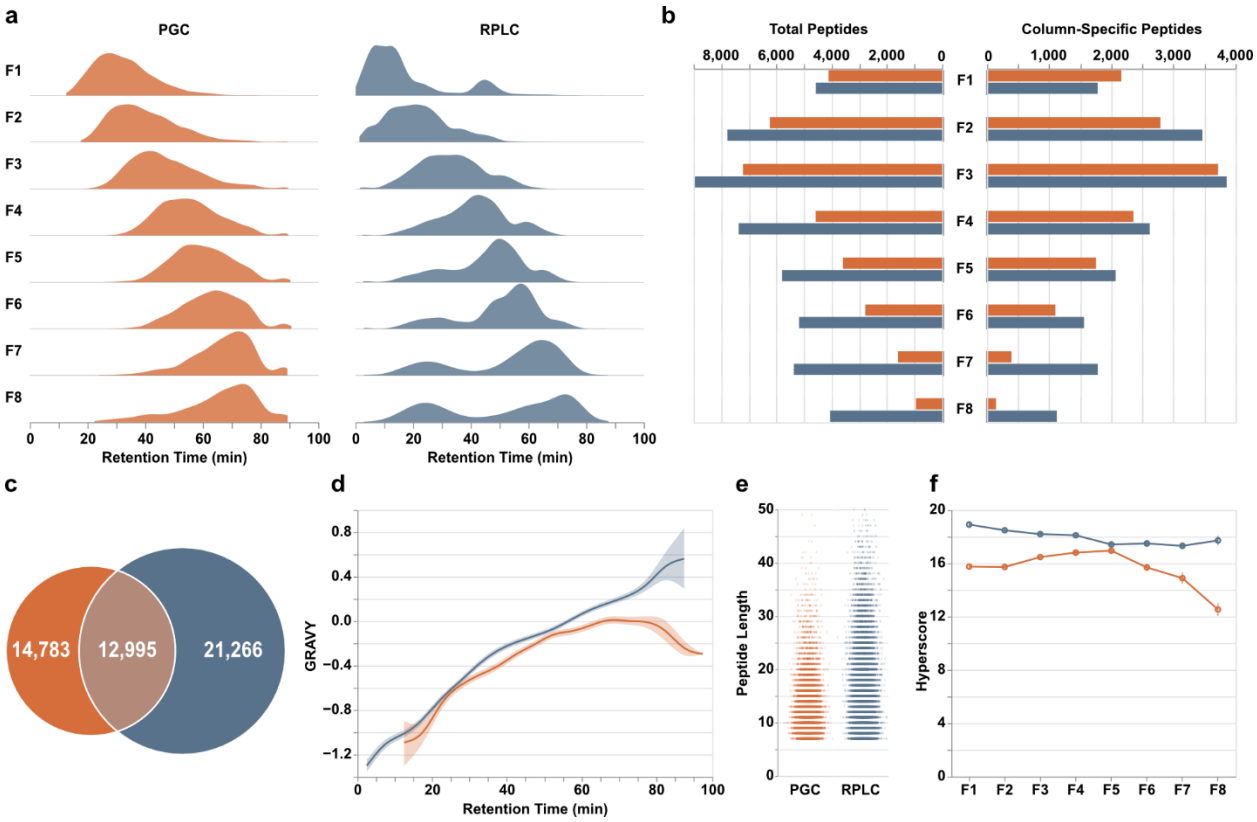


Figure 5.2 Peptide-level differences between PGC and RPLC analyses

a) Density plots displaying the time points during which peptides were identified. **b**) Total peptides (left) and number of column-specific peptides (right) identified in each fraction. "Column-specific" refers to peptides identified only through that single separation modality. **c**) Overlap of all peptides identified in all fractions. **d**) Relative hydrophilicity of all peptides identified within a given separation method. Grand average of hydropathy (GRAVY) calculations are grouped by retention time and averaged across fractions. **e**) Jitter plots displaying the length of peptides identified through both separations. **f**) Line plots displaying the average hyperscore, an MSFragger metric of confidence, for all peptides in a single fraction, partitioned according to the separation that retained them.

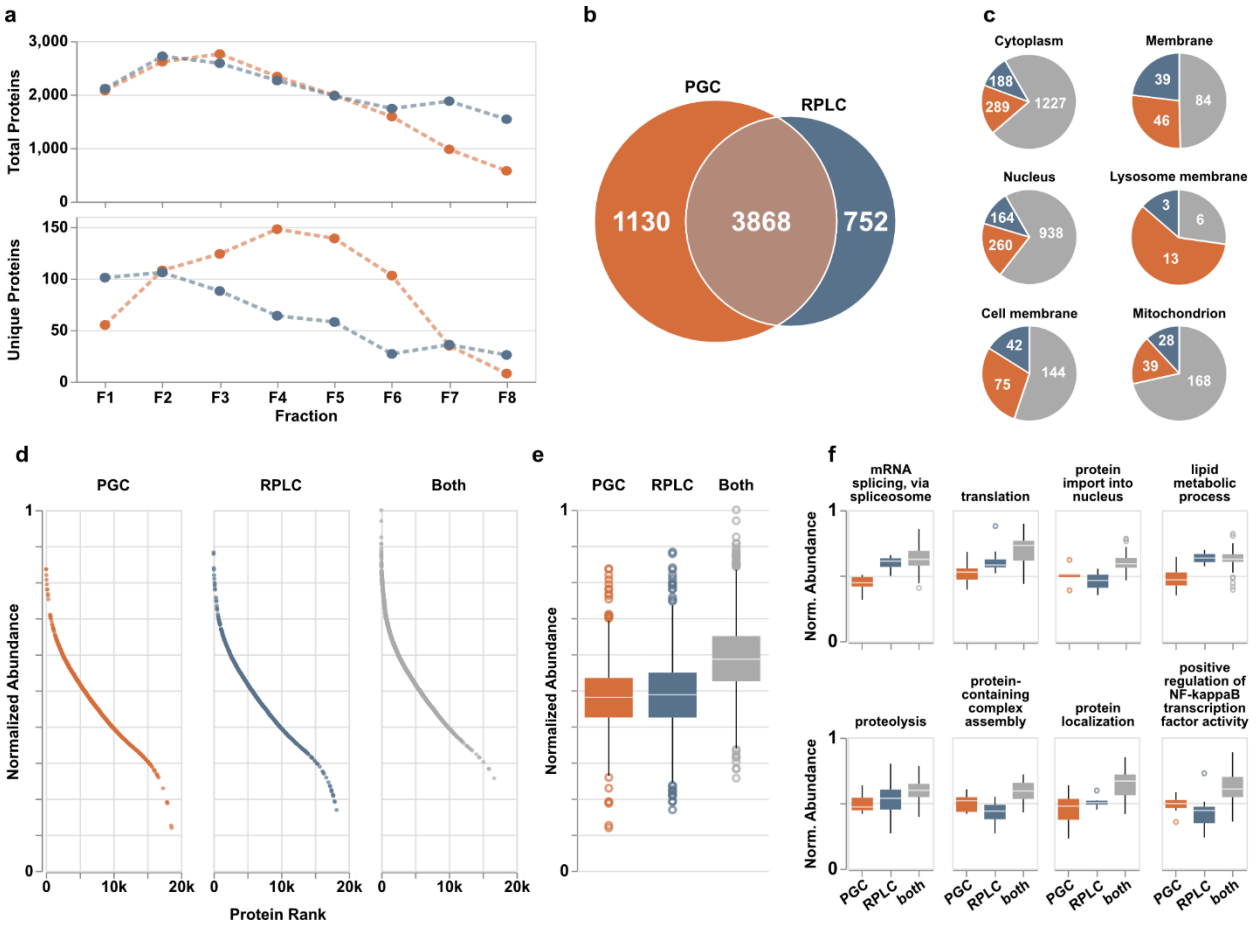


Figure 5.3 PGC analyses provide enhanced protein identification compared to RPLC alone. **a**) Quantities of total (top) and column-specific (bottom) proteins identified in each fraction. "Column-specific" refers to peptides identified only through that single separation modality. **b**) Overlap of proteins identified across all fractions. **c**) Six representative cellular compartments displaying the number of proteins localized within those compartments and through what separations they were identified; PGC (orange), RPLC (blue), both columns (gray). **d**) Proteins identified across all fractions sorted and ranked according to expected protein abundance within the human proteome. Expected abundances are normalized according to quantities estimated in the protein abundance database, Pax-DB[58] (see Supplemental Information). **e**) Boxplots displaying the distribution of protein abundances identified in either separation. **f**) Representative biological

processes identified across all fractions with boxplots displaying the distribution of protein abundances identified within those pathways.

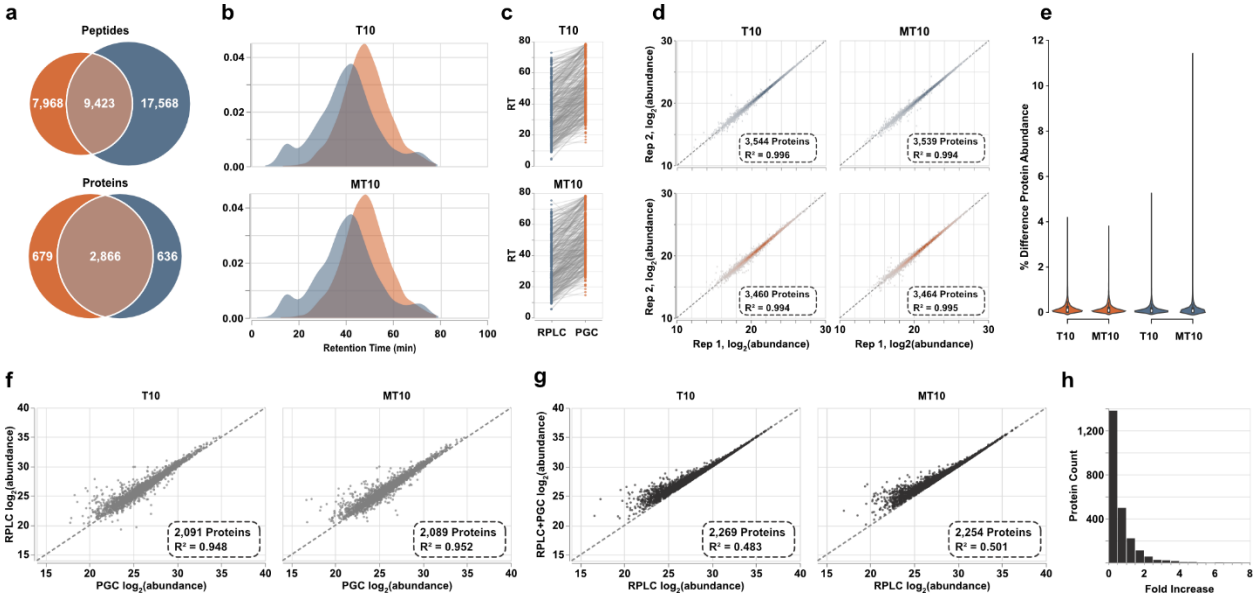


Figure 5.4 Spectral library-based DIA-MS analysis of prostate cell lines.

a) Overlap of peptides and proteins identified through both separations. Results are combined across the two cell lines analyzed. b) Density plots representing the elution profiles of peptides identified in both PGC and RPLC experiments. c) Retention times of representative peptides identified through both separation paradigms displaying the significantly later times associated with PGC analysis. d) Intrasample reproducibility of protein-level abundance calculated after DIA-MS analyses. e) Violin plots displaying the percent difference in protein abundance between technical replicates. f) Scatter plot displaying the agreement of protein abundances when using peptide precursors identified in both separation regimes. g) Protein abundances calculated using all proteotypic peptide precursors plotted against protein abundances estimated using only precursors found in RPLC analyses. h) The count of proteins showing increased abundance estimations after PGC peptides are included. Vertical bars represent count and horizontal axis is the binned fold increase as calculated by $Fold\ Increase = log_2(RPLC + PGC\ abundance) - log_2(RPLC\ abundance)$.

References

1. Ankney, J. A.; Muneer, A.; Chen, X., Relative and Absolute Quantitation in Mass Spectrometry–Based Proteomics. *Annual Review of Analytical Chemistry* **2016,** *11* (1), 49-77.

2. Brandi, J.; Noberini, R.; Bonaldi, T.; Cecconi, D., Advances in enrichment methods for mass spectrometry-based proteomics analysis of post-translational modifications. *Journal of Chromatography A* **2022,** *1678*, 463352.

3. Timp, W.; Timp, G., Beyond mass spectrometry, the next step in proteomics. *Science Advances* **2020,** *6* (2), eaax8978.

4. Parker, C. G.; Pratt, M. R., Click Chemistry in Proteomic Investigations. *Cell* **2020,** *180* (4), 605-632.

5. Gillet, L. C.; Leitner, A.; Aebersold, R., Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annual Review of Analytical Chemistry* **2016,** *9* (1), 449-472.

6. Addie, R. D.; Balluff, B.; Bovée, J. V. M. G.; Morreau, H.; McDonnell, L. A., Current State and Future Challenges of Mass Spectrometry Imaging for Clinical Research. *Analytical Chemistry* **2015,** *87* (13), 6426-6433.

7. Reinders, J.; Lewandrowski, U.; Moebius, J.; Wagner, Y.; Sickmann, A., Challenges in mass spectrometry-based proteomics. *PROTEOMICS* **2004,** *4* (12), 3686-3703.

8. Smits, A. H.; Vermeulen, M., Characterizing Protein–Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. *Trends in Biotechnology* **2016,** *34* (10), 825-834.

9. Karch, K. R.; Snyder, D. T.; Harvey, S. R.; Wysocki, V. H., Native Mass Spectrometry: Recent Progress and Remaining Challenges. *Annual Review of Biophysics* **2022,** *51* (1), 157-179.

10. Zhang, F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T., Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *PROTEOMICS* **2020,** *20* (17-18), 1900276.

11. Furtwängler, B.; Üresin, N.; Motamedchaboki, K.; Huguet, R.; Lopez-Ferrer, D.; Zabrouskov, V.; Porse, B. T.; Schoof, E. M., Real-Time Search Assisted Acquisition on a Tribrid Mass Spectrometer Improves Coverage in Multiplexed Single-Cell Proteomics. *Molecular & Cellular Proteomics* **2022**, 100219.

12. Lenz, C.; Urlaub, H., Separation methodology to improve proteome coverage depth. *Expert Review of Proteomics* **2014,** *11* (4), 409-414.

13. Mao, Z.; Hu, C.; Li, Z.; Chen, Z., A reversed-phase/hydrophilic bifunctional interaction mixed-mode monolithic column with biphenyl and quaternary ammonium stationary phases for capillary electrochromatography. *Analyst* **2019,** *144* (14), 4386-4394.

14. Liang, Y.; Zhang, L.; Zhang, Y., Monolithic Materials-Based RPLC-MS for Proteoform Separation and Identification. In *Proteoform Identification: Methods and Protocols*, Sun, L.; Liu, X., Eds. Springer US: New York, NY, 2022; pp 43-53.

15. Rozing, G., Micropillar array columns for advancing nanoflow HPLC. *Microchemical Journal* **2021,** *170*, 106629.

16. Badgett, M. J.; Boyes, B.; Orlando, R., Predicting the Retention Behavior of Specific O-Linked Glycopeptides. *J Biomol Tech* **2017,** *28* (3), 122-126.

17. Zhu, R.; Zacharias, L.; Wooding, K. M.; Peng, W.; Mechref, Y., Chapter Twenty-One - Glycoprotein Enrichment Analytical Techniques: Advantages and Disadvantages. In *Methods in Enzymology*, Shukla, A. K., Ed. Academic Press: 2017; Vol. 585, pp 397-429.

18. Oss, M.; Kruve, A.; Herodes, K.; Leito, I., Electrospray Ionization Efficiency Scale of Organic Compounds. *Analytical Chemistry* **2010,** *82* (7), 2865-2872.

19. Boersema, P. J.; Mohammed, S.; Heck, A. J. R., Hydrophilic interaction liquid chromatography (HILIC) in proteomics. *Analytical and Bioanalytical Chemistry* **2008,** *391* (1), 151-159.

20. Sun, Z.; Ji, F.; Jiang, Z.; Li, L., Improving deep proteome and PTMome coverage using tandem HILIC-HPRP peptide fractionation strategy. *Analytical and Bioanalytical Chemistry* **2019,** *411* (2), 459-469.

21. Badgett, M. J.; Boyes, B.; Orlando, R., Peptide retention prediction using hydrophilic interaction liquid chromatography coupled to mass spectrometry. *Journal of Chromatography A* **2018,** *1537*, 58-65.

22. Yeh, T.-T.; Ho, M.-Y.; Chen, W.-Y.; Hsu, Y.-C.; Ku, W.-C.; Tseng, H.-W.; Chen, S.-T.; Chen, S.-F., Comparison of different fractionation strategies for in-depth phosphoproteomics by liquid chromatography tandem mass spectrometry. *Analytical and Bioanalytical Chemistry* **2019,** *411* (15), 3417-3424.

23. Cui, Y.; Tabang, D. N.; Zhang, Z.; Ma, M.; Alpert, A. J.; Li, L., Counterion Optimization Dramatically Improves Selectivity for Phosphopeptides and Glycopeptides in Electrostatic Repulsion-Hydrophilic Interaction Chromatography. *Analytical Chemistry* **2021,** *93* (22), 7908-7916.

24. Cui, Y.; Yang, K.; Tabang, D. N.; Huang, J.; Tang, W.; Li, L., Finding the Sweet Spot in ERLIC Mobile Phase for Simultaneous Enrichment of N-Glyco and Phosphopeptides. *Journal of the American Society for Mass Spectrometry* **2019,** *30* (12), 2491-2501.

25. Alpert, A. J., Electrostatic Repulsion Hydrophilic Interaction Chromatography for Isocratic Separation of Charged Solutes and Selective Isolation of Phosphopeptides. *Analytical Chemistry* **2008,** *80* (1), 62-76.

26. Zhen, J.; Kim, J.; Zhou, Y.; Gaidamauskas, E.; Subramanian, S.; Feng, P., Antibody characterization using novel ERLIC-MS/MS-based peptide mapping. *mAbs* **2018,** *10* (7), 951-959.

27. Yan, J.; Ding, J.; Jin, G.; Duan, Z.; Yang, F.; Li, D.; Zhou, H.; Li, M.; Guo, Z.; Chai, W.; Liang, X., Profiling of Human Milk Oligosaccharides for Lewis Epitopes and Secretor Status by Electrostatic Repulsion Hydrophilic Interaction Chromatography Coupled with Negative-Ion Electrospray Tandem Mass Spectrometry. *Analytical Chemistry* **2019,** *91* (13), 8199-8206.

28. Qing, G.; Yan, J.; He, X.; Li, X.; Liang, X., Recent advances in hydrophilic interaction liquid interaction chromatography materials for glycopeptide enrichment and glycan separation. *TrAC Trends in Analytical Chemistry* **2020,** *124*, 115570.

29. Bermudez, A.; Pitteri, S. J., Enrichment of Intact Glycopeptides Using Strong Anion Exchange and Electrostatic Repulsion Hydrophilic Interaction Chromatography. In *Mass Spectrometry of Glycoproteins: Methods and Protocols*, Delobel, A., Ed. Springer US: New York, NY, 2021; pp 107-120.

30. Alpert, A. J., Effect of salts on retention in hydrophilic interaction chromatography. *Journal of Chromatography A* **2018,** *1538*, 45-53.

31. West, C.; Auroux, E., Deconvoluting the effects of buffer salt concentration in hydrophilic interaction chromatography on a zwitterionic stationary phase. *J Chromatogr A* **2016,** *1461*, 92-7.

32. West, C.; Elfakir, C.; Lafosse, M., Porous graphitic carbon: A versatile stationary phase for liquid chromatography. *Journal of Chromatography A* **2010,** *1217* (19), 3201-3216.

33. Bapiro, T. E.; Richards, F. M.; Jodrell, D. I., Understanding the Complexity of Porous Graphitic Carbon (PGC) Chromatography: Modulation of Mobile-Stationary Phase Interactions Overcomes Loss of Retention and Reduces Variability. *Analytical Chemistry* **2016,** *88* (12), 6190-6194.

34. Stavenhagen, K.; Hinneburg, H.; Kolarich, D.; Wuhrer, M., Site-Specific N- and O-Glycopeptide Analysis Using an Integrated C18-PGC-LC-ESI-QTOF-MS/MS Approach. In *High-Throughput Glycomics and Glycoproteomics: Methods and Protocols*, Lauc, G.; Wuhrer, M., Eds. Springer New York: New York, NY, 2017; pp 109-119.

35. Xu, J.; Liu, X.; Zhou, H., Recent advances in separation methods for post-translational modification proteomics. *Sepu* **2016,** *34* (12), 1199-1205.

36. Zhou, S.; Huang, Y.; Dong, X.; Peng, W.; Veillon, L.; Kitagawa, D. A. S.; Aquino, A. J. A.; Mechref, Y., Isomeric Separation of Permethylated Glycans by Porous Graphitic Carbon (PGC)-LC-MS/MS at High Temperatures. *Analytical Chemistry* **2017,** *89* (12), 6590-6597.

37. Zhou, S.; Dong, X.; Veillon, L.; Huang, Y.; Mechref, Y., LC-MS/MS analysis of permethylated N-glycans facilitating isomeric characterization. *Analytical and Bioanalytical Chemistry* **2017,** *409* (2), 453-466.

38. Huang, Y.; Zhou, S.; Zhu, J.; Lubman David, M.; Mechref, Y., LC-MS/MS isomeric profiling of permethylated N-glycans derived from serum haptoglobin of hepatocellular carcinoma (HCC) and cirrhotic patients. *ELECTROPHORESIS* **2017,** *38* (17), 2160-2167.

39. Ashwood, C.; Lin, C.-H.; Thaysen-Andersen, M.; Packer, N. H., Discrimination of Isomers of Released N- and O-Glycans Using Diagnostic Product Ions in Negative Ion PGC-LC-ESI-MS/MS. *Journal of the American Society for Mass Spectrometry* **2018,** *29* (6), 1194-1209.

40. Hinneburg, H.; Chatterjee, S.; Schirmeister, F.; Nguyen-Khuong, T.; Packer, N. H.; Rapp, E.; Thaysen-Andersen, M., Post-Column Make-Up Flow (PCMF) Enhances the Performance of Capillary-Flow PGC-LC-MS/MS-Based Glycomics. *Analytical Chemistry* **2019,** *91* (7), 4559-4567.

41. Ashwood, C.; Pratt, B.; MacLean, B. X.; Gundry, R. L.; Packer, N. H., Standardization of PGC-LC-MS-based glycomics for sample specific glycotyping. *The Analyst* **2019,** *144* (11), 3601-3612.

42. Ashwood, C.; Waas, M.; Weerasekera, R.; Gundry, R. L., Reference glycan structure libraries of primary human cardiomyocytes and pluripotent stem cell-derived cardiomyocytes reveal cell-type and culture stage-specific glycan phenotypes. *Journal of Molecular and Cellular Cardiology* **2020,** *139*, 33-46.

43. Wei, J.; Tang, Y.; Bai, Y.; Zaia, J.; Costello, C. E.; Hong, P.; Lin, C., Toward Automatic and Comprehensive Glycan Characterization by Online PGC-LC-EED MS/MS. *Analytical Chemistry* **2020,** *92* (1), 782-791.

44. Chen, C.-H.; Lin, Y.-P.; Ren, C.-T.; Shivatare, S. S.; Lin, N.-H.; Wu, C.-Y.; Chen, C.-H.; Lin, J.-L., Enhancement of fucosylated N-glycan isomer separation with an ultrahigh column temperature in porous graphitic carbon liquid chromatography-mass spectrometry. *Journal of Chromatography A* **2020,** *1632*, 461610.

45. Riley, N. M.; Bertozzi, C. R.; Pitteri, S. J., A Pragmatic Guide to Enrichment Strategies for Mass Spectrometry-based Glycoproteomics. *Molecular & Cellular Proteomics* **2021**, 100029.

46. Zhu, R.; Huang, Y.; Zhao, J.; Zhong, J.; Mechref, Y., Isomeric Separation of N-Glycopeptides Derived from Glycoproteins by Porous Graphitic Carbon (PGC) LC-MS/MS. *Analytical Chemistry* **2020,** *92* (14), 9556-9565.

47. Delafield, D. G.; Miles, H. N.; Ricke, W. A.; Li, L., Higher Temperature Porous Graphitic Carbon Separations Differentially Impact Distinct Glycopeptide Classes. *Journal of the American Society for Mass Spectrometry* **2022**.

48. Delafield, D. G.; Miles, H. N.; Liu, Y.; Ricke, W. A.; Li, L., Complementary proteome and glycoproteome access revealed through comparative analysis of reversed phase and porous graphitic carbon chromatography. *Analytical and Bioanalytical Chemistry* **2022**.

49. Liu, T. T.; Ewald, J. A.; Ricke, E. A.; Bell, R.; Collins, C.; Ricke, W. A., Modeling human prostate cancer progression in vitro. *Carcinogenesis* **2018,** *40* (7), 893-902.

50. 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. *Nature Methods* **2017,** *14* (5), 513-520.

51. Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M., DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods* **2020,** *17* (1), 41-44.

52. Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R., Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Analytical Chemistry* **2002,** *74* (20), 5383-5392.

53. 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* **2020,** *13*.

54. Meyer, J. G.; Schilling, B., Clinical applications of quantitative proteomics using targeted and untargeted data-independent acquisition techniques. *Expert Review of Proteomics* **2017,** *14* (5), 419-429.

55. Meyer, J. G.; Niemi, N. M.; Pagliarini, D. J.; Coon, J. J., Quantitative shotgun proteome analysis by direct infusion. *Nature Methods* **2020,** *17* (12), 1222-1228.

56. Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R., Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Molecular Systems Biology* **2018,** *14* (8), e8126.

57. Cox, J.; Hein, M. Y.; Luber, C. A.; Paron, I.; Nagaraj, N.; Mann, M., Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **2014,** *13* (9), 2513-26.

58. Wang, M.; Herrmann, C. J.; Simonovic, M.; Szklarczyk, D.; von Mering, C., Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **2015,** *15* (18), 3163-8.

Supplemental Methods

Cell Growth

Cell lines were grown and maintained in phenol-free DMEM/Ham's F12 (Gibco) supplemented with 5% fetal bovine serum (HyClone) and 1% penicillin-streptomycin solution (Gibco). T175 culture flasks were placed in an incubator at 5% CO2 and 98% humidity. Cells were grown to 90% confluency, washed with 1× phosphate-buffered saline (Cytiva), and harvested using a cell scraper. Approximately 3.5e6 cells were collected after culture. Cell pellets were washed twice using phosphate-buffered saline and stored at −80°C for subsequent processing.

Column Preparation

25 cm of capillary (360um o.d., 75um i.d.) were taken for column packing. Using a flame, a small 3cm portion of the capillary coating was removed, approximately 5 cm from one end. An emitter was pulled using a Sutter P2000 laser puller. Column shells were flushed with MeOH to ensure adequate spray. RPLC columns were packed with 3um packing material at 100psi for 5 seconds followed by packing with 1.7um packing material at 1500psi. Columns were packed until 15cm of the capillary was filled. Packing material for RPLC columns was suspended in chloroform. For PGC columns, this process is identical, save for the packing step. Packing material, 3um, was suspended in MeOH and packed at 1500psi until a final length of 15cm was reached.

Data Processing

DDA raw files were processed using FragPipe 18.0 with MSFragger 3.5 and IonQuant 1.8. All technical duplicates were assigned as part of the same experiment to yield one result file per fraction and all files were searched against the UniProt Human proteome (December 2021). MS1

and MS2 mass tolerance were set to 20ppm, enzymatic digestion was set to strict K and R cleavage and two allowed missed cleavage sites, peptide length 7-50 and mass 500-5000 were allowed. Carbamidomethyl of C was set as a fixed modification with the following modifications considered as variable: protein N-terminal acetylation (1 occurrence per peptide), peptide N-terminal acetylation (1 occurrence per peptide), oxidation of Met (3 occurrence per peptide), deamidation of Asn and Glu (3 occurrence per peptide). All results were filtered to a 1% FDR prior to analysis.

Spectral libraries were generated separately for RPLC and PGC separations and were imported to DIA-NN for analysis. Analyses were searched against the UniProt human proteome (December 2021) with Trypsin/P digestion and two allowed missed cleavage sites. N-terminal Met excision, carbamidomethylation of Cys, oxidation of Met, and N-terminal acetylation were selected as modifications with 4 variable modifications allowed per peptide. Peptides of length 7-50, charge 2-7, precursor range 400-1250m/z, fragment ion range 200-2000m/z were considered. FDR cutoff was set to 1%. Match between runs was selected and proteins were inferenced base on protein names. All other options were set as default. RPLC and PGC analyses were searched separately. For library free analyses within DIA-NN, the same parameters were used but no spectral libraries were provided. Rather, we allowed DIA-NN to create an *in silico* spectral library using the UniProt Human Proteome (December 2021).

Expected Protein Abundance

To aid in our understanding of profiling depth (i.e., whether a single separation paradigm samples proteins known to be lower in abundance or not), we utilized Pax-DB, an external resource containing empirical protein abundance measurements. We selected the integrated Homo sapiens – whole organism dataset (https://pax-db.org/dataset/9606/1502934799/, released 2021), which covers 99% of the human proteome and contains 19,338 abundance values. Abundance

values this dataset were converted to a log 10 scale using the equation $Log\ Abundance = log_{10}(PaxDB\ Abundance\ x\ 1000)$ and were then normalized to the maximum abundance value within the dataset to yield "normalized abundance." Throughout the main text, we refer to "expected" or "anticipated" abundance, referring to the relative abundance at which we expect these proteins to be seen in the average proteomic measurement. As we do not have quantitative data within our DDA analyses, the empirical abundance values for our samples cannot be stated.

Supplemental Tables

The supplemental tables listed below are found within a single Excel file with multiple, titled sheets. As a note, this data is not needed to recreate the analyses used within the manuscript; the original result files from MSFragger and DIA-NN have been uploaded to the github repository. Cloning the repository will let you recreate all figures.

Supplemental Figures

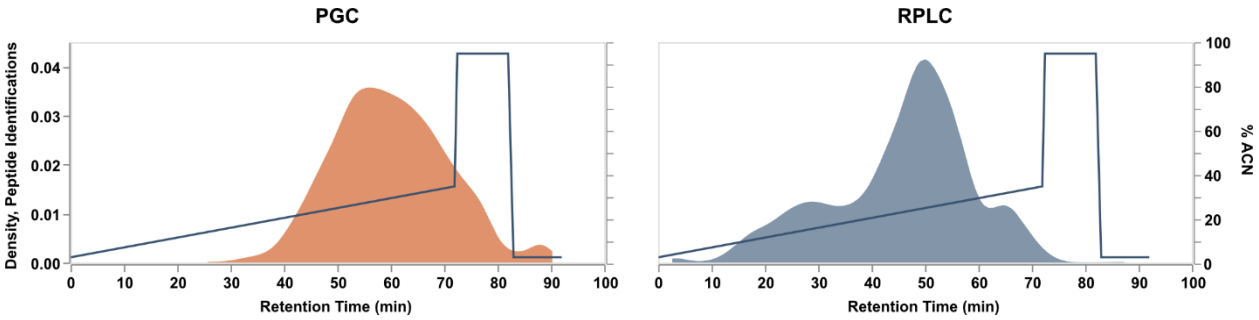


Figure S5.1 Comparison of peptide elution/identification times with respect to gradient composition.

A primary consideration raised in this work is the use of appropriate gradients. For consistency, both separations utilized a typical gradient used for RPLC analyses, which is indicated by the overlaid blue line and right Y-axis. As shown in this representative data (DDA analyses of fraction 5), RPLC separations (blue) were distributed across the useable gradient. However, PGC (left) analyses of the same sample with the same gradient displays a condensed elution profile with most identifications coming at 20%+ ACN, indicating typical proteomics gradients can work but are less suitable for these separations.

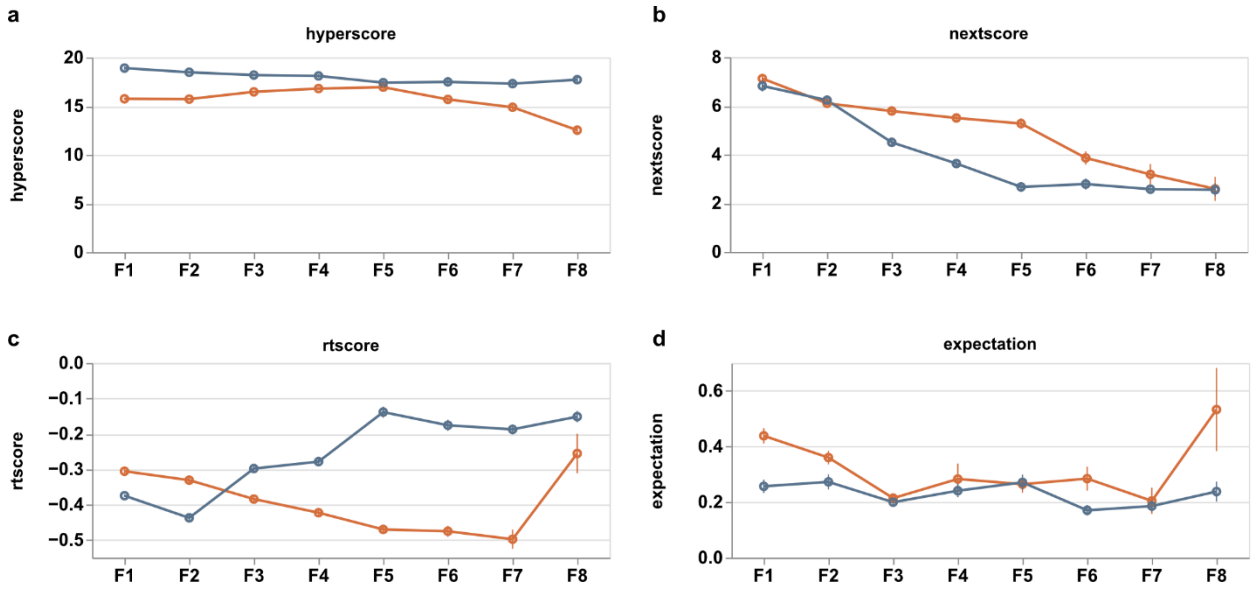


Figure S5.2 Quality metrics of peptides identified through MSFragger.

a) Comparison of hyperscore, a measure of similarity between experimental and theoretical spectra. The higher values for RPLC experiments (blue) indicate consistently better matches. b) Comparison of nextscore, the similarity of next-best scoring peptide to the experimental spectra. The higher values of PGC (orange) represents less well-defined differences in spectral quality between first- and second-best peptide matches. c) Comparison of rtscore, a metric indicating the match between theoretical and experimental retention times. The lower values of PGC (orange) indicate the retention times found largely do not match those expected. This highlights a prevalent bias in bioinformatic tools towards RPLC-based analyses. d) Comparison of expectation values from PeptideProphet; lower values indicate higher likelihood.

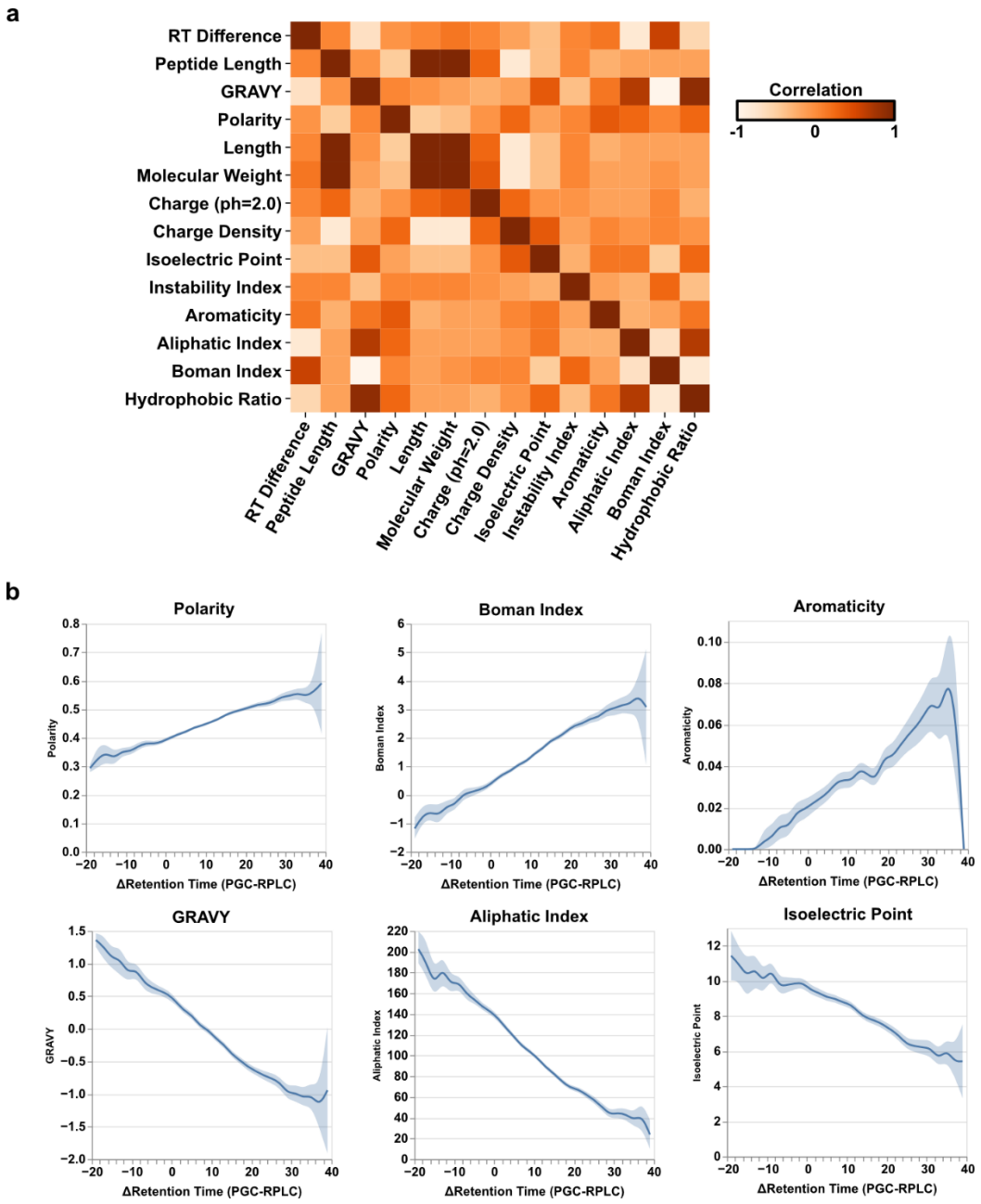


Figure S5.3 Peptide characteristics driving retention time differences in PGC separations.

a) Correlation matrix comparing numerous calculated peptide descriptors with respect to the determined RT difference. $\Delta RT = RT_{PGC} - RT_{RPLC}$ b) Line and error band plots of those peptide characteristics most responsible for RT difference PGC separations. Overall, the data suggest the hydrophilicity is a major driving factor, which is exacerbated by increased prevalence of aromatic residues and reduced prevalence of basic residues.

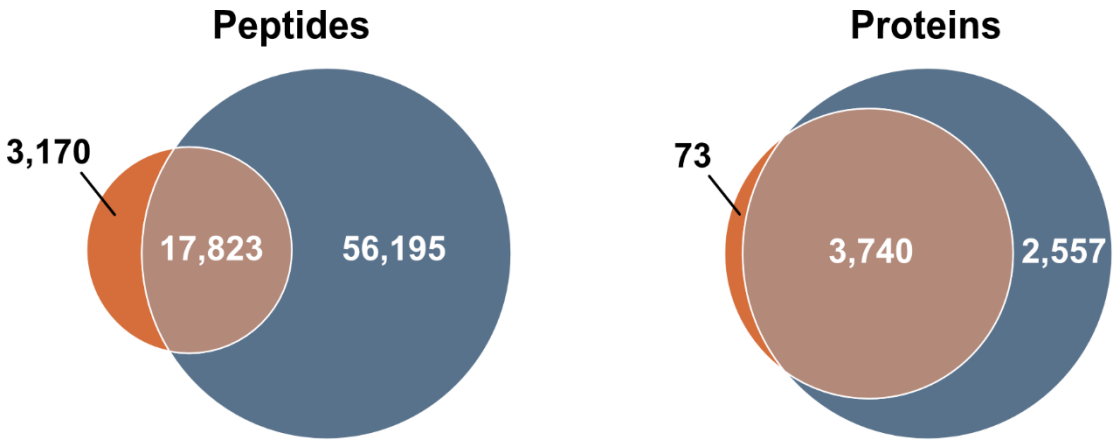


Figure S5.4 Overlap of peptide precursor and protein identifications identified in library free analysis.

While RPLC separations (blue) benefitted substantially from the use of library free analyses, as evidenced by the significant increase in precursor assignments, PGC separations (orange) only revealed moderately more peptides. In addition, library-free analyses resulted in far fewer identified proteins for PGC analyses compared to the DDA experiments. While the compressed elution profiles mentioned in the main text did not help identification rates, the substantial reduction of unique species for PGC separations confirm there is incongruency between the retention mechanism and the predictions made in DIA-NN.

Chapter 6: Recent Developments and Applications of Quantitative Proteomics Strategies for High-Throughput Biomolecular Analyses in Cancer Research

Miles, H. N.[†], **Delafield, D. G**.[†], & Li, L. (2021). Recent developments and applications of quantitative proteomics strategies for high-throughput biomolecular analyses in cancer research. *RSC Chemical Biology*.

[†] These authors contributed equally to this work.

Abstract

Innovations in medical technology and dedicated focus from the scientific community have inspired numerous treatment strategies for benign and invasive cancers. While these improvements often lend themselves to more positive prognoses and greater patient longevity, means for early detection and severity stratification have failed to keep pace. Detection and validation of cancer-specific biomarkers hinges on the ability to identify subtype-specific phenotypic and proteomic alterations and the systematic screening of diverse patient groups. For this reason, clinical and scientific research settings rely on high throughput and high sensitivity mass spectrometry methods to discover and quantify unique molecular perturbations in cancer patients. Discussed within is an overview of quantitative proteomics strategies and a summary of recent applications that enable revealing potential biomarkers and treatment targets in prostate, ovarian, breast, and pancreatic cancer in a high throughput manner.

Introduction

Mass spectrometry (MS) represents a unique and powerful technological platform in investigative biomolecular research. This high sensitivity regime grants access to the discovery and identification of small molecules,[1-3] endogenous peptides,[4-7] proteins,[8-10] and macromolecular complexes.[11-13] The utility of MS is enhanced through the facility of ionizing biomolecule species in solution via electrospray ionization[14] (ESI) and matrix-assisted laser desorption/ionization[15-17] (MALDI) that provides a means of producing ions from stationary supports and tissue sections. As well, numerous mass analyzers[18] have been developed to accommodate high-speed and high-resolution measurements. Realizing the full potential and flexibility of modern MS platforms, as well as their ability to decipher complex biological samples, focus has shifted towards improving instrument efficiency and sample throughput.

Gradual improvements in instrument operational speed, the advent of novel dissociation techniques[19-22] and implementation of multidimensional ion separation regimes[23-25] enable researchers to obtain greater levels of detail from complex mixtures than ever before. However, while shotgun proteomics provides a means for deep proteomic profiling, the typical time course and complexity of a single experiment[26] renders repetitive measurements of numerous samples untenable. For this reason, many have turned to multiplexed quantitative proteomics workflows to provide simultaneous deep proteomic profiling of numerous samples while retaining the ability to assign relative and absolute abundance information.

Quantitative proteomics, now comprised of several distinct strategies, operates under the principle that signal response from any given analyte is related to its abundance within the mixture. As such, should an analyte be identified in numerous samples, the relative intensity of the analyte's signal response in each sample can be used to provide a means of relative or absolute quantitation.

However, MS reporting signal is divided into numerous channels depending on the quantity and ionization efficiency of all present biomolecules, which indicates the high variability that can arise from even discrete sample changes. In remedy of this ailment—and to remove run-to-run variation—researchers have employed unique chemical modifiers that often incorporate stable isotopes to label biomolecules within solution. These labels result in a unique mass shift for each sample without altering their retention time in liquid-chromatography. These newly tagged analytes may then be combined, measured simultaneously via MS, and then evaluated for the relative abundance of all labeled channels.

These quantitative strategies have provided unique avenues towards the discovery and validation of cancer-specific biomarkers. The ability to analyze numerous samples simultaneously provides researchers not only with a means for high throughput sample profiling, but also a means to uncover what proteomic perturbations are relevant across patients, between control groups, and specific to disease severity and progression. These perturbations and quantitative differences are often discussed in language that is familiar to proteomic researchers but that may create confusion in those coming from adjacent fields of research. Within proteomics, and ubiquitous throughout this review, quantitative differences of proteins, peptides and other biomolecules are described as "up-regulated" or "down-regulated." These terms are used to describe those species with quantifiable differences against the control, often with statistical significance. Though readers may conjecture that up- or down-regulated protein species are the result of pathway regulation, these hypotheses are often not explored in proteomic literature and may be discussed elsewhere. For this reason, it is important to clarify that differences in regulation are meant only to indicate the quantitative findings presented by the original authors. Regardless of verbiage technicalities,

researchers often pursue quantitative proteomics as a facile avenue towards novel biological insight.

Given the significant heterogeneity associated with various cancer subtypes, researchers have sought to employ quantitative proteomics to a litany of biological questions. As seen within, these endeavors have provided seminal insights into the role post-translational modifications play in cancer progression, uncovering up- and down-regulation of biomolecules in disease groups, as well as the efficacy of using protein expression to monitor medical treatment. The true breadth of proteomic cancer research cannot be understated. While quantitative experiments date back several decades, we aim to present a mass spectrometry-centric review. High-throughput quantitative proteomics firmly gained prominence in the early 2000s, providing nearly twenty years of meaningful contributions to cancer detection, identification, and understanding. To provide readers with the most timely and topical review—as well as to provide discussion on future research interests, we have confined our literature review to applications seen within the past 5 years. This concise range enables us to provide critical suggestions for researchers seeking to begin or continue their unique cancer research. Here we present a brief introduction to quantitative proteomics methods and recent investigations of prostate, ovarian, breast, and pancreatic cancer.

Quantitative Strategies

Quantitative proteomics has experienced substantial growth over the last two decades, due in large part to the invention and development of high-speed, high-resolution mass spectrometry instrumentation. While there are numerous unique and technically driven means to pursue relative and absolute protein quantitation, most applications fall within one of four major categories: metabolic labeling, isotopic labeling, isobaric labeling, and label-free quantitation. Each method has been thoroughly reviewed and in-depth discussion can be found elsewhere. However, in order

to provide rationale behind each strategy for use in investigative cancer research, understanding the principles and key considerations of each is imperative.

Metabolic Labeling

Metabolic labeling is the earliest[27, 28] and arguably most traditional method of mass spectrometry-based quantitative proteomics. Taking after the classical Meselson-Stahl[29] experiment that proved the semiconservative nature of DNA replication, more routine use of mass spectrometry for peptide identification revealed that proteins, too, could be metabolically labeled with stable isotopes to provide 'heavy' and 'light' isotopologues. Within these experiments, adjacent cell cultures are provided with either unlabeled, naturally occurring amino acids or amino acids that have been labeled with stable isotopes; this also lends itself to the acronym SILAC, Stable Isotopic Labeling with Amino Acids in Cell Culture.[27, 30] Though SILAC has grown to incorporate numerous stable isotopes, the most traditional SILAC strategy is to grow a control group in the presence of ^{12}C-Lysine and ^{13}N- Arginine while providing the experimental group with ^{13}C-Lysine and ^{15}N-Arginine.[31] During culture growth, these light or heavy amino acids are incorporated into the protein backbones with no effect on protein function, viability or expression. Digesting these cellular proteins with a proteolytic enzyme (e.g. trypsin) produces peptides that contain a single labeled or unlabeled residue. Peptides are then combined and analyzed via MS, at which point their mass difference can be observed. Evaluating the intensities of the labeled and unlabeled peptide partners allows the relative abundance of peptides and proteins to be determined. Metabolic labeling strategies are of topical interest to groups seeking to reveal how altered growth conditions, drug administration, or environmental perturbations affect protein production and expression. Beyond relative quantitation of proteins and peptides, SILAC-like experiments have been used to probe post-translational modification production and turnover. However, the chief

considerations and drawbacks of these methods are 1) the small number of suitable amino acids that may be used for isotope incorporation; 2) poor separation of isotopic envelopes (causing errors in quantitative accuracy); and 3) the inability to incorporate isotopes to biological tissue and biofluid samples. In remedy, researchers may choose to tag proteins and peptides with isotopic labels after extraction and digestion.

Isotopic Tagging

Isotopic tagging, though similar in nature to metabolic incorporation, comes with a higher level of flexibility and customization.[31] Modern research settings have access to a broad array of stable isotopes, the most ubiquitous being ^{13}C, ^{15}N, ^{2}H, and ^{18}O. These isotopes enable researchers to synthesize their own chemical scaffold while varying the incorporation of these isotopes, creating an array of chemical tags with unique masses that may be functionalized and chemically bound to proteolytic peptides to provide them with a mass difference distinguishable via MS.[32, 33] In this way, the need for metabolic incorporation is completely removed and experimental peptides can be labeled after extraction and digestion. Similar to metabolic labeling, differences in MS1-level signal intensity between labeled species allow for determination of relative quantitation. Furthermore, isotopic labeling can be used as a means of absolute quantitation, whereby internal calibration curves are created and compared to experimental peptides. Overall, isotopic labeling presents highest utility in instances where the sample collection is relatively small because as sample number increases so does spectral complexity, which can create mass overlap between unique peptide species and produce erroneous quantitation estimates. These limitations in mind, the vast improvements in MS operational speeds, resolving power, and scanning depth begged the question as to whether more efficient chemical labels could leverage these instrumental

improvements and eliminate the spectral complexity found in complex isotopic tagging experiments.

Isobaric Labeling

As mass spectrometry technology continued to develop, it became obvious that the spectral complexity associated with high-throughput metabolic labeling and isotopic tagging experiments directly counteracted any instrumental improvements. As such, it became pertinent to find a method for absolute and relative quantitation that alleviate the ailments posed by multiplexed labeling methods while still retaining the facility in quantitative measurements. Remembering that isotopic tags may be constructed to provide a high number of labeling channels, each with a distinct mass difference of >1 Da, isobaric labels correct for this inherent mass difference using a balancing group.[34] When implemented, these isobaric labels display virtually indistinguishable masses at the MS^1 level, reducing the spectral complexity of high-throughput experiments. Upon selection of a labeled analyte, MS dissociation causes the isobaric tags to fragment and produce 'reporter ions' of unique mass. In this way labeled analytes may be selected and fragmented, providing identification and quantitative information in a single step. As a result, the reduced spectral complexity at the MS^1 level promotes greater profiling depth of complex samples and provides equivalent quantitative accuracy. The most popular examples of commercial isobaric labels are iTRAQ, Isotopic Tags for Relative and Absolute Quantitation[35] and TMT, Tandem Mass Tags.[36] However, the broad utility of isobaric labeling has garnered significant attention from the research community, resulting in numerous novel quantitative labeling strategies[37, 38] that promote quantitative accuracy at significantly reduced cost.

Label-free and Reaction Monitoring

Finally, in instances where sample labeling may not be preferred (i.e., precious samples, low-abundance molecules of interest, or instances where protein targets are known), label-free and reaction monitoring methods provide a suitable alternative.[39] Label-free quantitation serves to provide relative quantitation between samples by comparing area-under-curve for detected analytes. This method, though steadily improving with better instrumentation and software tools, is highly susceptible to changes in sample composition, can result in missing values, and is lower throughput than labeled methods. However, label-free quantitation does still represent a meaningful entry point in discovery-based quantitative proteomics, often providing deep sample profiling and elucidating targets for future analyses. In contrast, reaction monitoring workflows (e.g. multiple reaction monitoring, select reaction monitoring, etc.) may be considered one of the most accurate quantitative strategies, being most suited to targeted analyses and instances when internal standards are readily available. Though reaction monitoring strategies are often tailored to fit unique experimental conditions, all workflows bear resemblance to a basic strategy. First, serial dilutions of a purified or synthetic peptide standard are analyzed via targeted MS/MS. In these targeted analyses, the biomolecule(s) of interest are subjected to MS dissociation, with the various fragments observed and recorded. As each biomolecule will provide a unique transition/fragment, the prevalence of these transitions may be used as a proxy for overall biomolecule abundance. In this way, absolute and relative quantitation information can be determined without the need for chemical labeling while eliminating concerns over sample and spectral complexity. Often, it is preferential to incorporate an isotope-encoded standard,[40, 41] enabling rapid analyses and high quantitative accuracy. Given the variety of quantitative strategies,

it is of topical importance to evaluate their efficacy and provide understanding of quantitative accuracy.

Diagnostic Accuracy

As quantitative proteomics continues to mature, discussions over quantitative accuracy will continue to be a vanguard consideration. Recently, Dowle, *et al*.[42] provided an in-depth comparison of multiple quantitative strategies and should evaluated independently by interested parties. Within all quantitative strategies, the primary diagnostic for accuracy and utility are metrics built on specificity and sensitivity. Measures of specificity (i.e. proportion of correctly-identified trye positives) and sensitivity (i.e. proportion of correctly-identified true negatives), may be combined into a single metric. This receiver operating characteristic (ROC) is often viewed as a curve with high sensitivity and specificity representing a value close to 1. As demonstrated by Dowel, several commonly used quantitative strategies display high ROC values, providing detailed considerations of the method most appropriate for a range of experiments. This work may serve as a helpful guide when entering or expanding quantitative proteomics experiments.

Taken together, metabolic labeling, isotopic tagging, isobaric labeling, and label-free strategies provide a wealth of entry points into quantitative proteomics. This access in mind, the growing needs of the medical community combined with the ever-increasing access to mass spectrometry technology necessitate the utilization and expansion of investigational proteomics to aid in discovering and validating cancer-specific biomarkers.

Prostate Cancer

The second leading cancer type in men, prostate cancer is estimated to affect around 12 percent of all men during their lifetime and currently affects over 3 million men within the United States, with the majority of individuals diagnosed being at least 65 and older.[43] Androgen

deprivation, the first means of therapeutic intervention, can lead to the progression of castration-resistant prostate cancer (CRPC) in some men, a more aggressive stage of cancer resulting in poor prognosis and survival, with the majority of men developing metastases prior to or following diagnosis.[44] Further analyses of the literature have characterized these CRPC subtypes and demonstrated the growing emergence of CRPC phenotypes that have either low or negative AR expression for which there are few targeted therapeutics.[45] The growing heterogeneity in prostate cancer subtype underscores the urgency to elucidate and discover novel molecular mechanisms underlying pathogenesis for all subtypes. The use of mass spectrometry (MS)-based quantitative proteomics for prostate cancer research in recent years has been a driving force to exploit the factors underlying tumorigenesis and metastasis.

Cellular and Tissue Analyses

Investigations often profile quantitative differences in the proteome via patient-derived tissue samples, cellular models, or genetically engineered mouse models such as the transgenic adenocarcinoma of the mouse prostate (TRAMP) model. One such study by Zhang *et al.*[46] utilized a label-free approach to quantify differences in expression between the prostate glands of TRAMP versus wild-type littermates. Through generation and an in-depth analysis of the quantitative proteomics data, the authors were able to predict and validate the role of platelet-derived growth factor (PDGF)-B overexpression in increased proliferation, thereby highlighting the therapeutic potential of targeting PDGF signaling within prostate cancer. Other label-free approaches have utilized patient-derived tissue samples to profile global differences, including the work of Müller *et al.*[47] using formalin fixed, paraffin embedded tissue from radical prostatectomy, which focused on characterizing differences between nonmetastasizing tumors, metastasizing primary tumors and their distant nodal metastases. Although the analysis had only five biological replicates per tumor

type, significant differences in expression were measured that allowed for clear distinction of each and presented several potential proteins whose increased expression in metastatic tumors could be targeted in future therapeutic studies. However, a smaller sample set warrants further investigation into these proteins as potential targets with a larger sample cohort.

Methods that incorporate stable isotopes into the peptide backbone, such as SILAC, allow for direct comparison of identical peptides across sample types and is more robust to instrumentational variation compared to unlabeled approaches. Recently, SILAC was used to examine extracellular vesicles (EVs) and the impact that upregulated α(1,6)-fucosyltransferase (FUT8) expression had on biogenesis of these secreted biomolecules.[48] This was one of the first reports to map the systematic impact of an overexpressed glycosyltransferase on the EV proteome, specifically of a glycosyltransferase with known oncogenic activity.[49, 50] FUT8 overexpression showed a decrease in EVs produced compared to wild-type cells and further analysis of intact glycopeptides from LAPC4 EVs showed marked differences in glycosite occupancy between EV populations and revealed a shift in glycoform composition. Miao *et al.*[51] combined the SILAC approach with parallel-reaction-monitoring (PRM) methods to discern differential kinase expression in two bone metastasis-derived prostate lines, PC3 and PC3MLN4.[51] Of the kinases that were quantified and found to be differentially expressed, most notably different was mitogen-activated protein kinase kinase kinase kinase 4 (MAP4K4), a kinase previously observed to play a role in ovarian cancer.[52] One final example using the SILAC strategy by Sbrissa *et al.*[53] investigated the mechanisms of bone metastasis by determining CXCR4-interacting proteins through overexpression and knockdown of CXCR4 in PC3 cell lines. Proteomic analysis found one unexpected protein, phosphatidylinositol 4-kinase III α (PI4KIIIα), to be upregulated and it was found to localize with CXCR4 to lipid rafts and thus promote cancer cell invasion through

increasing phosphatidylinositol-4-phosphate production. The discovery of this novel interaction between chemokine receptor and PI4KIIIα and its regulation on tumor cell invasion requires more detailed experiments characterizing the specific molecular details regarding receptor-kinase communication.

Chemical or enzymatic isotopic labeling strategies allow for labeling of more than cell culture models to study prostate cancer. One approach by Lee *et al.*[54] used biotin ─ both as an isotopic label and for affinity purification ─ to systematically label cell-surface proteins that could serve to distinguish adenocarcinoma from neuroendocrine prostate cancer. From this proteogenomic investigation, they systematically validated two candidate antigens: FXYD domain containing ion transport regulator 3 (FXYD3) in prostate adenocarcinoma and CEA cell adhesion molecule 5 (CEACAM5) in neuroendocrine prostate cancer. While additional investigation into targeting these antigens is warranted, such a study demonstrates the utility of quantitative proteomics in discovering and validating new therapeutic targets for advanced prostate cancer.

Much of the quantitative research has shifted to the use of isobaric labeling strategies, which allow for increased multiplexing capabilities and decreased instrument variation through reduced overall runs. Zhou *et al.*[55] used 5-plex TMT labeling to perform a large-scale proteomic quantitation of core fucosylated glycopeptides after selective lectin affinity enrichment to differentiate non-aggressive and aggressive prostate cancer cell models (Figure 6.1). Over 20 fucosylated proteins were upregulated in the aggressive cell lines and were involved in processes such as cellular signaling, adhesion and extracellular communication. Identification of these fucosylated proteins and their upregulation in aggressive prostate cancer models establishes these proteins as potential targets for further examination into how their upregulation impacts the aggressive phenotype of the associated model. Another advantage to using TMT labeling is that

these tags can undergo synchronous precursor selection (SPS)-MS3 quantitation, which allow for more accurate quantitation. Zhou *et al*.[56] utilized a TMT-SPS-MS3 approach on patient-derived tissue samples with varying prostatic phenotypes to determine differential expression of protein complexes. Low-grade prostate cancer samples were found to have upregulation of complexes involved in RNA splicing and downregulation of those associated with cell adhesion, while high-grade prostate tissue samples had increased assembly of antiapoptotic complexes and a similar lower abundance of complexes involving cell adhesion. Such a comprehensive study of individual protein complexes may give way to determining what protein complexes are critical in distinguishing and diagnosing low- and high-grade cancers.

Comparable to TMT labeling, iTRAQ allows for both relative and absolute quantitation of labeled samples. Höti *et al.*[57] set out to examine the mechanisms underlying androgen resistance through a global proteomic approach using iTRAQ, labeling tryptic peptides from two prostate cancer cell models grown in triplicate. One main realization of the data was that androgen resistance cannot be treated with a single therapeutic, as the mechanisms driving resistance involve multiple independent pathways. While unfortunate, these findings did uncover some of the mechanisms driving resistance, including the PI3K/AKT signaling pathway, mitochondrial dysfunction of oxidative phosphorylation complexes and the multicatalytic 26S proteasome. Zhang *et al.*[58] used two sublines of PC-3M to distinguish unique characteristics of highly- and poorly-metastatic potential in prostate cancer. After validation, two proteins were found to potentially contribute to the higher metastatic potential, matrix metallopeptidase 1 (MMP1) and four and a half LIM domains 1 (FHL1). While FHL1 has been extensively studied in a variety of cancer types, the information collected here suggests a unique role of MMP1 for increasing metastatic potential in prostate cancer, presenting the opportunity for future inspection of both

MMP1 and other MMPs. Webber *et al*.[59] performed a stromal cell proteomics analysis to differentiate normal from tumor-reactive stromal phenotypes that drive disease progression. One compelling finding was the loss of aldehyde dehydrogenase (ALDH1A1) expression in altered versus normal stromal types, suggesting its potential role as a novel marker of disease-induced changes of the stromal environment. Additional investigations have turned to animal models, as prostate cancers grown *in vivo* reflect interactions that may otherwise be missed in cell culture models. The *Pten*-knockout mouse model[60] was recently examined by Zhang *et al.*[61] through the combined analysis of iTRAQ proteomics and microarray transcriptomics to identify associated molecular changes in mouse prostate carcinogenesis. Both transcriptomic and proteomic data found that immune and inflammation responses were greatly perturbed, in addition to mediations in central nodal activity through the Akt, NF-κB and P53 signaling pathways.

While tissue-based sampling allows for determination of mechanistic properties of the pathways contributing to tumorigenesis and metastasis, its highly invasive nature is discouraged unless necessary. Even if biopsies are obtained, these analyses are often limited by size constraints, as patient-derived tissues covering all stages of prostate cancer progression can be difficult to obtain in large numbers. Mouse models afford the opportunity to mimic tumor progression and metastasis *in vivo*, but there are still controversies surrounding prostate-based mouse models due to distinct anatomical differences.[62] Cell culture models avoid the translational constraints that other model organisms are bound to, but often omit stromal-epithelial interactions during cell growth, a process that has a great impact on tumor invasiveness and metastatic potential. Additionally, current cell-based models for prostate cancer often either only reflect advanced prostate cancer or require the use of multiple cell lines to cover multiple progression stages, introducing variability that complicates genetic-based analyses. Recent advances in cell-based

prostate models have been made that address some of the pitfalls outlined here,[63] so future quantitative studies should be selective in the models they choose when profiling.

Biofluid Analyses

There is a push to develop biomarker strategies involving the collection of biofluids, a less invasive and more cost-effective method of sample collection. Biofluids ─ such as blood, tissue-based fluid, saliva, or urine ─ allow for easier monitoring of patient outcomes, as disease progression and treatment responsiveness can be evaluated with frequent patient sampling. Such biofluid-based monitoring strategies are critical in prostate cancer patients, as a portion of men diagnosed with prostate cancer have tumors that are indolent. One study by Davalieva *et al.*[64] comparatively profiled urine samples using a label-free strategy from patients with prostate cancer, benign prostatic hyperplasia, bladder cancer and renal cancer to determine selective biomarkers for earlier diagnosis of prostate cancer. Of the most promising urinary biomarkers identified by the authors, nine had not yet been associated with prostate cancer, indicating their potential as novel biomarkers and necessitating further research into their associated pathways. Soekmadji *et al.*[65] profiled secretome differences of unlabeled, CD9-positive EVs from cell culture models treated with the hormone dihydrotestosterone (DHT). Their combined analyses determined that DHT treatment increases CD9-positive EV secretion and alters the content of secreted EVs, and in agreement with previous literature highlighting the potential of CD9 EVs as a biomarker for prostate cancer.

Reaction monitoring-based strategies are one label-free approach that are typically used after initial discovery for validation and accurate quantitation of biomarkers. Targeted analysis of urinary EVs was completed by Sequeiros *et al.*[66] using SRM to quantify 64 protein candidate biomarkers for prostate cancer. A two-protein combination (ADSV and TGM4) distinguished

patients with benign tissue from those with cancer, and a five-protein panel differentiated high- from low-grade prostate cancer (CD63, GLPK5, SPHM, PSA and PAPP), highlighting the advantages of targeted proteomics as a diagnostic tool in the clinic. Kim *et al.*[67] investigated expressed prostatic secretion samples using SRM-based quantitation to determine molecular signatures for extracapsular prostate cancer. From a pool of over 200 potential candidates, these researchers narrowed the candidate list to include 34 peptides representative of 27 unique proteins with promising results as biomarkers. Karasota *et al*.[68] evaluated the analytical performance of multiple SRM- and PRM-based strategies to quantitate kallikrein related peptidase 4 (KLK4) in a variety of biofluid samples. Secreted KLK4 was demonstrated to be present in seminal plasma, and for the first time was investigated as a potential biomarker in both seminal plasma and blood. Taken together, the label-free, targeted proteomics methods used for analysis of biofluids offer a reliable tool for biomarker validation and should thus be considered as useful tools for clinical development.

Fujita *et al*.[69] combined two strategies, initially using iTRAQ for urine samples to profile EVs from patients with a high Gleason score.[70] After quantifying 3528 proteins, candidate biomarkers were selected for further quantitation and validation using SRM/MRM. Fatty acid binding protein 5 (FABP5) was highlighted as the most promising biomarker from urinary EVs for the detection and diagnosis of high Gleason score prostate cancer, but further studies would be necessary for confirmation. Yan *et al.*[71] performed an iTRAQ-based analysis on the serum of prostate cancer patients with or without bone metastasis to find potential biomarkers indicative of these metastases. Of the 32 differentially expressed proteins identified, three ─ CD59, haptoglobin and tetranectin ─ were selected and validated to be related to prostate cancer bone metastasis, confirming their utility as serum biomarkers. Larkin *et al.*[72] implemented iTRAQ to enhance their

proteomic profiling of high-quality serum samples for biomarker discovery. After identification and validation using ELISA, two biomarkers, SAA and TSR1, showed promising results when used in combination with KLK3. However, these results were obtained in a small sample cohort, so further studies with a larger, more diverse sample set are necessary before serious consideration of these proteins as biomarkers.

The use of quantitative proteomic strategies on patient-derived biofluid samples show promise in the discovery and validation of new biomarkers. Specifically, the KLK family of proteins has been shown in the mentioned literature to have potential in many biofluids and may improve diagnostic accuracy further when combined with others. On the other hand, serum biomarkers in prostate cancer deserve a level of scrutiny as demonstrated by prostate-specific antigen (PSA), a currently approved biomarker whose elevation in serum is also associated with benign prostatic hyperplasia (BPH), resulting in high false positive rates.[73] Noting this, prostate cancer biomarkers should be rigorously tested against patients with BPH and other prostatic diseases to ensure accuracy. Such rigorous tests involving larger sample sets can be achieved using the quantitative strategies described above, indicating their potential to advance the knowledge within the field at a rapid pace.

Pancreatic Cancer

The seventh leading cause of cancer-related deaths in the world,[74] pancreatic cancer has rightfully garnered significant attention from clinical research communities. In-depth proteomic analyses have illuminated the highly dynamic nature of post-translational modifications,[75-77] while providing novel insights toward treatment monitoring and severity stratification. The promising results of these profiling experiments in hand, great success has come in the effort to employ

quantitative strategies to illuminate dysregulated protein expression, identify treatment pathways, and validate potential biomarkers.

Tissue Analyses

The prevalence of pancreatic cancer across the world's population has necessitated in-depth proteomic analyses of cancerous tissue and model systems. Model cell lines have enabled researchers to identify pertinent biomolecules specific to pancreatic cancer without the need for invasive, repetitive tissue resection. Naturally, the study of cell lines lends itself to the use of SILAC to perform quantitative investigations. Recently, Liu *et al*.[78] performed secretomic analyses of pancreatic cancer cells (PC-1), revealing 161 proteins with altered expression, including 55 proteins not previously reported. As well, they note a combination panel for cadherin 3 (CDH3), plasmogen activator, urokinase (PLAU), and lunatic fringe (LFNG) proteins that may be useful for improving cancer patient prognoses. Beyond this, Marchand *et al*.[79] employed a three-channel SILAC approach to reveal association of transcription factor EB (TFEB) with nuclear proteins upon inhibition of glycogen synthase kinase-3 (GSK3) and mammalian target of rapamycin (mTOR). Moving beyond SILAC experiments, Shi *et al*.[80] used isotopic dimethyl labeling to examine paracrine communication between pancreatic cancer cells and pancreatic stellate cells (PSCs). This experiment provided the knowledge that leukemia inhibitory factor (LIF) is a key paracrine factor from activated PSCs acting on cancer cells. Employing a novel approach, Roberts *et al*. [81] developed a cysteine-reactive fragment-based ligand library to coordinate novel small molecules that impair pancreatic cancer pathogenicity with druggable hotspots for potential cancer therapy. While numerous SILAC and isotopic tagging workflows exist outside the time frame of this review, the relatively small number of recent applications indicates an area of potential focus for researchers examining pancreatic cancer.

Isobaric labeling, however, has seen significant use in the study of pancreatic cancer. Zhang and colleagues[82] have provided a meaningful guide for those seeking to perform isobaric labeling experiments using the commercial TMT[36] offering from ThermoScientific. Beyond this, Perera *et al*.[83] employed TMT labeling to study pancreatic cancer cell metabolism, revealing the MiT/TFE proteins – MITF, TFE3 and TFEB – are decoupled from regulatory mechanisms, increasing expression levels of lysosomal catabolic function essential for pancreatic ductal adenocarcinoma (PDA) growth. As an alternative to TMT, An *et al*.[84] employed iTRAQ in the analysis of serum exosomes from chemotherapy patients (Figure 6.2). Of note, this study indicates patient-derived exosomes play a significant role in cancer metastasis. Furthermore, Li *et al*.[85] demonstrated monumental success in broad protein quantitation while analyzing Peripheral Blood Mononuclear Cells (PBMCs). This study, which employed iTRAQ labeling and 2D-LC-MS quantified 3,357 proteins, with 114 being distinguished as dysregulated in the cancer group. These examples of isobaric labeling indicate the broad utility for high throughput analyses of complex pancreatic cancer samples. However, a chief limitation of TMT and iTRAQ is cost, placing their use out-of-reach for many budding research groups. In remedy, Li and colleagues[34] developed Dimethyl Leucine (DiLeu) that provides greater multiplexity[86-88] than commercial options at a fraction of the cost. DiLeu is available as an isotopic,[89] isobaric,[34, 86, 87] and mass-defect[90] chemical tag and has even been modified to provide absolute quantitation.[89, 91] The mass-defect offering, mdDiLeu, has been successfully applied for simultaneous multiomic analysis of pancreatic cancer cells,[92] providing uncompromised access to high throughput small molecule and protein quantitation.

Label-free analyses, too, have seen routine utilization in pancreatic cancer investigations. Wang *et al*.[93] introduced the novel IonStar pipeline for accurate MS1-level protein quantitation.

This preliminary example quantified >4,000 proteins from 40 biological samples and identified 541 proteins dysregulated groups treated with birinapant and paclitaxel. Later Zhu *et al*.[94] applied the IonStar pipeline to elucidate the relations among relevant signaling pathways during gemcitabine and birinapant treatment. These applications highlight the utility of quantitative proteomics to evaluate treatment efficacy. In a similar vein, Singh *et al*.[95] presented a large-scale, label-free proteomics study to uncover the mechanism by which sanguinarine suppresses cancer proliferation. While quantifying >3,100 proteins, 37 biomolecules were identified as differentially expressed, highlighting the pleotropic effects of sanguinarine. Finally, Zhou *et al*.[96] employed parallel reaction monitoring (PRM) to identify 165 potential biomarkers in pancreatic cancer. During validation, brain acid soluble protein 1 (BASP1) was identified as a novel target for pancreatic cancer therapy and is shown to interact with Wilms tumor protein (WT1).

Biofluid Analyses

Considering the real-world application of investigational proteomics analyses, a topical concern is the need for invasive patient sampling. This in mind, researchers have long sought to identify cancer-specific analytes from biofluids, which may be sampled repeatedly at lower physical and monetary cost to patients. Though metabolic and isotopic labeling are not well represented in pancreatic cancer research in recent years, Jhaveri *et al*.[97] used a novel serum antibody–based SILAC immunoprecipitation approach, denoted as SASI, to identify specific targets expressed in cancer patients post-vaccine therapy. More popular, however, are applications utilizing isobaric labeling. Sogawa *et al*.[98] employed TMT labeling to ascertain that complement component 4 binding protein alpha (C4BPA) and polymeric immunoglobulin receptor (PIGR) expression was significantly higher in preoperative patients than postoperative. Naba *et al*.[99] identified unique expression levels in 35 proteins as pancreatic cancer islets progressed from

hyperplastic to angiogenic to insulomas. Yu *et al.*[100] employed iTRAQ to quantify 4,517 proteins in the exosomes of Panc02 and Panc02-H7 cells, notably revealing cancer-derived exosomes promote tumor metastasis. Lin *et al.*[101] and Liu *et al.*[102] further implemented iTRAQ for quantitative evaluations of pancreatic cancer patient serum. An important overlap of these two studies was the identification that apolipoprotein A-1 (APOA1) shows distinct expression in pancreatic cancer patients. Considering this trend was shared between patients expressing carbohydrate antigen (CA) 19-9 and those who are CA19-9-negative, APOA1 presents an area of significant interest moving forward.

Similar to the studies presented in pancreatic cancer tissue analyses, label-free quantitation has been routinely employed in quantification of biofluid proteins. Through this quantitative strategy, Ohmine *et al.*[103] successfully validated deoxycytidine kinase (dCK) as a good predictor of progression-free survival and an effective biomarker of gemcitabine sensitivity. Yoneyama *et al.*[104] identified insulin-like growth factor-binding proteins insulin-like growth factor binding protein 2 (IGFBP2) and IGFBP3 as compensatory biomarkers of pancreatic cancer in instances when CA19-9 screening is inconclusive. Park *et al.*[105] performed a large-scale validation of biomarkers, finding that APOA-IV, APOCIII, IGFBP2, and tissue inhibitor of metalloproteinase 1 (TIMP) were significantly altered in pancreatic cancer. Of note, a panel including CA19-9, APOA-IV, and TIMP1 showed improved performance in distinguishing early pancreatic cancer from pancreatitis. Do *et al.*[106] identified 18 biomarker candidates associated with malignancy in intraductal papillary mucinous neoplasms (IPMNs). Finally, Nigjeh *et al.*[107] developed an optimized data-independent acquisition (DIA) workflow to identify and quantify >14,000 peptides from ~2,300 plasma proteins (Figure 6.3).

As seen by the numerous examples of pancreatic cancer tissue and biofluid investigation, quantitative proteomics provides a facile entry point into the field of biomarker identification and validation (Table 2). Considering the agreement across several studies that proteins such as APOA1, APOA4, IGFBP and CA19-9 serve as rigorous biomarkers in pancreatic cancer, future studies should investigate the utility of high throughput label-free, PRM, or MRM screening of these biomolecules. Meaningful evaluation of MS-based protein assays in blind studies may demonstrate potential to accurately identify and diagnose pancreatic cancer at scale. Development of these workflows and associated technology will be vital to understanding the risk factors associated with disease onset and progression, as well as the success of current and novel treatment strategies.

Breast Cancer

The high rate of incidence associated with breast cancer, as well as targeted focus drawn from successful advocacy and research fundraising, have shed significant light on the mechanisms of breast cancer. Though this dedicated focus has reduced patient mortality and cancer rates in high income countries, developing nations display the opposite trend.[108] Beyond this, breast cancer is of continual interest to the medical community due to the high rate of recurrence and metastasis.[109, 110] For these reasons, many have turned to quantitative proteomics to aid in stratifying cancer subtypes and identifying potential biomarkers.

Tissue and Biofluid Analyses

Within the timeframe of this review, the majority of quantitative proteomic investigations have been centered on tissue analyses, often employing model cell lines or resected tumor tissue to determine protein expression. Though few applications have employed metabolic labeling for quantitative analyses, Tyanova *et al.*[111] presented a robust investigation that merged quantitative

mass spectrometry with traditional RNA- and DNA-based sequencing strategies. Analyzing 40 tumors that were either estrogen receptor positive, Her2 positive, or triple negative, the authors identified an average of >7,000 proteins on average, spanning 8 orders of magnitude in protein intensity. Within this study, they combined their quantitative results with microarray analyses and machine learning classification to identify potential subtype-specific therapies.

More popular than SILAC-like experiments, isobaric labeling has been extensively employed in breast cancer investigations. Suman *et al.*[112] employed iTRAQ to identify proteins associated with breast cancer subtypes. Notably, this study indicated fibronectin (FN1), alpha-2-macroglobulin (A2M), complement component-4-binding protein-alpha (C4BPA) and complement factor-B (CFB) were critical to subtype differentiation in both plasma and tissue samples. Calderon-Gonzalez *et al.*[113] further employed this technology to identify 306 differentially expressed proteins in breast cancer cell lines. As well, their study indicates large proline-rich protein (BAG6), ATP-dependent RNA helicase (DDX39), annexin A8 (ANXA8) and cytochrome c oxidase subunit 4 (COX4) may serve as useful biomarkers. Gajbhiye *et al.*[114] provided a novel DIA-iTRAQ strategy to uncover proteomic divergence in HER2-enriched cancer cell lines, which allowed for the creation and testing of a 21 protein panel to discriminate cancer and healthy controls. Turning to TMT labeling, Going *et al.*[115] and Clark *et al.*[116] utilized this strategy, illuminating the pathways of action of methoxyclcone in triple negative breast cancer (Figure 6.4) and classifying exosomal cargo proteins, respectively. As a cost-effective alternative to these iTRAQ and TMT labeling strategies, DiLeu tagging approach has also successfully been employed in identifying strategies for inhibiting cancer cell proliferation. Within this work, Liu, *et al.*[117] revealed that dynamic methylation of pyruvate kinase M2 (PKM2) directly affect the metabolic activity of cancer cells and promotes cell propagation, migration and metastasis. This

study, along with those detailed above, serve to indicate the importance of high-throughput quantitative cancer proteomics, outlining potential targets for future treatment strategies.

A significant entry into quantitative breast cancer proteomics was provided by Johansson *et al.*[118]. This study provided in-depth quantitation of 45 breast cancer tumors, spanning each of the 5 PAM50-based molecular classifications. Upon quantitation of 9,995 proteins, the authors used these proteome profiles to interpret multiple layers of systems measurements. While each of these studies offered unique insight into uncovering and validating potential biomarkers and investigative strategies, a chief concern among many is the long-term reproducibility of quantitative measurements. Using iTRAQ to quantify proteins from human-in-mouse xenograft tissue, Zhou *et al.*[119] demonstrated that the large majority of quantitative measurements hold consistent over time, but also raised some topical concerns. First, they observed higher variability in quantitation of hydrophilic peptides compared to those of average peptide character, likely due to poor retention of these peptides on column. Second, as researchers have their choice of dissociation methods, this study reveals stepped collision energy offers higher reproducibility between unique measurements. Finally, whereas most commercial software implements a form of scoring to determine the quality of a peptide spectral match (PSM), this study goes further and reveals that a stricter scoring mechanism improves reliability of time-course measurements. This study provides an excellent framework and series of considerations for individuals seeking to begin or improve quantitative mass spectrometry investigations.

Label-free analyses have also been routinely implemented for high throughput biomarker discovery and screening. Among these, Ntai *et al.*[120] compared the quantitation efficiency in bottom-up and top-down analyses of tumor xenografts. Tveitras *et al.*[121] performed comparative analyses of pre-metastatic and metastatic triple negative breast cancer xenograft tissue, uncovering

significant changes in expression of haptoglobin, fibrinogen, and thrombospondin-4 and transferrin receptor protein 1 between groups. Wang *et al.*[122] employed a DIA-select reaction monitoring (SRM) approach to reveal distinct proteomic and N-glycoproteomic divergence between normal, precancerous, and cancerous tissues. Gamez-Pozo *et al.*[123] integrated label-free MS quantitation with RT-qPCR to definitively distinguish estrogen receptor positive and triple negative cancer subtypes. Nie *et al.* [124] identified 98 differentially expressed proteins when comparing pure breast cancer stem cells and mature luminal cells. Finally, Warmoes *et al.*[125] elucidated 215 proteins that are significantly enriched in BRCA1-deficient secretome. This study highlights the potential of mass spectrometry to provide sensitive identification of biomarkers in instances when traditional ELISA screening may fall short.

These examples of successful quantitative proteomic analyses in breast cancer applications highlight the flexibility and facility of creating novel workflows to answer an array of biological problems. Knowing there have been a significant number of proteomic measurements made prior to the period in review, these examples of biomarker discovery and validation highlight how rigorous protein MS-based screening assays for the confident identification and stratification of breast cancer may be within reach (Table 3). Assays of this kind, devoid of the need for invasive and repetitive tissue sampling, provide a meaningful conduit towards aiding communities that have limited access to dedicated cancer screening centers and provide direct targets for potential novel therapies.

Ovarian Cancer

Although it has an estimated incidence rate of approximately 2% for 2020, ovarian cancer is the deadliest reproductive cancer in women, with an estimated mortality rate of 5% in women diagnosed with any cancer type and 64% for women diagnosed with ovarian cancer.[126] Much

emphasis has been placed on the continued research into mechanisms driving ovarian cancer, as late-stage diagnosis of advanced cancer contributes to the high mortality of ovarian cancer. Continued efforts have focused on the identification of critical mechanisms driving disease progression across ovarian cancer subtypes. Quantitative proteomic strategies have continued to increase the depth of knowledge surrounding ovarian cancer and its various subtypes to improve earlier identification strategies and highlight new therapeutic targets.

Cellular and Tissue Analyses

Because the majority of diagnosed ovarian cancer cases have already progressed to a more advanced stage, much quantitative research delves into tissue and cellular proteomic profiling to isolate and exploit dysregulated proteins. While only applicable to cellular-based models, SILAC has been implemented in ovarian cell lines and led to the discovery of critical modulators in ovarian disease progression. Musrap *et al.*[127] cultured the ovarian line OV-90 in adherent and non-adherent conditions using SILAC to compare the impacts of cancer aggregate formation on cellular proteomics. After quantifying 1533 proteins in total, they compared expression with other aggregate-forming lines and saw upregulation of CLCA1, which appeared to affect cancer cell aggregation after further siRNA experimentation. Grassi *et al.*[128] utilized SILAC to quantify epidermal growth factor (EGF)-induced epithelial-mesenchymal transition (EMT) to identify specific mechanisms of this process that may be dysregulated for metastatic purposes. 206 proteins were found to be differentially expressed, some of which included proteins associated with the G1 and G2 checkpoints of the cell cycle, indicating the role of EGF-induced EMT in cell cycle control mechanisms. Another investigation by Ji *et al.*[129] utilized the metabolic strategy to perform an integrated proteomic and N-glycoproteomic analysis of ovarian cancer lines that were either doxorubicin-sensitive or -resistant. They quantified 5509 protein groups and identified 1525 high-

confidence N-glycosites corresponding to 740 glycoproteins. Quantifying the protein abundance allowed these researchers to examine glycoprotein abundances and alterations, which provides unique information into the role of N-glycosylation in drug resistance.

Applicable to more than just cell culture-based models, isobaric labeling is commonly employed for quantitative experiments applied to ovarian cancer sample sets. Zhang *et al.*[130] used iTRAQ labeling to integrate quantitative proteomics with the transcriptomic profile of ovarian high-grade serous cancer (HGSC) patient biospecimens. Over 3500 proteins were quantified and used in tandem with genomic results to reveal a strong association between specific histone acetylation events and the homologous recombination deficient phenotype seen in patient samples. Hiramatsu *et al.*[131] comparatively profiled HGSC and endometrial carcinoma samples using iTRAQ-based quantitation. Comprehensive analysis revealed 356 quantifiable proteins and identified mitochondrial inner membrane protease subunit 2 (IMP2) and minichromosome maintenance complex component 2 (MCM2) to be modulators of rapid HGSC growth, illustrating the need to examine these two proteins in further ovarian cancer studies.

Alternatively, many other analyses have used the TMT-based isobaric strategy rather than iTRAQ labeling. Recently, Hu *et al.*[132] used an integrated proteomic and glycoproteomic approach with TMT-labeled peptides in their analysis of ovarian HGSC versus non-tumor tissues. These authors combined global proteomics, solid-phase extraction of glycosite-containing peptides (SPEG) and glycan identification via intact glycopeptide analysis to provide a comprehensive view into N-glycoproteomics within ovarian cancer. Their integrated approach yielded promising results, identifying tumor-specific glycosylation and revealing glycosylation enzymes that were correlated with altered glycosylation status. Yoshimura *et al.*[133] treated neighboring peritoneal mesothelial cells with a microRNA shown to be elevated in the serum of ovarian cancer patients

to determine its role in cancer invasion and metastasis. The TMT-based proteomics analysis exhibited increased expression of fibronectin and vitronectin, enhancing the ability of the cancer cells to invade the surrounding environment. A straightforward, quantitative comparison of TMT-labeled normal versus cancerous ovarian tissue was performed by Qu *et al.*[134] to find differentially expressed proteins that hold promise in elucidating disease progression. Initial analyses found 498 differentially expressed proteins and highlighted chloride intracellular channel protein 1 (CLIC1), which was examined further and ultimately determined to promote tumorigenesis, making it an attractive therapeutic target. Proteogenomic and phosphoproteomic analysis was performed by McDermott *et al.*[135] to characterize mechanisms driving ovarian HGSC functions down to the post-translational level. Global proteomic analysis led to the identification of 10,706 proteins and combined results described a role of histone acetylation as a marker for homologous recombination deficiency, confirming an association earlier proposed by Zhang *et al.*[130] Phosphoproteomics data provided understanding into proliferation-induced replication stress and the impact it has on chromosomal instability in HGSC, implying that mitotic and cyclin-dependent kinases could serve as therapeutic targets after future experimental validation.

Label-free quantitation is frequently employed for ovarian cancer analyses, as the global overview it provides of the proteome allows researchers to identify multiple pathways for further targeted analyses. Chuffa *et al.*[136] used this approach to determine the influence of melatonin treatment on an *in vivo* model of ovarian cancer. Comparative proteomics analyses showed that downregulation of processes involved in cancer signaling was promoted, underlining molecular targets for therapeutic intervention while indicating the feasibility of melatonin supplementation for ovarian cancer patients. Another comparative analysis by Júnior *et al.*[137] explored the effects of P-MAPA, IL-12 or a combination immunotherapy of the two on the SKOV-3 ovarian cancer

cell line. After confirming 532 proteins were identified across all groups, it was noted that combination therapy of P-MAPA and IL-12 was most efficient at regulating proteins involved in metabolic processes that may render cancer cells more vulnerable, suggesting that the use of the two therapies concomitantly is a plausible treatment strategy. Coscia *et al.*[138] used a quantitative, label-free approach in tandem with other quantitative strategies to probe the proteomes of platinum-resistant and -sensitive ovarian HGSC patient-derived tissues (Figure 6.5). Multi-level quantitative analyses revealed cancer/testis antigen family 45 (CT45) as a prognostic factor through mediation of chemosensitivity, thereby exposing it as an immunotherapy target.

The quantitative tissue analyses outlined here provide multiple protein targets for the development of new targeted therapies. The role of a defective DNA damage response in ovarian cancer is well established, so the multiple studies highlighting histone acetylation and its role in homologous recombination deficiency is supported by current literature and should be examined in therapeutic development.[139] Additional analyses that examine post-translational modifications simultaneously with proteomics should also be explored, as these studies may highlight other processes outside the DNA damage response that promote cancer progression. The experiments above outline the utility that quantitative proteomic approaches hold in advancing the knowledge of the ovarian cancer field.

Biofluid Analyses

Quantitative analyses that inspect biofluids of ovarian cancer samples provide valuable information about potential biomarkers that allow for earlier detection and diagnosis, a current area of the ovarian cancer field that is in dire need of new research breakthroughs. Isobaric labeling of ovarian biofluids allow scientists to relatively quantify biomarkers that may otherwise go undetected or are lost during depletion of abundant serum proteins such as albumin. Zhang *et al.*[140]

profiled exosomes derived from patient plasma using the TMT tagging strategy. When the 225 proteins identified across all samples were quantitatively compared, proteins associated with the coagulation cascade were found to be differentially expressed and may therefore be promising diagnostic factors for ovarian cancer. Zhang *et al.*[141] went on to further profile circulating exosomes of late-stage cancer patients using iTRAQ. After validation, they determined that apolipoprotein E (ApoE) multiplexed with epithelial cell adhesion molecule (EpCAM), plasminogen (PLG), serpin family C member 1 (serpinC1) and complement component 1q (C1q)were able to accurately diagnose ovarian cancer. It was also noted that activation of coagulation cascades was increased in the ovarian cancer cohort due to increased Factor X levels, demonstrating the impact that tumor-derived extracellular vesicles may have on other biological processes. Swiatly *et al.*[142] examined iTRAQ-labeled serum proteins from healthy control, benign ovarian tumor and ovarian cancer patients. Five proteins were found to be differentially expressed within the ovarian cancer group, and three of these coupled to current biomarkers CA125 and HE4 improved diagnostic discrimination between benign and malignant ovarian tumors. Russell *et al.*[143] used iTRAQ to screen preclinical serum samples for detection of early stage biomarkers and initially identified 90 differentially expressed proteins in ovarian cancer cases. A second targeted analysis of 20 selected candidates revealed Vitamin K-dependent protein Z (VKDP), an anticoagulant not previously associated with ovarian cancer, as either a novel independent early detection biomarker or concomitantly with CA125 to increase differential diagnostic capabilities.

Although label-free analyses suffer from longer instrument times and potential run-to-run variability, they provide the greatest profiling depth of the multiple quantitative strategies and are vital to finding new ovarian biomarkers. Barnabas *et al.*[144] performed deep proteome profiling of 187 uterine liquid biopsy-derived microvesicles to identify early detection biomarkers. Machine

learning algorithms identified a 9-protein signature that correctly identified all Stage I lesions, demonstrating the strength of the panel for future use in early diagnosis. Zhang *et al.*[145] studied the plasma proteins to isolate biomarkers related to chemoresistance of postoperative reoccurrence. These experiments found a total of six dysregulated proteins that could serve as predictive biomarkers for chemoresistance in ovarian cancer patients. The combination of plasma proteomics and metabolomics was utilized by Ahn *et al.*[146] to discover new molecular signatures of ovarian HGSC. Differential expression of 34 metabolites and 197 proteins was found, with three proteins (phosphopantothenoylcysteine synthetase (PPCS), peripheral myelin protein 2 (PMP2) and tubulin beta class I (TUBB)) and two metabolites (L-carnitine and PC-O) related to the carnitine system established as potential markers of cancer plasticity. Hüttenhain *et al.*[147] created a biomarker development strategy for large-scale SRM studies in ovarian cancer plasma samples. After developing a 5-protein signature for ovarian cancer and testing it against the current ELISA-based standard for biomarker tests, it was found that the SRM-based method had sensitivity measurements that exceeded the current ELISA standard, validating its potential for clinical development and use. Rauniyar *et al.*[148] also used a more targeted approach, combining data-independent acquisition methods with PRM to improve identification of ovarian cancer serum biomarkers. They demonstrated that ApoA-IV is a more reliable biomarker than previously determined by immunological assays in addition to the identification of C-reactive protein, transferrin and transthyretin as other available ovarian serum markers. Overall, this study validated the use of quantitative mass spectrometry as a more sensitive and reliable method of quantitation compared to immunological-based procedures.

While the quantitative research mentioned here has progressed ovarian cancer research, continuing studies are still necessary to delve deeper into specific mechanisms of novel markers

identified. During the review process, many of the identified studies had a tissue-based proteomics approach and minimal studies focused on biofluid samples (Table 4). More studies focusing on the use of biofluids in ovarian cancer research are critical in the development of novel biomarkers for earlier detection and treatment, and the lack of literature compared to tissue-based studies highlights a current area for further quantitative experimentation in ovarian cancer. In particular, studying the microvesicular proteome for the discovery of novel biomarkers has shown great potential both here and in other quantitative applications. Profiling of extracellular vesicles may prove to be a vital key in the prevention of late-stage diagnosis and increasing the overall survival rate of patients diagnosed with ovarian cancer.

Conclusions and Future Directions

The various quantitative strategies outlined here have demonstrated the growing utility of MS-based quantitation methods in cancer diagnosis and research. Quantitative analyses of prostate cancer have been frequently performed within the field due to the growing emergence of resistance to first-line treatments and false diagnoses associated with elevated PSA levels. Multiple members of the KLK family were identified as potential biomarkers and further strengthened when detected in combination with other proteins, suggesting their potential for clinical diagnosis. Targeted validation experiments in a cohort spanning all grades of prostate cancer as well as BPH should be performed before serious consideration is given to using these proteins as biomarkers. Pancreatic studies have been relatively successful in determining sets of robust biomarkers for diagnosis and patient stratification. APOA1, APOA4, IGFBP, and CA19-9 have been indicated in numerous peer-reviewed studies as critical components for pancreatic cancer screening. Future analyses should focus on high throughput reaction monitoring to rapidly screen for these biomarkers. Breast cancer research has seen limited quantitative proteomics studies in recent years, so future efforts of those investigating new biomarkers and determining mechanisms of carcinogenesis should consider quantitative proteomics strategies in their analyses. The small number of studies highlighted here contribute potential protein panels useful for breast cancer screening, but more large-scale studies that confirm the utility of these proteins as biomarkers are necessary. Ovarian research has seen large numbers of tissue- and cellular-based quantitation, but there is a lack of biofluid-based experiments. While tissue-based studies provide large amounts of information that guide knowledge of disease mechanisms, biofluid studies offer important insights that could facilitate the identification and development of protein biomarkers for clinical diagnosis.

Due to the lack of biomarkers that detect ovarian cancer at an earlier stage, studies covering biofluids are critical and present an understudied area within the ovarian field.

A common drawback of the quantitative studies addressed is that these investigations only determine up- or downregulation of differentially expressed proteins at a single point in time. Time-course evaluations monitoring the differential expression and dynamic changes of these proteins over time could prove to be more useful, as these studies would explain how expression levels change within a single patient over time. In combination with the expression levels across varying disease severity, there is a potential to determine a critical expression level for each stage of cancer progression that determines not only if the patient has cancer, but also the severity of that cancer relative to biomarker concentration levels. Rapid analyses of cancer samples via targeted monitoring strategies offer benefits over current immuno-based assays such as ELISA, demonstrating the advantage of MS-based quantitation for detection and prolonged patient monitoring. Another strategy for improving cancer diagnosis is the integration of additional analyses, such as transcriptomics, metabolomics, or analysis of post-translational modifications and associated crosstalk. Many of the studies outlined here utilized a combined approach to their investigations, leading to the successful identification of a specific protein or process with altered expression in both datasets. These integrated approaches help scientists identify mechanisms driving cancer metastasis and treatment resistance, thus demonstrating their growing utility in future studies. Additional efforts should be made towards understanding communication within the tumor microenvironment, as much remains to be known about the interactions that help a tumor transition from localized to metastatic ability. Finally, studies focusing on single-cell analyses should also be considered for future experiments, as the cellular diversity and heterogeneity

provided from such examinations may prove to be critical in understanding specific mechanisms that allow pathogenesis to advance.

Taken together, this review highlights the utility of various quantitative strategies, their associated limitations, and some directions for novel applications in cancer diagnosis and cancer research. As instrumental capabilities continue to grow, it will become necessary for researchers to develop and validate higher throughput labeling strategies that accommodate deeper proteomic profiling. Regardless of the application, quantitative proteomics represents a premier avenue towards cancer biomarker detection, identification, and validation. Continued efforts in the coming years will certainly be centered on the utility of mass spectrometry-based biomarker detection in clinical settings and the development of point-of-care biomolecule screening.

Acknowledgements

Support for this research was provided in part by the NIH grants U01CA231081, R01 DK071801, RF1 AG052324, and P41GM108538. LL acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Tables

Table 6.1 Summarized selection of prostate cancer biomarkers.

Proposed Biomarker	Source	Findings
Platelet-derived growth factor (PDGF)-B [46]	Prostatic tissue	Overexpressed with increased cancer proliferation
α(1,6)-fucosyltransferase (FUT8) [48-50]	LAPC4 and LNCaP cells	Increased FUT8 expression corresponded with decreased extracellular vesicle production
Mitogen-activated protein kinase kinase kinase kinase 4 (MAP4K4) [51, 52]	PC3 and PC3MLN4 cells	Differential expression in metastasis-derived cell lines
Phosphatidylinositol 4-kinase III α (PI4KIIIα) [53]	PC3 cells	Upregulated in PC3 cell lines; promotes cancer cell invasion
FXYD domain containing ion transport regulator 3 (FXYD3) [54]	PrAd, NEPC cell lines	Biomarker specific to prostate adenocarcinoma
CEA cell adhesion molecule 5 (CEACAM5) [54]	PrAd, NEPC cell lines	Biomarker specific to neuroendocrine cancer
Four and a half LIM domains 1 (FHL1), Matrix metallopeptidase 1 (MMP1) [58]	PC-3M sublines	Promote higher metastatic potential
Aldehyde dehydrogenase (ALDH1A1) [59]	Stromal tissue	Loss of expression in altered stromal cell types
Actin-depolymerizing factor (ADSV), transglutaminase 4 (TGM4) [66]	Urine	Differentiates benign and cancerous tissue
CD63 Molecule (CD63), glycerol kinase 5 (GLPK5), SPHM sulfohydrolase (SPHM), Prostate-specific antigen (PSA) and pappalysin 1 (PAPP) [66]	Urine	Distinguishes high- and low- grade cancer
Kallikrein related peptidase 4 (KLK4) [68]	Seminal Fluid	Biomarker available in seminal fluid
Fatty acid binding protein 5 (FABP5)	Urine	Utility in detecting, diagnosing high gleason score prostate cancer
CD59 molecule (CD59), haptoglobin and tetranectin [71]	Serum	Expression correlated to bone metastasis

Table 6.2 Summarized selection of pancreatic cancer biomarkers.

Proposed Biomarker	Source	Findings
Cadherin 3 (CDH3), plasmogen activator, urokinase (PLAU), lunatic fringe (LFNG) [78]	PC-1 cell secretome	Potential for improving cancer patient prognoses
Transcription factor EB (TFEB) [79]	HEK293, PANC1, MIA PaCa-2 cells	Association with nuclear protein upon inhibition of GSK3
Leukemia inhibitory factor (LIF) [80]	Pancreatic stellate cells	Denoted as major paracrine factor
Melanocyte inducing transcription factor (MITF), transcription factor binding to IGHM enhancer 3 (TFE3) and transcription factor EB (TFEB) [83]	Tissue, PDA cells	Decoupled from regulatory mechanisms, promote catabolic function
Brain acid soluble protein 1 (BASP1) [96]	Tissue	Novel cancer therapy target
Complement component 4 binding protein alpha (C4BPA), polymeric immunoglobulin receptor (PIGR) [98]	Serum	Higher expression in preoperative patients than postoperative
Apolipoprotein A-1 (APOA1) [101, 102]	Serum	Distinct expression in both CA 19-9 positive and CA 19-9-defficient patients
Deoxycytidine kinase (dCK) [103]	PK9, CFPac-1, PK1, SUIT-2, and AsPC-1 cells	Predictor of progression-free survival, biomarker of gemcitabine sensitivity
Insulin like growth factor binding protein 2 (IGFBP2) and IGFBP3 [104]	Plasma	Compensatory biomarkers when CA 19-9 screening is inconclusive
Insulin-like growth factor binding protein 2 (IGFBP2) tissue inhibitor of metalloproteinase 1 (TIMP1), Apolipoprotein A IV (APOA-IV), Apolipoprotein CIII APOCIII [105]	Blood	Protein panel highly effective in early detection of pancreatic cancer

Table 6.3 Summarized selection of breast cancer biomarkers.

Proposed Biomarker	Source	Findings
Fibronectin (FN1), alpha-2-macroglobulin (A2M), complement component-4-binding protein-alpha (C4BPA) and complement factor-B (CFB) [112]	Tumor tissue	Critical for subtype differentiation
Large proline-rich protein (BAG6), ATP-dependent RNA helicase (DDX39), annexin A8 (ANXA8) and cytochrome c oxidase subunit 4 (COX4) [113]	MCF7 and T47D, MDA-MB-231, and SK-BR-3 cells	Putative biomarkers for breast cancer
Methylated pyruvate kinase M2 (PKM2) [117]	MCF7, MDA-MB-231, HEK293T cells	Promotes cell propagation, migration and metastasis
Haptoglobin, fibrinogen, and thrombospondin-4 and transferrin receptor protein 1 [121]	Pre-/metastatic xenograft tissue	Reveal N-glycoproteomic divergence between normal, precancerous, and cancerous tissues

Table 6.4 Summarized selection of ovarian cancer biomarkers.

Proposed Biomarker	Source	Findings
Calcium-activated chloride channel 1 (CLCA1) [127]	OV-90 cells	Affects cancer cell regulation
Insulin-like growth factor 2 (IMP2) and minichromosome maintenance complex component 2 (MCM2) [131]	HGSC and endometrial tissue	Modulators of rapid high-grade serous cancer growth
Fibronectin and vitronectin [133]	Peritoneal mesothelial cells	Increased expression promotes cancer cell invasion
Chloride intracellular channel protein 1 (CLIC1) [134]	Tissue	Determined to promote tumorigenesis
Histone acetylation [130, 135]	Tumor tissue	Marker for homologous recombination deficiency
Phospholinoleate–palmitoleate anhydride (P-MAPA), interleukin 12 (IL-12) [137]	SKOV-3	Combination immunotherapy is a plausible treatment strategy
Cancer/testis antigen family 45 (CT45) [138]	Tissue	Found to be a prognostic factor
Apolipoprotein E (ApoE), epithelial cell adhesion molecule (EpCAM), plasminogen (PLG), serpin family C member 1 (serpinC1) and complement component 1q (C1q) [141]	Circulating exosomes	Diagnostic markers of ovarian cancer
Vitamin K-dependent protein Z (VKDP)	Preclinical serum	Novel, early detection biomarker
Phosphopantothenoylcysteine synthetase (PPCS), peripheral myelin protein 2 (PMP2) and tubulin beta class I (TUBB) [146]	Blood, plasma	Potential markers of cancer plasticity
Apolipoprotein IV (ApoA-IV) [148]	Serum	More reliable biomarker compared to benchmark proteins

Figures

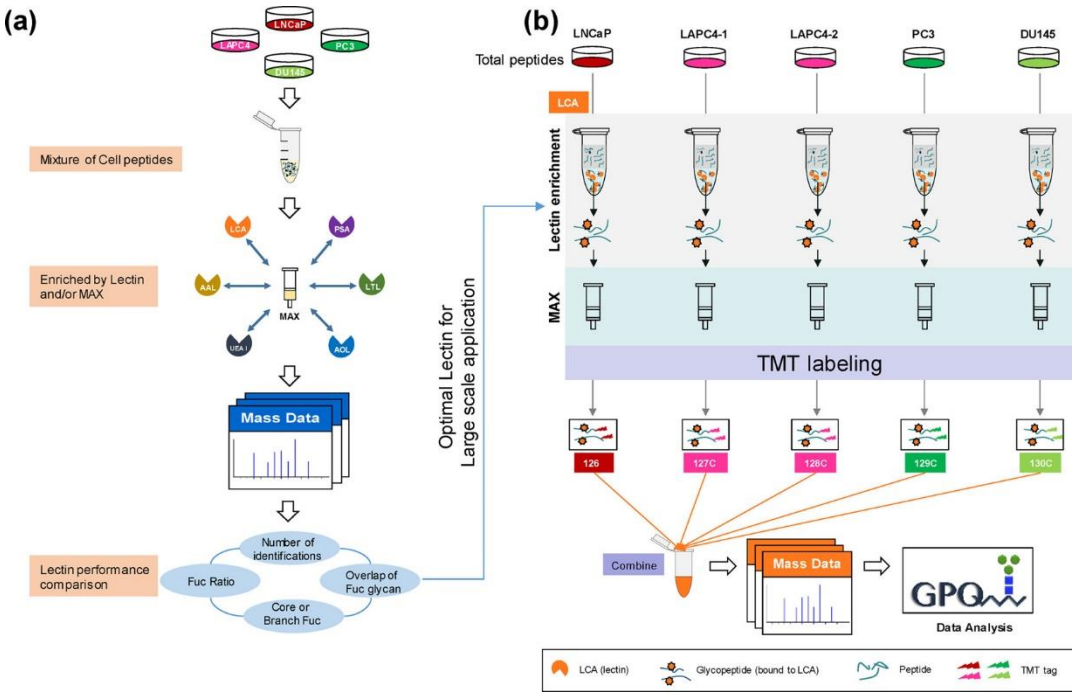


Figure 6.1 Example workflow detailing quantitative investigation of prostate cancer protein modifications.

Complete workflow utilized by Zhou *et al.*[55] detailing the quantitative approach to investigate site-specific fucosylation and glycoproteins associated with aggressive prostate cancer phenotypes. The optimized enrichment strategy used to identify glycopeptides contributing to prostate cancer aggressiveness shows promise for application in a variety of cancer glycosylation studies but should also be applied to other prostate cancer models to determine its utility across sample types. *Reprinted with permission.*

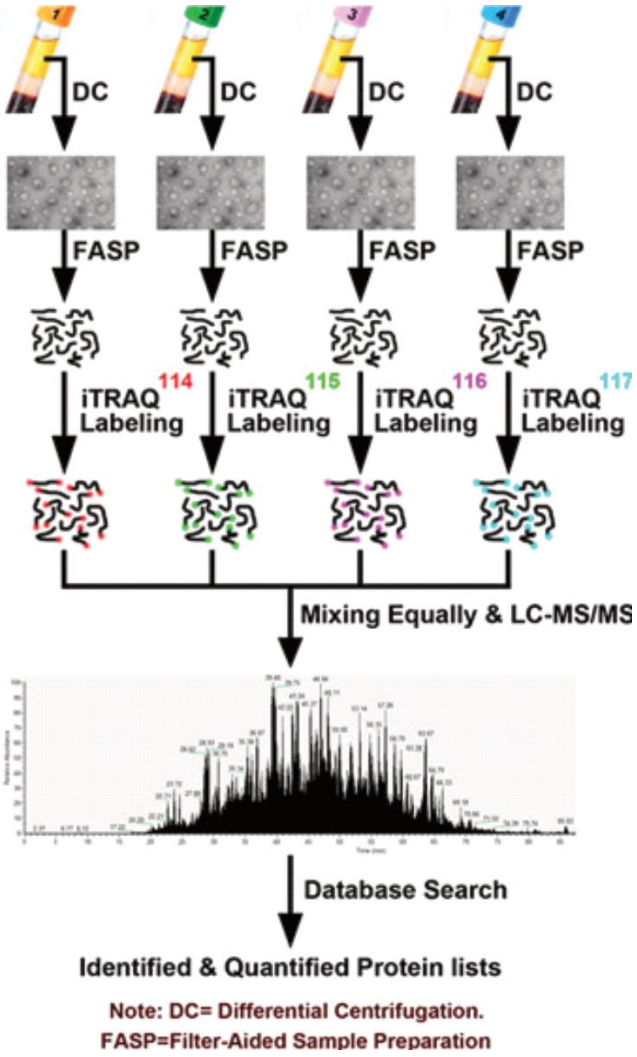


Figure 6.2 Quantitative analysis of chemotherapy patient exosomes via iTRAQ.

Workflow described by An, *et al*.[84] for the quantitative analysis of chemotherapy patient exosomes through iTRAQ labeling and quantitative mass spectrometry. This example of a facile isobaric labeling proteomics experiment provides deep proteomic profiling of multiple complex samples with lower spectral complexity than isotopic labeling methods. *Reprinted with permission.*

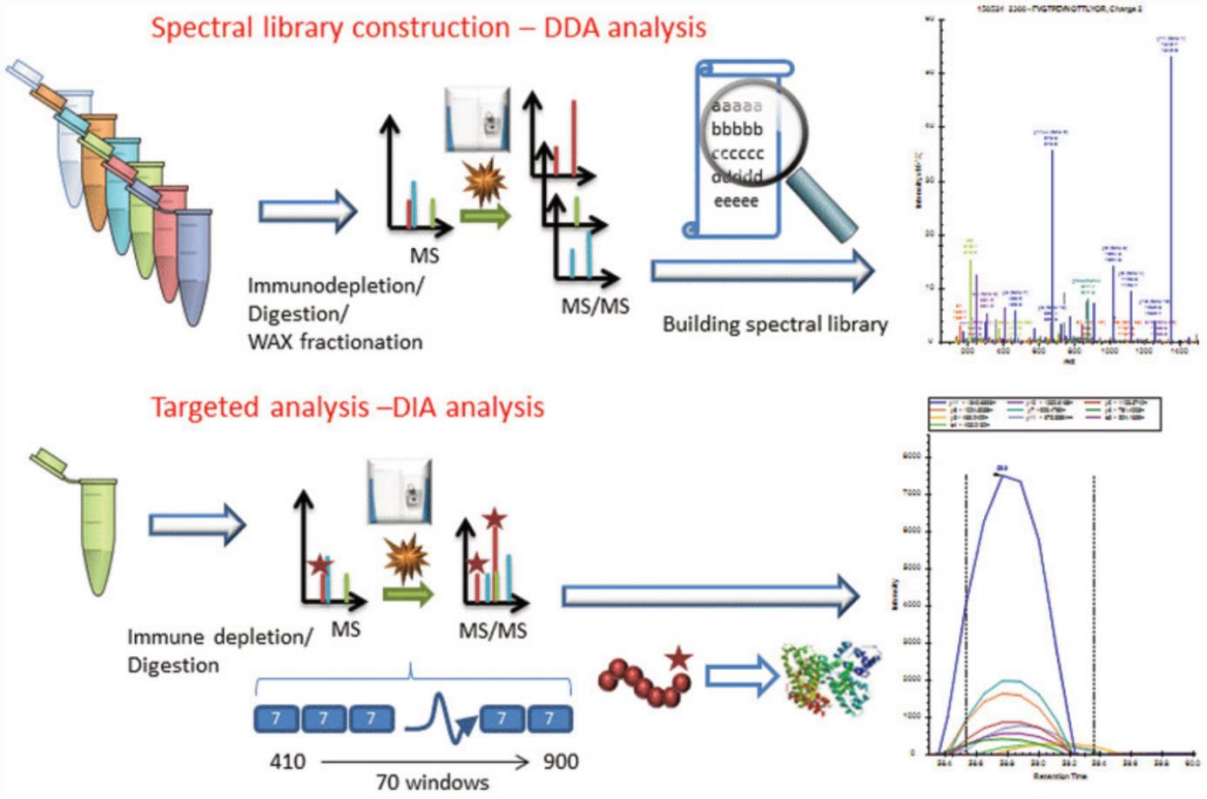


Figure 6.3 Isobaric labeling in the context of high-throughput quantitative proteomics.

Workflow implemented by Nigjeh, *et al.*[107] Quantitative workflows utilizing isobaric labels present the greatest propensity for deep proteome profiling. However, these workflows are limited by their instrument acquisition speed and cycle time required to select and fragment top precursors. For this reason, implementation of DIA strategies presents the ability to sequence a greater number of peptides in the same amount of time. Though the data processing methods are significantly more involved, DIA workflows are sure to be of critical importance to proteome profiling in the coming years. *Reprinted with permission.*

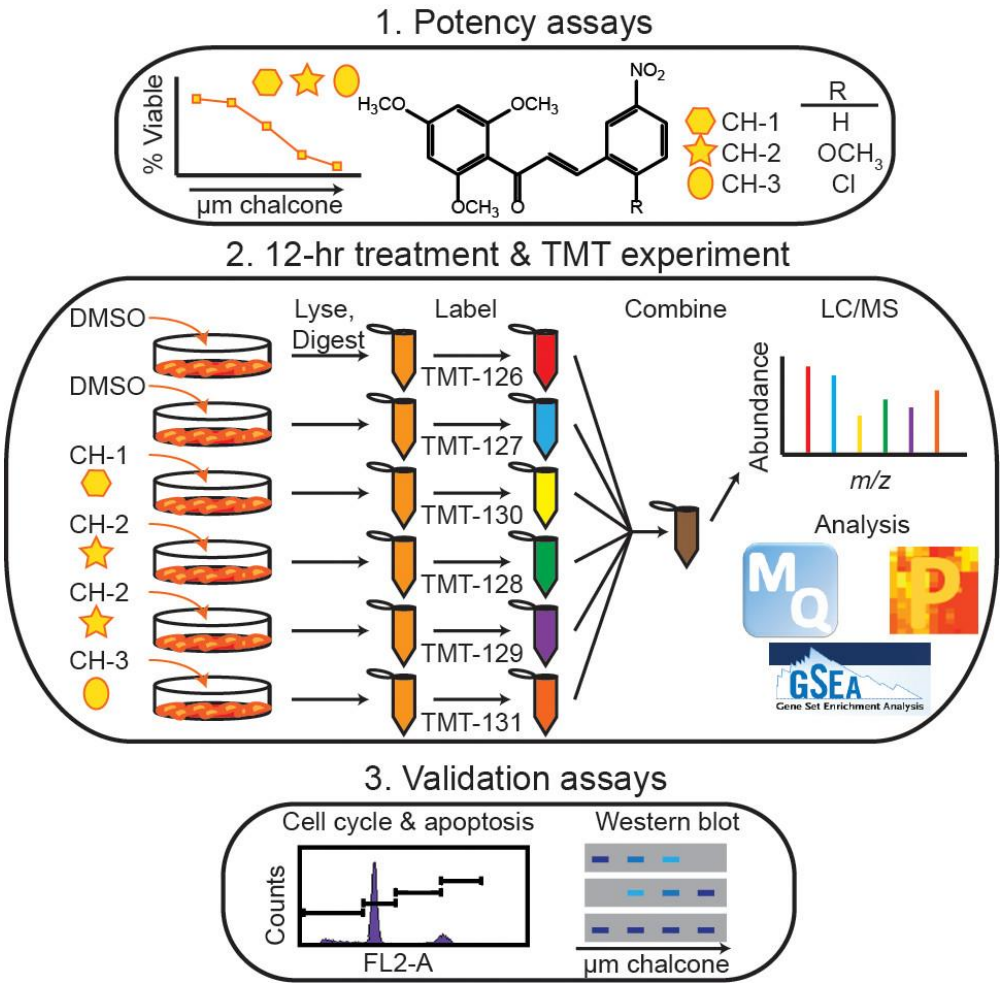


Figure 6.4 Example of quantitative analyses to determine treatment efficacy.

Representative workflow established by Going, *et al.*[115] As quantitative proteomics is critical for discovering and validated biomolecules of interest during periods of disease and treatment, this workflow represents an example of how treatment strategies may be controlled and systematically evaluated. While SILAC methods would be useful in situations where cell growth is monitored, isotopic labeling methods may be considered inherently lower throughput due to the increases in spectral complexity they may provide. *Reprinted with permission.*

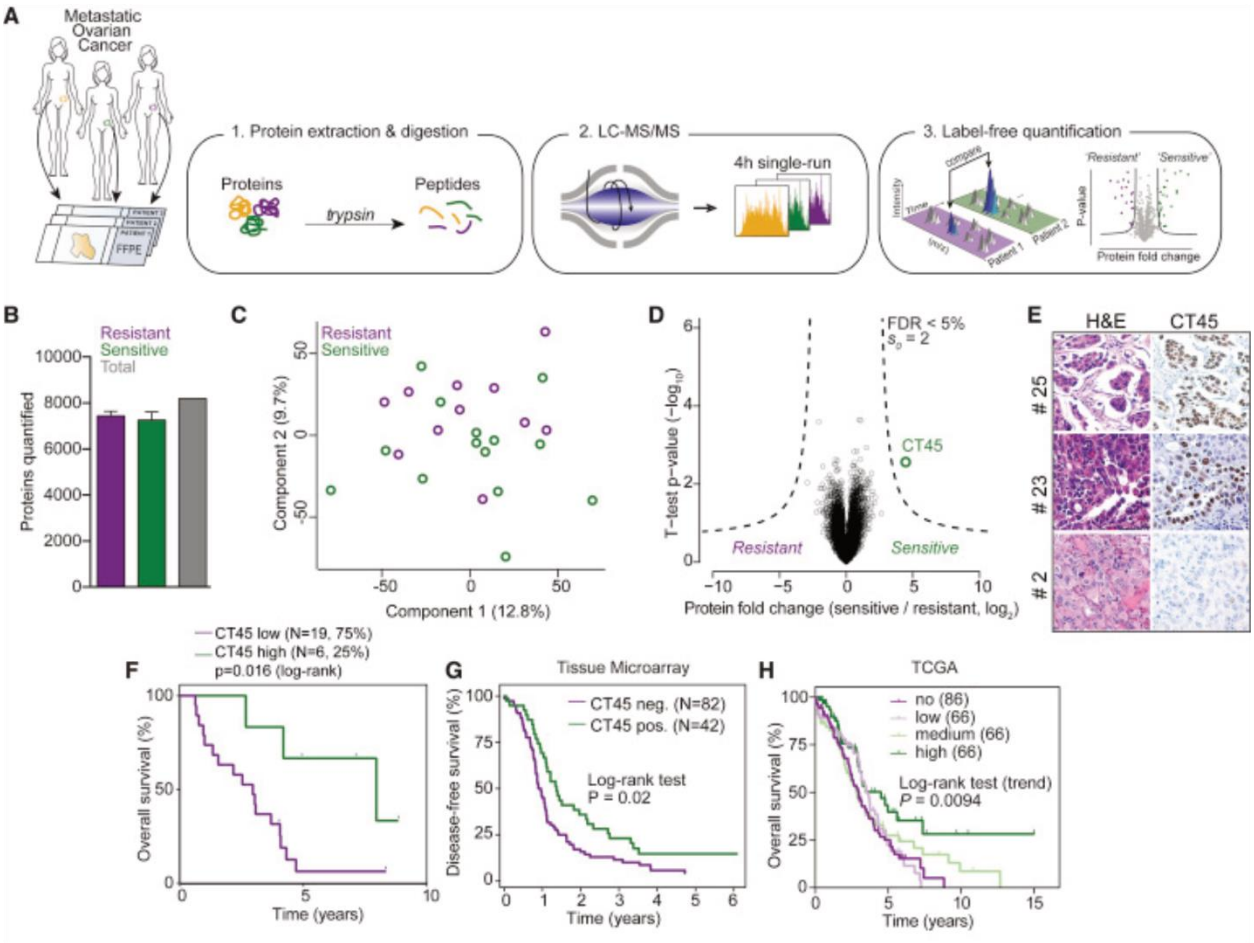


Figure 6.5 Quantitative proteomics workflow to stratify chemotherapeutic response.

Analysis by Coscia *et al.*[138] to determine proteomic differences in ovarian cancer tissue samples either resistant or sensitive to platinum-based chemotherapeutics. This strategy identified CT45 as a chemosensitivity modulator and demonstrates the ability of quantitative methods to identify factors that play a role in therapeutic resistance. *Reprinted with permission.*

References

1. Aksenov, A. A.; da Silva, R.; Knight, R.; Lopes, N. P.; Dorrestein, P. C., Global chemical analysis of biology by mass spectrometry. *Nature Reviews Chemistry* **2017,** *1* (7), 0054.

2. Ren, J.-L.; Zhang, A.-H.; Kong, L.; Wang, X.-J., Advances in mass spectrometry-based metabolomics for investigation of metabolites. *RSC Advances* **2018,** *8* (40), 22335-22350.

3. Paglia, G.; Astarita, G., Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. *Nature Protocols* **2017,** *12* (4), 797-813.

4. DeLaney, K.; Buchberger, A. R.; Atkinson, L.; Gründer, S.; Mousley, A.; Li, L., New techniques, applications and perspectives in neuropeptide research. *J Exp Biol* **2018,** *221* (Pt 3), jeb151167.

5. DeLaney, K.; Li, L., Data Independent Acquisition Mass Spectrometry Method for Improved Neuropeptidomic Coverage in Crustacean Neural Tissue Extracts. *Anal Chem* **2019,** *91* (8), 5150-5158.

6. Li, G.; Delafield, D. G.; Li, L., Improved structural elucidation of peptide isomers and their receptors using advanced ion mobility-mass spectrometry. *TrAC Trends in Analytical Chemistry* **2019**.

7. Livnat, I.; Tai, H.-C.; Jansson, E. T.; Bai, L.; Romanova, E. V.; Chen, T.-t.; Yu, K.; Chen, S.-a.; Zhang, Y.; Wang, Z.-y.; Liu, D.-d.; Weiss, K. R.; Jing, J.; Sweedler, J. V., A d-Amino Acid-Containing Neuropeptide Discovery Funnel. *Analytical Chemistry* **2016,** *88* (23), 11868-11876.

8. Ryan, D. J.; Spraggins, J. M.; Caprioli, R. M., Protein identification strategies in MALDI imaging mass spectrometry: a brief review. *Current Opinion in Chemical Biology* **2019,** *48*, 64-72.

9. Srzentić, K.; Fornelli, L.; Tsybin, Y. O.; Loo, J. A.; Seckler, H.; Agar, J. N.; Anderson, L. C.; Bai, D. L.; Beck, A.; Brodbelt, J. S.; van der Burgt, Y. E. M.; Chamot-Rooke, J.; Chatterjee, S.; Chen, Y.; Clarke, D. J.; Danis, P. O.; Diedrich, J. K.; D'Ippolito, R. A.; Dupré, M.; Gasilova, N.; Ge, Y.; Goo, Y. A.; Goodlett, D. R.; Greer, S.; Haselmann, K. F.; He, L.; Hendrickson, C. L.; Hinkle, J. D.; Holt, M. V.; Hughes, S.; Hunt, D. F.; Kelleher, N. L.; Kozhinov, A. N.; Lin, Z.; Malosse, C.; Marshall, A. G.; Menin, L.; Millikin, R. J.; Nagornov, K. O.; Nicolardi, S.; Paša-Tolić, L.; Pengelley, S.; Quebbemann, N. R.; Resemann, A.; Sandoval, W.; Sarin, R.; Schmitt, N. D.; Shabanowitz, J.; Shaw, J. B.; Shortreed, M. R.; Smith, L. M.; Sobott, F.; Suckau, D.; Toby, T.; Weisbrod, C. R.; Wildburger, N. C.; Yates, J. R.; Yoon, S. H.; Young, N. L.; Zhou, M., Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (9), 1783-1802.

10. Brown, K. A.; Melby, J. A.; Roberts, D. S.; Ge, Y., Top-down Proteomics: Challenges, Innovations, and Applications in Basic and Clinical Research. *Expert Review of Proteomics* **2020**, null-null.

11. Takemori, A.; Butcher, D. S.; Harman, V. M.; Brownridge, P.; Shima, K.; Higo, D.; Ishizaki, J.; Hasegawa, H.; Suzuki, J.; Yamashita, M.; Loo, J. A.; Loo, R. R. O.; Beynon, R. J.; Anderson, L. C.; Takemori, N., PEPPI-MS: Polyacrylamide-Gel-Based Prefractionation for Analysis of Intact Proteoforms and Protein Complexes by Mass Spectrometry. *Journal of Proteome Research* **2020,** *19* (9), 3779-3791.

12. Griffiths, R. L.; Sisley, E. K.; Lopez-Clavijo, A. F.; Simmonds, A. L.; Styles, I. B.; Cooper, H. J., Native mass spectrometry imaging of intact proteins and protein complexes in thin tissue sections. *International Journal of Mass Spectrometry* **2019,** *437*, 23-29.

13. VanAernum, Z. L.; Gilbert, J. D.; Belov, M. E.; Makarov, A. A.; Horning, S. R.; Wysocki, V. H., Surface-Induced Dissociation of Noncovalent Protein Complexes in an Extended Mass Range Orbitrap Mass Spectrometer. *Analytical Chemistry* **2019,** *91* (5), 3611-3618.

14. Mann, M., The ever expanding scope of electrospray mass spectrometry—a 30 year journey. *Nature Communications* **2019,** *10* (1), 3744.

15. Karas, M.; Bachmann, D.; Hillenkamp, F., Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Analytical Chemistry* **1985,** *57* (14), 2935-2939.

16. Buchberger, A. R.; DeLaney, K.; Johnson, J.; Li, L., Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. *Analytical Chemistry* **2018,** *90* (1), 240-265.

17. Soltwisch, J.; Heijs, B.; Koch, A.; Vens-Cappell, S.; Höhndorf, J.; Dreisewerd, K., MALDI-2 on a Trapped Ion Mobility Quadrupole Time-of-Flight Instrument for Rapid Mass Spectrometry Imaging and Ion Mobility Separation of Complex Lipid Profiles. *Analytical Chemistry* **2020,** *92* (13), 8697-8703.

18. Arevalo, R., Jr.; Ni, Z.; Danell, R. M., Mass spectrometry and planetary exploration: A brief review and future projection. *J Mass Spectrom* **2020,** *55* (1), e4454-e4454.

19. Maitre, P.; Scuderi, D.; Corinti, D.; Chiavarino, B.; Crestoni, M. E.; Fornarini, S., Applications of Infrared Multiple Photon Dissociation (IRMPD) to the Detection of Posttranslational Modifications. *Chemical Reviews* **2020,** *120* (7), 3261-3295.

20. Stiving, A. Q.; VanAernum, Z. L.; Busch, F.; Harvey, S. R.; Sarni, S. H.; Wysocki, V. H., Surface-Induced Dissociation: An Effective Method for Characterization of Protein Quaternary Structure. *Analytical Chemistry* **2019,** *91* (1), 190-209.

21. Brodbelt, J. S.; Morrison, L. J.; Santos, I., Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chemical Reviews* **2020,** *120* (7), 3328-3380.

22. Riley, N. M.; Westphall, M. S.; Hebert, A. S.; Coon, J. J., Implementation of Activated Ion Electron Transfer Dissociation on a Quadrupole-Orbitrap-Linear Ion Trap Hybrid Mass Spectrometer. *Analytical Chemistry* **2017,** *89* (12), 6358-6366.

23. Brandão, P. F.; Duarte, A. C.; Duarte, R. M. B. O., Comprehensive multidimensional liquid chromatography for advancing environmental and natural products research. *TrAC Trends in Analytical Chemistry* **2019,** *116*, 186-197.

24. Ranjbar, L.; Foley, J. P.; Breadmore, M. C., Multidimensional liquid-phase separations combining both chromatography and electrophoresis – A review. *Analytica Chimica Acta* **2017,** *950*, 7-31.

25. Lv, W.; Shi, X.; Wang, S.; Xu, G., Multidimensional liquid chromatography-mass spectrometry for metabolomic and lipidomic analyses. *TrAC Trends in Analytical Chemistry* **2019,** *120*, 115302.

26. Michalski, A.; Cox, J.; Mann, M., More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC−MS/MS. *Journal of Proteome Research* **2011,** *10* (4), 1785-1793.

27. Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics* **2002,** *1* (5), 376-386.

28. Ong, S.-E.; Mann, M., Mass spectrometry–based proteomics turns quantitative. *Nature Chemical Biology* **2005,** *1* (5), 252-262.

29. Meselson, M.; Stahl, F. W., The replication of DNA in Escherichia coli. *Proceedings of the National Academy of Sciences* **1958,** *44* (7), 671-682.

30. Zhu, H.; Hunter, T. C.; Pan, S.; Yau, P. M.; Bradbury, E. M.; Chen, X., Residue-specific Mass Signatures for the Efficient Detection of Protein Modifications by Mass Spectrometry. *Analytical Chemistry* **2002,** *74* (7), 1687-1694.

31. Lau, H.-T.; Suh, H. W.; Golkowski, M.; Ong, S.-E., Comparing SILAC- and stable isotope dimethyl-labeling approaches for quantitative proteomics. *Journal of proteome research* **2014,** *13* (9), 4164-4174.

32. Kovanich, D.; Cappadona, S.; Raijmakers, R.; Mohammed, S.; Scholten, A.; Heck, A. J. R., Applications of stable isotope dimethyl labeling in quantitative proteomics. *Analytical and Bioanalytical Chemistry* **2012,** *404* (4), 991-1009.

33. Wu, Y.; Wang, F.; Liu, Z.; Qin, H.; Song, C.; Huang, J.; Bian, Y.; Wei, X.; Dong, J.; Zou, H., Five-plex isotope dimethyl labeling for quantitative proteomics. *Chemical Communications* **2014,** *50* (14), 1708-1710.

34. Xiang, F.; Ye, H.; Chen, R.; Fu, Q.; Li, L., N,N-Dimethyl Leucines as Novel Isobaric Tandem Mass Tags for Quantitative Proteomics and Peptidomics. *Analytical Chemistry* **2010,** *82* (7), 2817-2825.

35. Wiese, S.; Reidegeld, K. A.; Meyer, H. E.; Warscheid, B., Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* **2007,** *7* (3), 340-50.

36. Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C., Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Analytical Chemistry* **2003,** *75* (8), 1895-1904.

37. Delafield, D. G.; Li, L., Recent Advances in Analytical Approaches for Glycan and Glycopeptide Quantitation. *Molecular & Cellular Proteomics* **2020**, mcp.R120.002095.

38. Rauniyar, N.; Yates, J. R., 3rd, Isobaric labeling-based relative quantification in shotgun proteomics. *Journal of proteome research* **2014,** *13* (12), 5293-5309.

39. Neilson, K. A.; Ali, N. A.; Muralidharan, S.; Mirzaei, M.; Mariani, M.; Assadourian, G.; Lee, A.; van Sluyter, S. C.; Haynes, P. A., Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* **2011,** *11* (4), 535-53.

40. Kuzyk, M. A.; Smith, D.; Yang, J.; Cross, T. J.; Jackson, A. M.; Hardie, D. B.; Anderson, N. L.; Borchers, C. H., Multiple Reaction Monitoring-based, Multiplexed, Absolute Quantitation of 45 Proteins in Human Plasma*. *Molecular & Cellular Proteomics* **2009,** *8* (8), 1860-1877.

41. Addona, T. A.; Abbatiello, S. E.; Schilling, B.; Skates, S. J.; Mani, D. R.; Bunk, D. M.; Spiegelman, C. H.; Zimmerman, L. J.; Ham, A.-J. L.; Keshishian, H.; Hall, S. C.; Allen, S.; Blackman, R. K.; Borchers, C. H.; Buck, C.; Cardasis, H. L.; Cusack, M. P.; Dodder, N. G.; Gibson, B. W.; Held, J. M.; Hiltke, T.; Jackson, A.; Johansen, E. B.; Kinsinger, C. R.; Li, J.; Mesri, M.; Neubert, T. A.; Niles, R. K.; Pulsipher, T. C.; Ransohoff, D.; Rodriguez, H.; Rudnick, P. A.; Smith, D.; Tabb, D. L.; Tegeler, T. J.; Variyath, A. M.; Vega-Montoto, L. J.; Wahlander, Å.; Waldemarson, S.; Wang, M.; Whiteaker, J. R.; Zhao, L.; Anderson, N. L.; Fisher, S. J.; Liebler, D. C.; Paulovich, A. G.; Regnier, F. E.; Tempst, P.; Carr, S. A., Multi-site assessment of the precision and reproducibility of multiple reaction monitoring–based measurements of proteins in plasma. *Nature Biotechnology* **2009,** *27* (7), 633-641.

42. Dowle, A. A.; Wilson, J.; Thomas, J. R., Comparing the Diagnostic Classification Accuracy of iTRAQ, Peak-Area, Spectral-Counting, and emPAI Methods for Relative Quantification in Expression Proteomics. *Journal of Proteome Research* **2016,** *15* (10), 3550-3562.

43. SEER*Explorer: An interactive website for SEER cancer statistics . Surveillance Research Program. https://seer.cancer.gov/explorer/. (accessed Dec. 4th).

44. Kirby, M.; Hirst, C.; Crawford, E. D., Characterising the castration-resistant prostate cancer population: a systematic review. *International Journal of Clinical Practice* **2011,** *65* (11), 1180-1192.

45. Vellky, J. E.; Ricke, W. A., Development and prevalence of castration-resistant prostate cancer subtypes. *Neoplasia* **2020,** *22* (11), 566-575.

46. Zhang, Y.; Wang, D.; Li, M.; Wei, X.; Liu, S.; Zhao, M.; Liu, C.; Wang, X.; Jiang, X.; Li, X.; Zhang, S.; Bergquist, J.; Wang, B.; Yang, C.; Mi, J.; Tian, G., Quantitative Proteomics of TRAMP Mice Combined with Bioinformatics Analysis Reveals That PDGF-B Regulatory Network Plays a Key Role in Prostate Cancer Progression. *Journal of Proteome Research* **2018,** *17* (7), 2401-2411.

47. Müller, A.-K.; Föll, M.; Heckelmann, B.; Kiefer, S.; Werner, M.; Schilling, O.; Biniossek, M. L.; Jilg, C. A.; Drendel, V., Proteomic Characterization of Prostate Cancer to Distinguish Nonmetastasizing and Metastasizing Primary Tumors and Lymph Node Metastases. *Neoplasia* **2018,** *20* (2), 140-151.

48. Clark, D. J.; Schnaubelt, M.; Hoti, N.; Hu, Y.; Zhou, Y.; Gooya, M.; Zhang, H., Impact of Increased FUT8 Expression on the Extracellular Vesicle Proteome in Prostate Cancer Cells. *Journal of Proteome Research* **2020,** *19* (6), 2195-2205.

49. Wang, X.; Chen, J.; Li, Q. K.; Peskoe, S. B.; Zhang, B.; Choi, C.; Platz, E. A.; Zhang, H., Overexpression of α (1,6) fucosyltransferase associated with aggressive prostate cancer. *Glycobiology* **2014,** *24* (10), 935-944.

50. Höti, N.; Lih, T.-S.; Pan, J.; Zhou, Y.; Yang, G.; Deng, A.; Chen, L.; Dong, M.; Yang, R.-B.; Tu, C.-F.; Haffner, M. C.; Kay Li, Q.; Zhang, H., A Comprehensive Analysis of FUT8 Overexpressing Prostate Cancer Cells Reveals the Role of EGFR in Castration Resistance. *Cancers* **2020,** *12* (2), 468.

51. Miao, W.; Yuan, J.; Li, L.; Wang, Y., Parallel-Reaction-Monitoring-Based Proteome-Wide Profiling of Differential Kinase Protein Expression during Prostate Cancer Metastasis in Vitro. *Analytical Chemistry* **2019,** *91* (15), 9893-9900.

52. Collins, C. S.; Hong, J.; Sapinoso, L.; Zhou, Y.; Liu, Z.; Micklash, K.; Schultz, P. G.; Hampton, G. M., A small interfering RNA screen for modulators of tumor cell motility

identifies MAP4K4 as a promigratory kinase. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** *103* (10), 3775-3780.

53. Sbrissa, D.; Semaan, L.; Govindarajan, B.; Li, Y.; Caruthers, N. J.; Stemmer, P. M.; Cher, M. L.; Sethi, S.; Vaishampayan, U.; Shisheva, A.; Chinni, S. R., A novel cross-talk between CXCR4 and PI4KIIIα in prostate cancer cells. *Oncogene* **2019,** *38* (3), 332-344.

54. Lee, J. K.; Bangayan, N. J.; Chai, T.; Smith, B. A.; Pariva, T. E.; Yun, S.; Vashisht, A.; Zhang, Q.; Park, J. W.; Corey, E.; Huang, J.; Graeber, T. G.; Wohlschlegel, J.; Witte, O. N., Systemic surfaceome profiling identifies target antigens for immune-based therapy in subtypes of advanced prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America* **2018,** *115* (19), E4473-E4482.

55. Zhou, J.; Yang, W.; Hu, Y.; Höti, N.; Liu, Y.; Shah, P.; Sun, S.; Clark, D.; Thomas, S.; Zhang, H., Site-Specific Fucosylation Analysis Identifying Glycoproteins Associated with Aggressive Prostate Cancer Cell Lines Using Tandem Affinity Enrichments of Intact Glycopeptides Followed by Mass Spectrometry. *Analytical Chemistry* **2017,** *89* (14), 7623-7630.

56. Zhou, B.; Yan, Y.; Wang, Y.; You, S.; Freeman, M. R.; Yang, W., Quantitative proteomic analysis of prostate tissue specimens identifies deregulated protein complexes in primary prostate cancer. *Clinical Proteomics* **2019,** *16* (1), 15-15.

57. Höti, N.; Shah, P.; Hu, Y.; Yang, S.; Zhang, H., Proteomics analyses of prostate cancer cells reveal cellular pathways associated with androgen resistance. *PROTEOMICS* **2017,** *17* (6), 1600228-1600228.

58. Zhang, S.; Zheng, C.; Yao, S.; Wang, Z.; Xu, L.; Yang, R.; Meng, X.; Wu, J.; Zhou, L.; Sun, Z., Proteomic analysis of human prostate cancer PC-3M-1E8 cells and PC-3M-2B4 cells of same origin but with different metastatic potential. *PLOS ONE* **2018,** *13* (10), e0206139-e0206139.

59. Webber, J. P.; Spary, L. K.; Mason, M. D.; Tabi, Z.; Brewis, I. A.; Clayton, A., Prostate stromal cell proteomics analysis discriminates normal from tumour reactive stromal phenotypes. *Oncotarget; Vol 7, No 15* **2016**.

60. Wang, S.; Gao, J.; Lei, Q.; Rozengurt, N.; Pritchard, C.; Jiao, J.; Thomas, G. V.; Li, G.; Roy-Burman, P.; Nelson, P. S.; Liu, X.; Wu, H., Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer. *Cancer Cell* **2003,** *4* (3), 209-221.

61. Zhang, J.; Kim, S.; Li, L.; Kemp, C. J.; Jiang, C.; Lü, J., Proteomic and transcriptomic profiling of Pten gene-knockout mouse model of prostate cancer. *The Prostate* **2020,** *80* (7), 588-605.

62. Hensley, P. J.; Kyprianou, N., Modeling prostate cancer in mice: limitations and opportunities. *J Androl* **2012,** *33* (2), 133-144.

63. Liu, T. T.; Ewald, J. A.; Ricke, E. A.; Bell, R.; Collins, C.; Ricke, W. A., Modeling human prostate cancer progression in vitro. *Carcinogenesis* **2019,** *40* (7), 893-902.

64. Davalieva, K.; Kiprijanovska, S.; Kostovska, I. M.; Stavridis, S.; Stankov, O.; Komina, S.; Petrusevska, G.; Polenakovic, M., Comparative proteomics analysis of urine reveals down-regulation of acute phase response signaling and LXR/RXR activation pathways in prostate cancer. *Proteomes* **2018,** *6* (1).

65. Soekmadji, C.; Riches, J. D.; Russell, P. J.; Ruelcke, J. E.; McPherson, S.; Wang, C.; Hovens, C. M.; Corcoran, N. M.; Hill, M. M.; Nelson, C. C., Modulation of paracrine signaling by CD9 positive small extracellular vesicles mediates cellular growth of androgen deprived prostate cancer. *Oncotarget* **2017,** *8* (32), 52237-52255.

66. Sequeiros, T.; Rigau, M.; Chiva, C.; Montes, M.; Garcia-Grau, I.; Garcia, M.; Diaz, S.; Celma, A.; Bijnsdorp, I.; Campos, A.; Di Mauro, P.; Borrós, S.; Reventós, J.; Doll, A.; Paciucci, R.; Pegtel, M.; de Torres, I.; Sabidó, E.; Morote, J.; Olivan, M., Targeted proteomics in urinary extracellular vesicles identifies biomarkers for diagnosis and prognosis of prostate cancer. *Oncotarget* **2017,** *8* (3), 4960-4976.

67. Kim, Y.; Jeon, J.; Mejia, S.; Yao, C. Q.; Ignatchenko, V.; Nyalwidhe, J. O.; Gramolini, A. O.; Lance, R. S.; Troyer, D. A.; Drake, R. R.; Boutros, P. C.; Semmes, O. J.; Kislinger, T., Targeted proteomics identifies liquid-biopsy signatures for extracapsular prostate cancer. *Nature Communications* **2016,** *7* (1), 11906-11906.

68. Karakosta, T. D.; Soosaipillai, A.; Diamandis, E. P.; Batruch, I.; Drabovich, A. P., Quantification of Human Kallikrein-Related Peptidases in Biological Fluids by Multi-Platform Targeted Mass Spectrometry Assays. *Molecular & Cellular Proteomics* **2016**, mcp.M115.057695.

69. Fujita, K.; Kume, H.; Matsuzaki, K.; Kawashima, A.; Ujike, T.; Nagahara, A.; Uemura, M.; Miyagawa, Y.; Tomonaga, T.; Nonomura, N., Proteomic analysis of urinary extracellular vesicles from high Gleason score prostate cancer. *Scientific Reports* **2017,** *7* (1), 42961-42961.

70. Epstein, J. I.; Egevad, L.; Amin, M. B.; Delahunt, B.; Srigley, J. R.; Humphrey, P. A., The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol* **2016,** *40* (2), 244-52.

71. Yan, B.; Chen, B.; Min, S.; Gao, Y.; Zhang, Y.; Xu, P.; Li, C.; Chen, J.; Luo, G.; Liu, C., iTRAQ-based Comparative Serum Proteomic Analysis of Prostate Cancer Patients with or without Bone Metastasis. *Journal of Cancer* **2019,** *10* (18), 4165-4177.

72. Larkin, S. E. T.; Johnston, H. E.; Jackson, T. R.; Jamieson, D. G.; Roumeliotis, T. I.; Mockridge, C. I.; Michael, A.; Manousopoulou, A.; Papachristou, E. K.; Brown, M. D.; Clarke, N. W.; Pandha, H.; Aukim-Hastie, C. L.; Cragg, M. S.; Garbis, S. D.; Townsend, P. A., Detection of candidate biomarkers of prostate cancer progression in serum: a depletion-free 3D LC/MS quantitative proteomics pilot study. *British Journal of Cancer* **2016,** *115* (9), 1078-1086.

73. Loeb, S.; Bjurlin, M. A.; Nicholson, J.; Tammela, T. L.; Penson, D. F.; Carter, H. B.; Carroll, P.; Etzioni, R., Overdiagnosis and overtreatment of prostate cancer. *Eur Urol* **2014,** *65* (6), 1046-55.

74. Rawla, P.; Sunkara, T.; Gaduputi, V., Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. *World J Oncol* **2019,** *10* (1), 10-27.

75. Rauth, S.; Karmakar, S.; Shah, A.; Nimmakayala, R. K.; Bhatia, R.; Muniyan, S.; Kumar, S.; Dutta, S.; Datta, K.; Batra, S. K.; Ponnusamy, M. P., Abstract 4438: Role of post translational modification of PAF1/PD2 in gemcitabine resistance of pancreatic cancer. *Cancer Research* **2019,** *79* (13 Supplement), 4438-4438.

76. Ansari, D.; Torén, W.; Zhou, Q.; Hu, D.; Andersson, R., Proteomic and genomic profiling of pancreatic cancer. *Cell Biology and Toxicology* **2019,** *35* (4), 333-343.

77. Yang, S.-z.; Xu, F.; Yuan, K.; Sun, Y.; Zhou, T.; Zhao, X.; McDonald, J. M.; Chen, Y., Regulation of pancreatic cancer TRAIL resistance by protein O-GlcNAcylation. *Laboratory Investigation* **2020,** *100* (5), 777-785.

78. Liu, P.; Weng, Y.; Sui, Z.; Wu, Y.; Meng, X.; Wu, M.; Jin, H.; Tan, X.; Zhang, L.; Zhang, Y., Quantitative secretomic analysis of pancreatic cancer cells in serum-containing conditioned medium. *Scientific Reports* **2016,** *6* (1), 37606.

79. Marchand, B.; Arsenault, D.; Raymond-Fleury, A.; Boisvert, F.-M.; Boucher, M.-J., Glycogen Synthase Kinase-3 (GSK3) Inhibition Induces Prosurvival Autophagic Signals in Human Pancreatic Cancer Cells. *Journal of Biological Chemistry* **2015,** *290* (9), 5592-5605.

80. Shi, Y.; Gao, W.; Lytle, N. K.; Huang, P.; Yuan, X.; Dann, A. M.; Ridinger-Saison, M.; DelGiorno, K. E.; Antal, C. E.; Liang, G.; Atkins, A. R.; Erikson, G.; Sun, H.; Meisenhelder, J.; Terenziani, E.; Woo, G.; Fang, L.; Santisakultarm, T. P.; Manor, U.; Xu, R.; Becerra, C. R.; Borazanci, E.; Von Hoff, D. D.; Grandgenett, P. M.; Hollingsworth, M. A.; Leblanc, M.; Umetsu, S. E.; Collisson, E. A.; Scadeng, M.; Lowy, A. M.; Donahue, T. R.; Reya, T.; Downes, M.; Evans, R. M.; Wahl, G. M.; Pawson, T.; Tian, R.; Hunter, T., Targeting LIF-mediated paracrine interaction for pancreatic cancer therapy and monitoring. *Nature* **2019,** *569* (7754), 131-135.

81. Roberts, A. M.; Miyamoto, D. K.; Huffman, T. R.; Bateman, L. A.; Ives, A. N.; Akopian, D.; Heslin, M. J.; Contreras, C. M.; Rape, M.; Skibola, C. F.; Nomura, D. K.,

Chemoproteomic Screening of Covalent Ligands Reveals UBA5 As a Novel Pancreatic Cancer Target. *ACS Chemical Biology* **2017,** *12* (4), 899-904.

82. Liu, C.-W.; Zhang, Q., Isobaric Labeling-Based LC-MS/MS Strategy for Comprehensive Profiling of Human Pancreatic Tissue Proteome BT - Tissue Proteomics: Methods and Protocols. Sarwal, M. M.; Sigdel, T. K.; Sarwal, M. M.; Sigdel, T. K., Eds. New York, NY, 2018; pp 215-224.

83. Perera, R. M.; Stoykova, S.; Nicolay, B. N.; Ross, K. N.; Fitamant, J.; Boukhali, M.; Lengrand, J.; Deshpande, V.; Selig, M. K.; Ferrone, C. R.; Settleman, J.; Stephanopoulos, G.; Dyson, N. J.; Zoncu, R.; Ramaswamy, S.; Haas, W.; Bardeesy, N., Transcriptional control of autophagy–lysosome function drives pancreatic cancer metabolism. *Nature* **2015,** *524* (7565), 361-365.

84. An, M.; Lohse, I.; Tan, Z.; Zhu, J.; Wu, J.; Kurapati, H.; Morgan, M. A.; Lawrence, T. S.; Cuneo, K. C.; Lubman, D. M., Quantitative Proteomic Analysis of Serum Exosomes from Patients with Locally Advanced Pancreatic Cancer Undergoing Chemoradiotherapy. *Journal of Proteome Research* **2017,** *16* (4), 1763-1772.

85. Li, H.; Mao, Y.; Xiong, Y.; Zhao, H. H.; Shen, F.; Gao, X.; Yang, P.; Liu, X.; Fu, D., A Comprehensive Proteome Analysis of Peripheral Blood Mononuclear Cells (PBMCs) to Identify Candidate Biomarkers of Pancreatic Cancer. *Cancer Genomics - Proteomics* **2019,** *16* (1), 81-89.

86. Frost, D. C.; Greer, T.; Xiang, F.; Liang, Z.; Li, L., Development and characterization of novel 8-plex DiLeu isobaric labels for quantitative proteomics and peptidomics. *Rapid communications in mass spectrometry : RCM* **2015,** *29* (12), 1115-1124.

87. Frost, D. C.; Rust, C. J.; Robinson, R. A. S.; Li, L., Increased N,N-Dimethyl Leucine Isobaric Tag Multiplexing by a Combined Precursor Isotopic Labeling and Isobaric Tagging Approach. *Analytical Chemistry* **2018,** *90* (18), 10664-10669.

88. Frost, D. C.; Greer, T.; Li, L., High-resolution enabled 12-plex DiLeu isobaric tags for quantitative proteomics. *Analytical chemistry* **2015,** *87* (3), 1646-1654.

89. Greer, T.; Lietz, C. B.; Xiang, F.; Li, L., Novel isotopic N,N-dimethyl leucine (iDiLeu) reagents enable absolute quantification of peptides and proteins using a standard curve approach. *Journal of the American Society for Mass Spectrometry* **2015,** *26* (1), 107-119.

90. Zhong, X.; Frost, D. C.; Li, L., High-Resolution Enabled 5-plex Mass Defect-Based N,N-Dimethyl Leucine Tags for Quantitative Proteomics. *Analytical Chemistry* **2019,** *91* (13), 7991-7995.

91. Zhong, X.; Yu, Q.; Ma, F.; Frost, D. C.; Lu, L.; Chen, Z.; Zetterberg, H.; Carlsson, C. M.; Okonkwo, O.; Li, L., HOTMAQ: a multiplexed absolute quantification method for targeted proteomics. *Analytical Chemistry* **2019**.

92. Hao, L.; Johnson, J.; Lietz, C. B.; Buchberger, A.; Frost, D.; Kao, W. J.; Li, L., Mass Defect-Based N,N-Dimethyl Leucine Labels for Quantitative Proteomics and Amine Metabolomics of Pancreatic Cancer Cells. *Analytical Chemistry* **2017,** *89* (2), 1138-1146.

93. Wang, X.; Niu, J.; Li, J.; Shen, X.; Shen, S.; Straubinger, R. M.; Qu, J., Temporal effects of combined birinapant and paclitaxel on pancreatic cancer cells investigated via large-scale, ion-current-based quantitative proteomics (IonStar). *Molecular and Cellular Proteomics* **2018,** *17* (4), 655-671.

94. Zhu, X.; Shen, X.; Qu, J.; Straubinger, R. M.; Jusko, W. J., Multi-Scale Network Model Supported by Proteomics for Analysis of Combined Gemcitabine and Birinapant Effects in Pancreatic Cancer Cells. *CPT: Pharmacometrics & Systems Pharmacology* **2018,** *7* (9), 549-561.

95. Singh, C. K.; Kaur, S.; George, J.; Nihal, M.; Pellitteri Hahn, M. C.; Scarlett, C. O.; Ahmad, N., Molecular signatures of sanguinarine in human pancreatic cancer cells: A large scale label-free comparative proteomics approach. *Oncotarget* **2015,** *6* (12), 10335-10348.

96. Zhou, Q.; Andersson, R.; Hu, D.; Bauden, M.; Kristl, T.; Sasor, A.; Pawłowski, K.; Pla, I.; Hilmersson, K. S.; Zhou, M.; Lu, F.; Marko-Varga, G.; Ansari, D., Quantitative proteomics identifies brain acid soluble protein 1 (BASP1) as a prognostic biomarker candidate in pancreatic cancer tissue. *EBioMedicine* **2019,** *43*, 282-294.

97. Jhaveri, D. T.; Kim, M.-S.; Thompson, E. D.; Huang, L.; Sharma, R.; Klein, A. P.; Zheng, L.; Le, D. T.; Laheru, D. A.; Pandey, A.; Jaffee, E. M.; Anders, R. A., Using Quantitative Seroproteomics to Identify Antibody Biomarkers in Pancreatic Cancer. *Cancer Immunology Research* **2016,** *4* (3), 225 LP - 233.

98. Sogawa, K.; Takano, S.; Iida, F.; Satoh, M.; Tsuchida, S.; Kawashima, Y.; Yoshitomi, H.; Sanda, A.; Kodera, Y.; Takizawa, H.; Mikata, R.; Ohtsuka, M.; Shimizu, H.; Miyazaki, M.; Yokosuka, O.; Nomura, F., Identification of a novel serum biomarker for pancreatic cancer, C4b-binding protein α-chain (C4BPA) by quantitative proteomic analysis using tandem mass tags. *British Journal of Cancer* **2016,** *115* (8), 949-956.

99. Naba, A.; Clauser, K. R.; Mani, D. R.; Carr, S. A.; Hynes, R. O., Quantitative proteomic profiling of the extracellular matrix of pancreatic islets during the angiogenic switch and insulinoma progression. *Scientific Reports* **2017,** *7* (1), 40495.

100. Yu, Z.; Zhao, S.; Ren, L.; Wang, L.; Chen, Z.; Hoffman, R. M.; Zhou, J., Pancreatic cancer-derived exosomes promote tumor metastasis and liver pre-metastatic niche formation. *Oncotarget* **2017,** *8* (38), 63461-63483.

101. Lin, C.; Wu, W.-C.; Zhao, G.-C.; Wang, D.-S.; Lou, W.-H.; Jin, D.-Y., ITRAQ-based quantitative proteomics reveals apolipoprotein A-I and transferrin as potential serum

markers in CA19-9 negative pancreatic ductal adenocarcinoma. *Medicine (Baltimore)* **2016,** *95* (31), e4527-e4527.

102. Liu, X.; Zheng, W.; Wang, W.; Shen, H.; Liu, L.; Lou, W.; Wang, X.; Yang, P., A new panel of pancreatic cancer biomarkers discovered using a mass spectrometry-based pipeline. *British Journal of Cancer* **2017,** *117* (12), 1846-1854.

103. Ohmine, K.; Kawaguchi, K.; Ohtsuki, S.; Motoi, F.; Ohtsuka, H.; Kamiie, J.; Abe, T.; Unno, M.; Terasaki, T., Quantitative Targeted Proteomics of Pancreatic Cancer: Deoxycytidine Kinase Protein Level Correlates to Progression-Free Survival of Patients Receiving Gemcitabine Treatment. *Molecular Pharmaceutics* **2015,** *12* (9), 3282-3291.

104. Yoneyama, T.; Ohtsuki, S.; Honda, K.; Kobayashi, M.; Iwasaki, M.; Uchida, Y.; Okusaka, T.; Nakamori, S.; Shimahara, M.; Ueno, T.; Tsuchida, A.; Sata, N.; Ioka, T.; Yasunami, Y.; Kosuge, T.; Kaneda, T.; Kato, T.; Yagihara, K.; Fujita, S.; Huang, W.; Yamada, T.; Tachikawa, M.; Terasaki, T., Identification of IGFBP2 and IGFBP3 As Compensatory Biomarkers for CA19-9 in Early-Stage Pancreatic Cancer Using a Combination of Antibody-Based and LC-MS/MS-Based Proteomics. *PLOS ONE* **2016,** *11* (8), e0161009.

105. Park, J.; Lee, E.; Park, K.-J.; Park, H.-D.; Kim, J.-W.; Woo, H. I.; Lee, K. H.; Lee, K.-T.; Lee, J. K.; Park, J.-O.; Park, Y. S.; Heo, J. S.; Choi, S. H.; Choi, D. W.; Jang, K.-T.; Lee, S.-Y., Large-scale clinical validation of biomarkers for pancreatic cancer using a mass spectrometry-based proteomics approach. *Oncotarget* **2017,** *8* (26), 42761-42771.

106. Do, M.; Han, D.; Wang, J. I.; Kim, H.; Kwon, W.; Han, Y.; Jang, J. Y.; Kim, Y., Quantitative proteomic analysis of pancreatic cyst fluid proteins associated with malignancy in intraductal papillary mucinous neoplasms. *Clinical Proteomics* **2018,** *15* (1), 1-22.

107. Nigjeh, E. N.; Chen, R.; Brand, R. E.; Petersen, G. M.; Chari, S. T.; von Haller, P. D.; Eng, J. K.; Feng, Z.; Yan, Q.; Brentnall, T. A.; Pan, S., Quantitative Proteomics Based on Optimized Data-Independent Acquisition in Plasma Analysis. *Journal of Proteome Research* **2017,** *16* (2), 665-676.

108. Ginsburg, O. M.; Love, R. R., Breast cancer: a neglected disease for the majority of affected women worldwide. *Breast J* **2011,** *17* (3), 289-295.

109. Miah, S.; Banks, C. A.; Adams, M. K.; Florens, L.; Lukong, K. E.; Washburn, M. P., Advancement of mass spectrometry-based proteomics technologies to explore triple negative breast cancer. *Mol Biosyst* **2016,** *13* (1), 42-55.

110. Sallam, R. M., Proteomics in cancer biomarkers discovery: challenges and applications. *Dis Markers* **2015,** *2015*, 321370.

111. Tyanova, S.; Albrechtsen, R.; Kronqvist, P.; Cox, J.; Mann, M.; Geiger, T., Proteomic maps of breast cancer subtypes. *Nature Communications* **2016,** *7* (1), 10259.

112. Suman, S.; Basak, T.; Gupta, P.; Mishra, S.; Kumar, V.; Sengupta, S.; Shukla, Y., Quantitative proteomics revealed novel proteins associated with molecular subtypes of breast cancer. *Journal of Proteomics* **2016,** *148*, 183-193.

113. Calderón-González, K. G.; Valero Rustarazo, M. L.; Labra-Barrios, M. L.; Bazán-Méndez, C. I.; Tavera-Tapia, A.; Herrera-Aguirre, M. E.; Sánchez del Pino, M. M.; Gallegos-Pérez, J. L.; González-Márquez, H.; Hernández-Hernández, J. M.; León-Ávila, G.; Rodríguez-Cuevas, S.; Guisa-Hohenstein, F.; Luna-Arias, J. P., Determination of the protein expression profiles of breast cancer cell lines by quantitative proteomics using iTRAQ labelling and tandem mass spectrometry. *Journal of Proteomics* **2015,** *124*, 50-78.

114. Gajbhiye, A.; Dabhi, R.; Taunk, K.; Vannuruswamy, G.; RoyChoudhury, S.; Adhav, R.; Seal, S.; Mane, A.; Bayatigeri, S.; Santra, M. K.; Chaudhury, K.; Rapole, S., Urinary proteome alterations in HER2 enriched breast cancer revealed by multipronged quantitative proteomics. *PROTEOMICS* **2016,** *16* (17), 2403-2418.

115. Going, C. C.; Tailor, D.; Kumar, V.; Birk, A. M.; Pandrala, M.; Rice, M. A.; Stoyanova, T.; Malhotra, S.; Pitteri, S. J., Quantitative Proteomic Profiling Reveals Key Pathways in the Anticancer Action of Methoxychalcone Derivatives in Triple Negative Breast Cancer. *Journal of Proteome Research* **2018,** *17* (10), 3574-3585.

116. Clark, D. J.; Fondrie, W. E.; Liao, Z.; Hanson, P. I.; Fulton, A.; Mao, L.; Yang, A. J., Redefining the Breast Cancer Exosome Proteome by Tandem Mass Tag Quantitative Proteomics and Multivariate Cluster Analysis. *Analytical Chemistry* **2015,** *87* (20), 10462-10469.

117. Liu, F.; Ma, F.; Wang, Y.; Hao, L.; Zeng, H.; Jia, C.; Wang, Y.; Liu, P.; Ong, Irene M.; Li, B.; Chen, G.; Jiang, J.; Gong, S.; Li, L.; Xu, W., PKM2 methylation by CARM1 activates aerobic glycolysis to promote tumorigenesis. *Nature Cell Biology* **2017,** *19* (11), 1358-1370.

118. Johansson, H. J.; Socciarelli, F.; Vacanti, N. M.; Haugen, M. H.; Zhu, Y.; Siavelis, I.; Fernandez-Woodbridge, A.; Aure, M. R.; Sennblad, B.; Vesterlund, M.; Branca, R. M.; Orre, L. M.; Huss, M.; Fredlund, E.; Beraki, E.; Garred, Ø.; Boekel, J.; Sauer, T.; Zhao, W.; Nord, S.; Höglander, E. K.; Jans, D. C.; Brismar, H.; Haukaas, T. H.; Bathen, T. F.; Schlichting, E.; Naume, B.; Geisler, J.; Hofvind, S.; Engebråten, O.; Geitvik, G. A.; Langerød, A.; Kåresen, R.; Mælandsmo, G. M.; Sørlie, T.; Skjerven, H. K.; Park, D.; Hartman-Johnsen, O.-J.; Luders, T.; Borgen, E.; Kristensen, V. N.; Russnes, H. G.; Lingjærde, O. C.; Mills, G. B.; Sahlberg, K. K.; Børresen-Dale, A.-L.; Lehtiö, J.; Consortia Oslo Breast Cancer Research, C., Breast cancer quantitative proteome and proteogenomic landscape. *Nature Communications* **2019,** *10* (1), 1600.

119. Zhou, J.-Y.; Chen, L.; Zhang, B.; Tian, Y.; Liu, T.; Thomas, S. N.; Chen, L.; Schnaubelt, M.; Boja, E.; Hiltke, T.; Kinsinger, C. R.; Rodriguez, H.; Davies, S. R.; Li, S.; Snider, J. E.; Erdmann-Gilmore, P.; Tabb, D. L.; Townsend, R. R.; Ellis, M. J.; Rodland, K. D.; Smith, R. D.; Carr, S. A.; Zhang, Z.; Chan, D. W.; Zhang, H., Quality Assessments of Long-Term Quantitative Proteomic Analysis of Breast Cancer Xenograft Tissues. *Journal of Proteome Research* **2017,** *16* (12), 4523-4530.

120. Ntai, I.; LeDuc, R. D.; Fellers, R. T.; Erdmann-Gilmore, P.; Davies, S. R.; Rumsey, J.; Early, B. P.; Thomas, P. M.; Li, S.; Compton, P. D.; Ellis, M. J. C.; Ruggles, V. K.; Fenyö, D.; Boja, E. S.; Rodriguez, H.; Townsend, R. R.; Kelleher, N. L., Integrated bottom-up and top-down proteomics of patient-derived breast tumor xenografts. *Molecular and Cellular Proteomics* **2016,** *15* (1), 45-56.

121. Tveitarås, M. K.; Selheim, F.; Sortland, K.; Reed, R. K.; Stuhr, L., Protein expression profiling of plasma and lungs at different stages of metastatic development in a human triple negative breast cancer xenograft model. *PLoS ONE* **2019,** *14* (5), 1-16.

122. Wang, Z.; Liu, H.; Yan, Y.; Yang, X.; Zhang, Y.; Wu, L., Integrated Proteomic and N-Glycoproteomic Analyses of Human Breast Cancer. *Journal of Proteome Research* **2020**.

123. Gámez-Pozo, A.; Berges-Soria, J.; Arevalillo, J. M.; Nanni, P.; López-Vacas, R.; Navarro, H.; Grossmann, J.; Castaneda, C. A.; Main, P.; Díaz-Almirón, M.; Espinosa, E.; Ciruelos, E.; Fresno Vara, J. Á., Combined Label-Free Quantitative Proteomics and microRNA Expression Analysis of Breast Cancer Unravel Molecular Differences with Clinical Implications. *Cancer Research* **2015,** *75* (11), 2243-2253.

124. Nie, S.; McDermott, S. P.; Deol, Y.; Tan, Z.; Wicha, M. S.; Lubman, D. M., A quantitative proteomics analysis of MCF7 breast cancer stem and progenitor cell populations. *PROTEOMICS* **2015,** *15* (22), 3772-3783.

125. Warmoes, M.; Lam, S. W.; van der Groep, P.; Jaspers, J. E.; Smolders, Y. H. C. M.; Boer, d. L.; Pham, V. T.; Piersma, S. R.; Rottenberg, S.; Boven, E.; Jonkers, J.; van Diest, P. J.; Jimenez, C. R., Secretome proteomics reveals candidate non-invasive biomarkers of BRCA1 deficiency in breast cancer. *Oncotarget* **2016,** *7* (39), 63537-63548.

126. Siegel, R. L.; Miller, K. D.; Jemal, A., Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians* **2020,** *70* (1), 7-30.

127. Musrap, N.; Tuccitto, A.; Karagiannis, G. S.; Saraon, P.; Batruch, I.; Diamandis, E. P., Comparative proteomics of ovarian cancer aggregate formation reveals an increased expression of calcium-activated chloride channel regulator 1 (CLCA1). *Journal of Biological Chemistry* **2015,** *290* (28), 17218-17227.

128. Grassi, M. L.; Palma, C. d. S.; Thomé, C. H.; Lanfredi, G. P.; Poersch, A.; Faça, V. M., Proteomic analysis of ovarian cancer cells during epithelial-mesenchymal transition

(EMT) induced by epidermal growth factor (EGF) reveals mechanisms of cell cycle control. *Journal of Proteomics* **2017,** *151*, 2-11.

129. Ji, Y.; Wei, S.; Hou, J.; Zhang, C.; Xue, P.; Wang, J.; Chen, X.; Guo, X.; Yang, F., Integrated proteomic and N-glycoproteomic analyses of doxorubicin sensitive and resistant ovarian cancer cells reveal glycoprotein alteration in protein abundance and glycosylation. *Oncotarget* **2017,** *8* (8), 13413-13427.

130. Zhang, H.; Liu, T.; Zhang, Z.; Payne, S. H.; Zhang, B.; McDermott, J. E.; Zhou, J. Y.; Petyuk, V. A.; Chen, L.; Ray, D.; Sun, S.; Yang, F.; Chen, L.; Wang, J.; Shah, P.; Cha, S. W.; Aiyetan, P.; Woo, S.; Tian, Y.; Gritsenko, M. A.; Clauss, T. R.; Choi, C.; Monroe, M. E.; Thomas, S.; Nie, S.; Wu, C.; Moore, R. J.; Yu, K. H.; Tabb, D. L.; Fenyö, D.; Vineet, V.; Wang, Y.; Rodriguez, H.; Boja, E. S.; Hiltke, T.; Rivers, R. C.; Sokoll, L.; Zhu, H.; Shih, I. M.; Cope, L.; Pandey, A.; Zhang, B.; Snyder, M. P.; Levine, D. A.; Smith, R. D.; Chan, D. W.; Rodland, K. D.; Carr, S. A.; Gillette, M. A.; Klauser, K. R.; Kuhn, E.; Mani, D. R.; Mertins, P.; Ketchum, K. A.; Thangudu, R.; Cai, S.; Oberti, M.; Paulovich, A. G.; Whiteaker, J. R.; Edwards, N. J.; McGarvey, P. B.; Madhavan, S.; Wang, P.; Whiteley, G. A.; Skates, S. J.; White, F. M.; Kinsinger, C. R.; Mesri, M.; Shaw, K. M.; Stein, S. E.; Fenyo, D.; Rudnick, P.; Snyder, M.; Zhao, Y.; Chen, X.; Ransohoff, D. F.; Hoofnagle, A. N.; Liebler, D. C.; Sanders, M. E.; Shi, Z.; Slebos, R. J. C.; Zimmerman, L. J.; Davies, S. R.; Ding, L.; Ellis, M. J. C.; Townsend, R. R., Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* **2016,** *166* (3), 755-765.

131. Hiramatsu, K.; Yoshino, K.; Serada, S.; Yoshihara, K.; Hori, Y.; Fujimoto, M.; Matsuzaki, S.; Egawa-Takata, T.; Kobayashi, E.; Ueda, Y.; Morii, E.; Enomoto, T.; Naka, T.; Kimura, T., Similar protein expression profiles of ovarian and endometrial high-grade serous carcinomas. *British Journal of Cancer* **2016,** *114* (5), 554-561.

132. Hu, Y.; Pan, J.; Shah, P.; Ao, M.; Thomas, S. N.; Liu, Y.; Chen, L.; Schnaubelt, M.; Clark, D. J.; Rodriguez, H.; Boja, E. S.; Hiltke, T.; Kinsinger, C. R.; Rodland, K. D.; Li, Q. K.; Qian, J.; Zhang, Z.; Chan, D. W.; Zhang, H.; Pandey, A.; Paulovich, A.; Hoofnagle, A.; Zhang, B.; Mani, D. R.; Liebler, D. C.; Ransohoff, D. F.; Fenyo, D.; Tabb, D. L.; Levine, D. A.; Kuhn, E.; White, F. M.; Whiteley, G. A.; Zhu, H.; Shih, I.-M.; Bavarva, J.; McDermott, J. E.; Whiteaker, J.; Ketchum, K. A.; Clauser, K. R.; Ruggles, K.; Elburn, K.; Ding, L.; Hannick, L.; Zimmerman, L. J.; Watson, M.; Thiagarajan, M.; Ellis, M. J. C.; Oberti, M.; Mesri, M.; Sanders, M. E.; Borucki, M.; Gillette, M. A.; Snyder, M.; Edwards, N. J.; Vatanian, N.; Rudnick, P. A.; McGarvey, P. B.; Mertins, P.; Townsend, R. R.; Thangudu, R. R.; Smith, R. D.; Rivers, R. C.; Slebos, R. J. C.; Payne, S. H.; Davies, S. R.; Cai, S.; Stein, S. E.; Carr, S. A.; Skates, S. J.; Madhavan, S.; Liu, T.; Chen, X.; Zhao, Y.; Wang, Y.; Shi, Z., Integrated Proteomic and Glycoproteomic Characterization of Human High-Grade Serous Ovarian Carcinoma. *Cell Reports* **2020,** *33* (3).

133. Yoshimura, A.; Sawada, K.; Nakamura, K.; Kinose, Y.; Nakatsuka, E.; Kobayashi, M.; Miyamoto, M.; Ishida, K.; Matsumoto, Y.; Kodama, M.; Hashimoto, K.; Mabuchi, S.;

Kimura, T., Exosomal miR-99a-5p is elevated in sera of ovarian cancer patients and promotes cancer cell invasion by increasing fibronectin and vitronectin expression in neighboring peritoneal mesothelial cells. *BMC Cancer* **2018,** *18* (1), 1-13.

134. Qu, H.; Chen, Y.; Cao, G.; Liu, C.; Xu, J.; Deng, H.; Zhang, Z., Identification and validation of differentially expressed proteins in epithelial ovarian cancers using quantitative proteomics. *Oncotarget* **2016,** *7* (50), 83187-83199.

135. McDermott, J. E.; Arshad, O. A.; Petyuk, V. A.; Fu, Y.; Gritsenko, M. A.; Clauss, T. R.; Moore, R. J.; Schepmoes, A. A.; Zhao, R.; Monroe, M. E.; Schnaubelt, M.; Tsai, C.-F.; Payne, S. H.; Huang, C.; Wang, L.-B.; Foltz, S.; Wyczalkowski, M.; Wu, Y.; Song, E.; Brewer, M. A.; Thiagarajan, M.; Kinsinger, C. R.; Robles, A. I.; Boja, E. S.; Rodriguez, H.; Chan, D. W.; Zhang, B.; Zhang, Z.; Ding, L.; Smith, R. D.; Liu, T.; Rodland, K. D.; Consortium, C. T. A., Proteogenomic Characterization of Ovarian HGSC Implicates Mitotic Kinases, Replication Stress in Observed Chromosomal Instability. *Cell reports. Medicine* **2020,** *1* (1), 100004-100004.

136. Chuffa, L. G. A.; Lupi Júnior, L. A.; Seiva, F. R. F.; Martinez, M.; Domeniconi, R. F.; Pinheiro, P. F. F.; dos Santos, L. D.; Martinez, F. E., Quantitative Proteomic Profiling Reveals That Diverse Metabolic Pathways Are Influenced by Melatonin in an in Vivo Model of Ovarian Carcinoma. *Journal of Proteome Research* **2016,** *15* (10), 3872-3882.

137. Júnior, L. A. L.; Cucielo, M. S.; Domeniconi, R. F.; Dos Santos, L. D.; Silveira, H. S.; Da Silva Nunes, I.; Martinez, M.; Martinez, F. E.; Fávaro, W. J.; Chuffa, L. G. D. A., P-MAPA and IL-12 Differentially Regulate Proteins Associated with Ovarian Cancer Progression: A Proteomic Study. *ACS Omega* **2019,** *4* (26), 21761-21777.

138. Coscia, F.; Lengyel, E.; Duraiswamy, J.; Ashcroft, B.; Bassani-Sternberg, M.; Wierer, M.; Johnson, A.; Wroblewski, K.; Montag, A.; Yamada, S. D.; López-Méndez, B.; Nilsson, J.; Mund, A.; Mann, M.; Curtis, M., Multi-level Proteomics Identifies CT45 as a Chemosensitivity Mediator and Immunotherapy Target in Ovarian Cancer. *Cell* **2018,** *175* (1), 159-170.e16.

139. Gee, M. E.; Faraahi, Z.; McCormick, A.; Edmondson, R. J., DNA damage repair in ovarian cancer: unlocking the heterogeneity. *Journal of ovarian research* **2018,** *11* (1), 50-50.

140. Zhang, W.; Ou, X.; Wu, X., Proteomics profiling of plasma exosomes in epithelial ovarian cancer: A potential role in the coagulation cascade, diagnosis and prognosis. *International Journal of Oncology* **2019,** *54* (5), 1719-1733.

141. Zhang, W.; Peng, P.; Ou, X.; Shen, K.; Wu, X., Ovarian cancer circulating extracelluar vesicles promote coagulation and have a potential in diagnosis: an iTRAQ based proteomic analysis. *BMC cancer* **2019,** *19* (1), 1095-1095.

142. Swiatly, A.; Horala, A.; Matysiak, J.; Hajduk, J.; Nowak-Markwitz, E.; Kokot, Z. J., Understanding ovarian cancer: iTRAQ-based proteomics for biomarker discovery. *International Journal of Molecular Sciences* **2018,** *19* (8).

143. Russell, M. R.; Walker, M. J.; Williamson, A. J. K.; Gentry-Maharaj, A.; Ryan, A.; Kalsi, J.; Skates, S.; D'Amato, A.; Dive, C.; Pernemalm, M.; Humphryes, P. C.; Fourkala, E. O.; Whetton, A. D.; Menon, U.; Jacobs, I.; Graham, R. L. J., Protein Z: A putative novel biomarker for early detection of ovarian cancer. *International Journal of Cancer* **2016,** *138* (12), 2984-2992.

144. Barnabas, G. D.; Bahar-Shany, K.; Sapoznik, S.; Helpman, L.; Kadan, Y.; Beiner, M.; Weitzner, O.; Arbib, N.; Korach, J.; Perri, T.; Katz, G.; Blecher, A.; Brandt, B.; Friedman, E.; Stockheim, D.; Jakobson-Setton, A.; Eitan, R.; Armon, S.; Brand, H.; Zadok, O.; Aviel-Ronen, S.; Harel, M.; Geiger, T.; Levanon, K., Microvesicle proteomic profiling of uterine liquid biopsy for ovarian cancer early detection. *Molecular and Cellular Proteomics* **2019,** *18* (5), 865-875.

145. Zhang, Z.; Qin, K.; Zhang, W.; Yang, B.; Zhao, C.; Zhang, X.; Zhang, F.; Zhao, L.; Shan, B., Postoperative recurrence of epithelial ovarian cancer patients and chemoresistance related protein analyses. *Journal of Ovarian Research* **2019,** *12* (1), 1-8.

146. Ahn, H.-S.; Yeom, J.; Yu, J.; Kwon, Y.-I.; Kim, J.-H.; Kim, K., Convergence of Plasma Metabolomics and Proteomics Analysis to Discover Signatures of High-Grade Serous Ovarian Cancer. *Cancers* **2020,** *12* (11), 3447-3447.

147. Hüttenhain, R.; Choi, M.; Martin de la Fuente, L.; Oehl, K.; Chang, C.-Y.; Zimmermann, A.-K.; Malander, S.; Olsson, H.; Surinova, S.; Clough, T.; Heinzelmann-Schwarz, V.; Wild, P. J.; Dinulescu, D. M.; Niméus, E.; Vitek, O.; Aebersold, R., A Targeted Mass Spectrometry Strategy for Developing Proteomic Biomarkers: A Case Study of Epithelial Ovarian Cancer. *Molecular & Cellular Proteomics* **2019,** *18* (9), 1836 LP-1850.

148. Rauniyar, N.; Peng, G.; Lam, T. K. T.; Zhao, H.; Mor, G.; Williams, K. R., Data-Independent Acquisition and Parallel Reaction Monitoring Mass Spectrometry Identification of Serum Biomarkers for Ovarian Cancer. *Biomarker Insights* **2017,** *12*.

Chapter 7: Proteomic Fingerprinting of Prostate Cancer Progression Through Library-Free DIA-MS Reveals Systematic and Conserved Pathway Dysregulation

Delafield, D. G., Miles, H. N., Liu, T.T., Ricke, W. A., & Li, L. Proteomic Fingerprinting of Prostate Cancer Progression Through Library-Free DIA-MS Reveals Systematic and Conserved Pathway Dysregulation. In submission. 2023.

Abbreviations

ACN (acetonitrile), ACTN1 (actin alpha 1), AGC (automatic gain control), AKT (protein kinase b), ANLN (anillin), AURKB (aurora kinase b), BCA (bicinchoninic acid), BCAM25 (basal cell adhesion molecule), BCAT1 (branched chain amino acid transferase), BCaP (benign prostate-derived cancer progression), BPH (benign prostate hypertrophy), BUB1 (mitotic checkpoint serine/threonine kinase), CALD126 (caldesmon), CAV1 (caveolin-1), CDC20 (cell division cycle 20), CEP78 (centrosomal protein), CLDN1 (caludin-1), COL17A1 (collagen 17), CV (coefficient of variance), DIA (data-independent acquisition), DIA-NN (DIA-neural network), DPS (DNA-binding protein from starved cells), DRE (digital rectal exam), EGFR (epidermal growth factor receptor), FA (formic acid), FAH (fumarylacetoacetate hydrolase), FDR (false discovery rate), FSTL118 (follistatin-related protein 1), G6PD (Glucose-6-phosphate dehydrogenase), GCLC (glutamate-cysteine ligase catalytic subunit), GCLM (glutamate-cysteine ligase modifier subunit), GO (gene ontology), GOT (glutamic oxaloacetate transaminase), GPX (glutathione peroxidase), GSH (glutathione), GSN (gelsolin), GSS (glutathione synthase), GSTM4 (glutathione S-transferase Mu), HCD (higher-energy C-trap dissociation), HIF (hypoxia induced factor), HLA (human leukocyte antigen), HYOU1 (hypoxia upregulated 1), IFI35 (interferon induced protein), IFIT (interferon induced protein with tripartate motif), IFN (interferon), INA (internexin), INCENP (inner centromere protein), ISG (interferon-stimulated gene), ISRE (interferon-stimulated response element), IVL (involucrin), JAK (janus kinase), KIF (kinesin superfamily protein), KNL1 (kinetochore null protein), KNSTRN (kinetochore localized astrin), KRT (keratin), LC-MS (liquid chromatography-mass spectrometry), LOY (loss of Y), MBR (match between runs), MPRIP (myosin phosphatase rho), MS (mass spectrometry), MTRR (5-Methyltetrahydrofolate-Homocysteine Methyltransferase Reductase), MYH (myosin heavy chain

14), MYL (myosin light polypeptide), MYO (mysoin), NUF2 (kinetochore protein nuf2), OAS (Oligoadenylate synthase), PDK1 (3-phosphoinositide-dependent kinase 1), PFKM (ATP-dependent 6-phosphofructokinase, muscle), PGM1 (Phosphoglucomutase 1), PI3K (Phosphoinositide 3-kinase), PKN1 (protein kinase N1), PPL (periplakin), PPP2R5D (protein phosphatase 2 regulatory subunit B'delta), PRC1 (protein regulator of cytokinesis), PSA (prostate specific antigen), PSAT1 (phosphoserine aminotransferase 1), PSMB8 (proteasome 20S subunit beta 8), PSPH (phosphoserine phosphatase), PYCR (Pyrroline-5-carboxylate reductase), RF1 (release factor 1), ROS (reactive oxygen species), RPLC (reversed-phase liquid chromatography), RPTOR (regulatory associated protein of MTOR complex 1), SLC (solute carrier protein), SPC24 (kinetochore\xa0protein Spc24), SPDL1 (spindly-like protein 1), STAT (signal transducer and activator of transcription), STOM (stomatin), TCA (tricarboxylic acid), TPM (tropomyosin), TPP1 (Tripeptidyl-peptidase 1), UBE2 (ubiquitin conjugating enzyme E2), mTOR (mammalian target of rapamycin), rUGM (Urogenital mesenchymal)

Abstract

Prostate cancer is the second-leading form of cancer in men and diagnoses are expected to increase sharply as risk factors become more prevalent and global populations age. Despite this, clinical diagnoses continue to rely on variable physical examinations, invasive tissue biopsy, and serological screening of poor-specificity biomarkers. There has never been a more pressing need to develop methodologies for accurate diagnosis and stratification, mandating a comprehensive understanding of the biomolecular drivers of prostate cancer. As research towards this goal has traditionally been limited by the lack of model organisms that mimic the progressive genetic, phenotypic, and molecular characteristics human cancer development, the identification and investigation of suitable alternatives remains a principal concern. The recently developed benign prostate hypertrophy-1 cancer progression (BCaP) cell model provides direct reprieve from traditional shortcomings of prostate cancer investigation and enables the confident association of proteomic reorganization with cancer phenotype. Here, we analyze three components of this cell model to illuminate biomolecular alterations as benign prostate cancer transitions and progresses through malignancy. Using library-free data-independent acquisition mass spectrometry, we identify 95,144 peptides and quantify 6,614 proteins with 1,242 shown to be significantly dysregulated in malignant phenotypes. Dysregulated proteins were grouped into 7 distinct diagnostic expression patterns, revealing the progressive reorganization of critical biological processes such as kinetochore formation, cytoskeletal organization, metabolic processing, and interferon signaling. We also provide a topical comparison of transcript and protein level analyses, articulating the importance of proteomic measurements and the need for regular, multimodal analyses. Together, this study presents a primary investigation of the proteomic perturbations

observed in a novel progressive prostate cancer cell model and suggest protein communities useful biomarker validation, therapeutic targeting, and protein-centric cancer diagnosis.

Introduction

Prostate cancer (PCa) is the second most prevalent form of cancer in men[1] and exhibits the highest age-adjusted rate of incidence in numerous populations[2]. As the risk of prostate cancer emergence positively correlates with age[3], maturing demographics in numerous in population-dense and high cancer rate countries[2] significantly exacerbate this global health concern and further emphasizes the need for early detection and treatment. Standard mitigation practices involve regular urological screenings for men considered at risk – digital rectal examinations (DREs) being the standard medical diagnostic[4]. While DRE remains the predominant form of prostate cancer detection, limitations have been noted in consistency[5] and reliability[6]. As well, DREs are elective procedures that only serve as a primary screen, which brings patient reluctance prior to symptom manifestation[7]. While DREs serve to identify irregularities prostate volume and tissue density that accompany prostate cancer development tumor presentation, the unique physiological alterations seen in both healthy aging and diseased prostates present diagnostic challenges that cannot be surmounted by a physical examination alone.

The diversity of physiologies associated with benign and cancerous prostate enlargement presents the principal challenge. Benign prostatic hyperplasia (BPH), which affects >50% of men past the age of 50[8, 9], presents the hallmarks of an enlarged prostate, dysuria and nocturia[10]; these symptoms overlap with those seen in PCa. Furthermore, PCa is a heterogenous, progressive cancer type with four clinically recognized stages[11], each of which presents unique physiological traits. Today, these conditions and phenotypes are diagnosed through a combination of tumor classification, serological PSA evaluation, histological imaging, and Gleason scoring[11]. Biopsy and pathology are

paramount in diagnoses, but limitations are apparent when considering invasiveness, subjectivity[12], and global availability. More confounding is the reliance of PSA as an effective PCa biomarker[13, 14], which has been highly scrutinized for its poor specifity[15-17]. As there is a clear need to directly address these clinical limitations and pitfalls, we posit recent biological and analytical developments may be leveraged to surmount traditional obstacles in PCa investigation and provide a suitable path towards understanding the biomolecular drivers of prostate cancer.

A primary obstacle in PCa investigation has traditionally been the availability of a suitable cell or animal model that mimics the diverse, progressive physiology of human PCa. Though many entries have proved useful in suggesting relevant biological underpinnings[18], unraveling the complex events that inspire PCa initiation and development will likely depend on the use of temporal, human-derived models. Recently, a cell model that mimics the genetic, phenotypic, and molecular characteristics of human prostate cancer was presented by Ricke and colleagues[19]. This BPH-1-derived Cancer Progression (BCaP) cell model enables the study of discrete PCa phenotypes which may be analyzed independently or considered as longitudinal counterparts. This ability to confidently relate biomolecular observations to PCa severity provides a significant step forward in the effort to illuminate the biological foundations of PCa progression and makes the BCaP model a worthwhile subject in contemporary analyses.

A secondary obstacle in PCa investigation is the pragmatic selection of biomolecular analysis. Dedicated focus has established the robustness and sensitivity of transcriptional analyses and validated their utility in identifying relevant biological targets in PCa[20-22]. However, these analyses are fundamentally limited by the lack of a quantitative correlation between transcript evidence and protein abundance[23-25], obfuscating biomolecular understanding and revelation of

active biological participants. Mass spectrometry (MS)-based proteomic analyses, on the other hand, offer direct reprieve from these limitations and is considered the tool-of-choice for contemporary biomolecular interrogation. Critically, data-independent acquisition mass spectrometry (DIA-MS) has found favor in quantitative proteomics investigations[26] as it provides a facile avenue towards deep proteomic profiling and quantitative accuracy, and experimental reproducibility[27, 28]. Given the routine employment in biomarker investigation[29], disease state stratification[26], and therapeutic monitoring[30], DIA-MS proteomics represents the most facile, reproducible avenue towards representative biomolecular profiling.

Within this work, we validate the utility of library-free DIA-MS to illuminate relevant proteomic dysregulation and reorganization events correlated to distinct prostate cancer phenotypes. Biological triplicate assessment of benign non-tumorigenic (BCaP-NT1), aggressive tumorigenic (BCaP-T10), and aggressive metastatic tumorigenic (BCaP-MT10) cell lines provided confident identification of 95,144 proteotypic peptide sequences mapping to 7,820 proteins. 6,614 proteins were reproducibly identified and quantified in all samples, with 1,242 proteins shown to be dysregulated beyond statistical significance. Clustering analyses revealed these proteins may be grouped into 7 diagnostic patterns, enabling facile assignment of proteins as putative cancer progression biomarkers or phenotypic identifiers. These proteomic fingerprints highlight the systematic disruption and suppression of kinetochore formation and cytoskeletal organization that accompany cancer development; in contrast, other processes such as glutathione production and metabolic flux are shown to be upregulated. We also present a topical comparison of proteomic and transcriptomic analyses, articulating the importance of protein-level measurements in conferring biomolecule abundance and uncovering pertinent reorganization. As shown, the silencing of critical pathways such as interferon signaling and antigen presentation are only

observed at the protein level where transcriptional analyses would indicate no dysregulation. Taken together, this study provides initial reports on the proteomic fingerprints of discrete prostate cancer physiologies. Further validation and employment of the protein panels discussed within may provide a seminal, protein-centric approach towards accurate, minimally invasive prostate cancer identification and stratification.

Experimental Procedures

Experimental Design

The BPH-1-derived Cancer Progression (BCaP) cell model was developed to mimic the molecular and genetic characteristics of prostate cancer progression and may be reviewed elsehwere[19]. To provide a rational evaluation of biomolecular regulation with respect to increasing cancer severity, three discrete phenotypes were selected for analysis: BCaP-NT1 (non-tumorigenic), BCaP-T10 (aggressive tumorigenic), and BCaP-MT10 (aggressive metastatic, tumorigenic) (Figure 7.1). The remainder of the 5 available BCaP phenotypes are to be evaluated in future studies. Three biological replicates of NT1, T10, and MT10 cell lines were grown in parallel under identical conditions. Cells were harvested, washed, and lysed prior to a standard in-solution protein extraction and tryptic digestion workflow. Each of the 9 biological samples was analyzed in technical duplicate via DIA-MS without prefractionation. Data were analyzed using the free software DIA-Neural Network (DIA-NN)[31], providing a means of performing library-free data analysis by creating *in silico* spectral libraries from the reference UniProt human proteome. Results were then processed and visualized using custom code, which is available for public use as described below.

Statistical Rationale

A critical component of DIA data analysis is the validation of true positive precursor assignment. Using an empirical spectral library, precursors are confidently identified prior to library construction, allowing reliable precursor identification in later DIA analyses. As we employed a library-free approach, however, false-discovery rate estimations must be inferred directly. DIA-NN provides built-in false discovery rate (FDR) estimation at both the peptide and protein level, which has been systematically validated[31] and is relied upon in this current work. The final collection of proteins used for quantitation and discussion are those found below a strict 1% protein-level FDR cutoff, identified in every biological sample, and evidenced by at least one proteotypic peptide. All peptides discussed within the manuscript have likewise cleared a 1% peptide-level FDR cutoff. DIA-NN provides automatic protein-level quantitation using the MaxLFQ algorithm, which has been previously validated for quantitative accuracy[32]. In comparison of protein expression levels, the low coefficient of variance (CV) rationalized a $\log_2$ fold change threshold of ±1 from the control group, NT1. Statistical significance was determined via Student's t-test and P-values were subjected Benjamini-Hochberg correction to account for multiple comparisons.

Materials

Acetone (179124), sodium dodecyl sulfate (SDS, 7173C), dithiothreitol (DTT, D9779), and iodoacetic acid (IAA, I6125) were purchased from Millipore Sigma. Water (H_2O, 223623) acetonitrile (ACN, A955), methanol (MeOH, A456), chloroform (C298), formic acid (FA, A117), tris base (BP152), urea (U15), and hydrochloric acid (HCl, A144SI), bicinchoninic acid (BCA) assay (23225), and peptide assay (23275) were purchased from Thermo Scientific. Trypisn (V5113) was purchased from Promega (Madison, WI). RPLC packing materials (4451IP, 4472IP)

were purchased from Osaka Soda Co (Osaka, Japan). Capillary tubing (1068150019) was purchased from PolyMicro. All other reagent manufacturers are listed.

Cell Culture

BCaP cell lines were generated as described previously[19]. Briefly, BPH-1 and urogenital mesenchymal (rUGM) cells were combined and grafted into adult male athymic mice. Without treatment, these cells resulted in a benign graft whereas longitudinal treatment with testosterone and estrodiol (T+E2) resulted in unique tumors mimicking stage-specific cancer characteristics found in humans. The most aggressive tumorigenic line (T10) was regrafted and found to form lymph node metastases without treatment (Figure 7.1). Cells harvested from these were grown and maintained in phenol-free DMEM/Ham's F12 media (Gibco) supplemented with 5% fetal bovine serum (HyClone) and 1% penicillin-streptomycin solution (Gibco). Culture was performed inside T175 culture flasks, incubated at 5% CO2 and 98% humidity, Cells were grown to a confluency of 90%, washed with 1x phosphate-buffered saline (PBS, Cytiva), released with trypsin, and harvested using a cell scraper. Harvested cell count was 2e6 for NT1 and 3e6 for the T10 and MT10 lines. Cell pellets were washed twice using phosphate-buffered saline and stored at −80°C for subsequent processing.

RNA Microarray

Microarray preparation and analysis has been described previously[33]. Briefly, messenger RNA was harvested from cultured cell lines, processed, labeled, and hybridized to Affymetrix HuGene-1.0-st-v1 expression microarrays with 22,211 gene probes. Data were collected by Affymetrix (Santa Clara, CA) and processed to determine differences between NT1, T10, and MT10 lines. Due to sample size, statistics could not be utilized in gene selection.

Protein Extraction and Digestion

Cell pellets were allowed to thaw on ice for 30 minutes and resuspended in 4 volumes of 4% SDS, 50mM Tris-HCl solution prior to lysis via ultrasonication. Homogenate was centrifuged at 14,000rpm (~20,800rcf) and the supernatant was extracted from any remaining cell debris. Protein concentration was estimated via BCA assay and volumes were estimated provide approximate total protein contents. Disulfide bonds were reduced by adding 1.25M DTT in a 1:278 (v:v) ratio and incubated at 55°C for 30 minutes. Solutions were left to stand at ambient conditions until they reached room temperature, followed by 1:10 (v:v) additions of 19mg/mL IAA. Solutions were left to incubate at room temperature for 15 minutes in the dark. Four volumes of ice cold 80% acetone were added to the protein solutions and left to incubate at -20°C overnight. Three sequential precipitate washes were facilitated by centrifugation at 14,000g (~20,800rcf), 4°C for 15 minutes followed by addition of 4 volumes of cold 80% acetone and -20°C incubation for 30 minutes. Upon final acetone removal, samples were left to air dry for 2-3 minutes and then resuspended in 4 volumes of 50mM Tris-HCl, 8M urea, 1x protease and phosphatase inhibitor cocktail. Protein aliquots were digested with trypsin 1:100 for 4 hours at 37°C followed by a second 1:50 trypsin addition and overnight incubation at 37°C. Digests were desalted, and peptide concentration was estimated via peptide assay.

LC-MS/MS

Peptide digest samples were analyzed using an Orbitrap Fusion Lumos mass spectrometer coupled to an Ultimate 3000 nanoflow liquid chromatography system. Analytes were separated on a 15cm-long, custom-packed C18 capillary column, the preparation of which is noted in the supplemental information. Using a flow rate of 350nL/min, H_2O+0.1% FA as buffer A, and ACN+0.1% FA as buffer B, the following 150-minute gradient was used for all analyses: 3% B

minutes 0-18, linear gradient to 35% B at minute 126, 75% B minutes 126.5-130, 95% B minutes 130.5-140, 3% B minutes 140.5-150. A custom DIA-MS method with the following parameters was used for all analyses. MS1 settings: resolution, 120,000; scan range 400-1250m/z; AGC target, 1e6; maximum injection time, 50ms. MS2 settings: precursor mass range, 400-1250m/z; isolation window, 24mz; mass defect, 1.0005; activation type, HCD; collision energy, 30; detector, Orbitrap; resolution, 60,000; scan range, 200-2000; AGC target, 1e5; maximum injection time, 45ms; loop control, N; number of spectra (N), 9. The option to inject ion for all available parallelizable time was set to True.

Data Analysis

MS data files were processed using DIA-NN 1.8. An *in silico* spectral library was created from the reference UniProt human proteome (downloaded February 2022) to facilitate library-free search. Trpysin/P digested peptides with up to 2 missed cleavages were considered; methionine oxidation and n-terminal acetylation were considered variable modifications with up to 4 allowable modifications per peptide. Precursors were confined to those of length 7-50, charge state 1-7, and m/z 300-1400; fragment m/z 50-2,000 were used for identification. MS1 and MS2 accuracy, as well as retention time scan window were determined automatically by DIA-NN, match between runs (MBR) was enabled, protein inference was set to 'Protein Names,' and all other settings left as default. Results were filtered to a q-value=0.01 (1% FDR) prior to analysis. Gene ontology analysis was performed using Metascape[34].

Data and Code Availability

The .raw MS data has been uploaded to the MassIVE repository with the primary accession code MSV000091469 (https://doi.org/doi:10.25345/C5QN5ZN2J). Data may also be found through ProteomeXchange using the identifier PXD040776. Summary tables containing relevant

peptide and protein information have also been provided in a tabular format as Supplemental Data 1-3. All files and custom code needed to reproduce the analyses and figures presented within this manuscript can be accessed at htttps://github.com/lingjunli-research/bcap_dia_profiling.

Results

DIA-MS Profiling Depth and Quantitative Accuracy

Whereas traditional quantitative DIA-MS workflows would employ empirical spectral libraries, decoupling biomolecular discovery from quantitative investigation, our use of a library-free approach implies both aspects must be simultaneously evaluated to ensure correct discrimination of results. The bulk of peptide identifications provided through DIA-NN come through matches to the comprehensive *in silico* library initialized as part of the workflow – namely those with unambiguous evidence and appreciable abundance. However, an advantageous component of DIA-NN is the implementation of match between runs (MBR), allowing previously missed precursor identifications to be rescued based on confident identification in another sample. These two aspects combine to enable highly reproducible peptide and protein identifications across the experimental dataset. This is evidenced in our data as any 3 biological replicates display excellent correlation of precursor masses confidently identified, being evenly distributed across the chosen mass range and highly conserved among the three treatment groups (Figure 7.2**a**). This results in highly reproducible protein group identifications, with >93% of all protein groups being confidently identified in each of three triplicates and similar quantities of proteins identified across the three cell lines (Figure 7.2**b**). Altogether, we observe 95,144 peptide sequences belonging to 7,820 protein groups that may be evaluated for confidence and quantitative accuracy.

It is imperative we establish intra- and inter-sample quantitative reproducibility in order to establish the reliability of protein-level estimations, rationalize any quantitative thresholds, and

avoid incorrect interpretation of protein expression levels. DIA-NN provides direct protein quantitation through implementing the MaxLFQ algorithm[32], which provides robust normalization and recalibration of protein abundance across samples. Comparing to the number of protein groups evidenced in every biological replicate for a given sample, technical duplicate analyses provided reproducible ($R^2 \geq 0.987$), correlational protein abundance estimations of ≥96.8% all protein observations (Figure 7.2**c**). These data highlight the quantitative information will be largely conserved when replicates are averaged together for comparison and suggest the potential for low variance, as discussed below. Filtering data to meet a strict 1% peptide- and protein-level FDR cutoff, our final dataset contained an average 74,847 peptide precursors per sample with an average of 71,406 found to be proteotypic. The presence of at least one proteotypic peptide was a principal requirement during protein filtering, as well as surpassing a 1% protein-level FDR cutoff (Figure 7.2**d**). It is worth noting that FDR filtering alone removed only a small number of proteins from final consideration (Figure 7.2**e**), highlighting the breadth of confident peptide identifications observed for each protein. Proteotypic peptide requirements reduced the final protein collection by ~25% but this ensures any proteins suggested as useful for disease stratification are not mis-identified. Finally, the 6,614 confident protein identifications shared between all biological replicates and each treatment group (Figure 7.2**f,** supplemental data 1) were taken for comparison and discussion.

Prostate Cancer Biomolecular Fingerprinting

To establish rational criteria for consideration of proteomic dysregulation, we evaluated the variance associated from replicate measurements. For all samples, the average coefficient of variance was found to be 5.7-6.0% (median 3.6-3.8%, supplemental Figure S7.1) when using the average and deviation of biological replicate measurements in each sample. This low variance

enabled us to establish a 1-fold change (2x change in abundance, as calculated by MaxLFQ) threshold for dysregulation, further discriminated by corrected statistical significance determination. With the non-tumorigenic (NT1) treatment group serving as the control, our analyses revealed both the aggressive tumorigenic (T10) and aggressive metastatic, tumorigenic (MT10) groups were relatively consistent in the number of upregulated proteins, approximately 420 (Figure 7.3a). T10 displayed a similar number of downregulated proteins, 410, while this quantity increased in the MT10 group by approximately 30% to 540 proteins (Figure 7.3a). In total, our data provided a pool of 1,242 statistically dysregulated proteins that may be further investigated to illuminate potentially significant biomolecular aberrations that signal PCa stage and progression.

Realizing the extent of protein dysregulation was seemingly more advanced in the latter physiology, we anticipated this observation reflected the progressive nature of these BCaP cell lines and highlighted proteome reorganization was conserved or correlated with prostate cancer progression. Pearson correlation analysis initially confirmed this expectation as the NT1 and MT10 phenotypes bore greater similarity to the T10 group than they did to one another (Figure 7.3b, supplemental Figure S7.2). We further investigated the congruency of protein regulation between T10 and MT10 groups and reveal 44% of all dysregulated proteins were similarly altered in both groups (i.e. proteins upregulated in T10 stay upregulated in MT10 and vice versa, supplemental Figure S7.3). These observations are encouraging and cause us to speculate that these proteins are likely to provide relevant markers of advanced PCa or signals of malignancy. Furthermore, because less than half of all significantly dysregulated proteins are well-correlated within the disease groups, we hypothesize a holistic evaluation of the entire 1,242 protein panel would reveal nuanced regulation useful for stratifying these three phenotypes. This hypothesis was confirmed

through hierarchical clustering, which revealed correct separation of all biological groups and reaffirmed phenotypic relationship (Figure 7.3c). We validated this clustering and rearticulated the distinction of disease states through unsupervised dimensional reduction (supplemental Figure S7.4).

Encouraged by the unambiguous grouping of biological replicates within each disease state, we sought to identify the collection of proteins most responsible for ascertaining differences. Our data had so far suggested a correlational – though not always proportional – dysregulation for some proteins in the T10 and MT10 lines (supplemental Figure S7.3). If we permit the consideration of a longitudinal relationship between the three phenotypes analyzed here (i.e., NT1 comes before T10 and T10 comes before MT10), the consistency in protein dysregulation can be considered reflective of putative oncogenes or tumor suppressor genes. Based on this, we hypothesized all remaining dysregulated proteins may be clustered according to their magnitude and direction and then represented longitudinally to reveal putative protein panels useful for biomarker validation and therapeutic targeting. We evaluated this hypothesis by performing row-wise hierarchical clustering of protein expression, manually inspecting cluster thresholds to obtain the most logical protein groupings. Ultimately, the allowed granularity revealed 7 major clusters discussed below.

The dysregulation of proteins within these clusters enable understanding as to which protein populations are useful for disease identification, those that may be investigated for potential biomarkers, and which proteins may serve to suggest phenotype but are not viable as markers of progression. For example, of the clusters displayed in Figure 7.3d, clusters 2 (112 proteins) and 6 (141 proteins) demonstrate progressive downregulation and upregulation in accordance with cancer aggressiveness, respectively; logically, these are topical targets for

biomarker validation or for use in therapeutic monitoring. Clusters 4 (456 proteins) and 7 (302 proteins) display conserved dysregulation in the T10 and MT10 groups, which implies these protein panels may be used as signals of malignancy and can be evaluated as oncogenes or suppressor genes. On the other hand, clusters 3 (54 proteins) and 5 (136 proteins) show distinct protein expression in each of the tested cell lines, suggesting expansion of these populations could prove useful in assigning phenotype. Regardless of the speculated potential of these protein clusters, we cannot infer the biological significance or relevance from dysregulation patterns alone. We therefore further analyzed these proteins to reveal protein communities of interest and to illuminate potential underpinnings of PCa progression.

Identification of Protein Communities Related to PCa Malignancy, Progression and Phenotype

The proteomic reorganization observed within our analyses can be grouped into 7 unique clusters based on the magnitude and direction of protein regulation. However, the extent of dysregulation alone is not sufficient to suggest importance in PCa-specific investigations, and each cluster should therefore be investigated separately to establish biological relevance and ascertain proteins that show promise for future investigation (Figure 7.3d). We surveyed the activity and relatedness of these proteins by submitting each cluster to gene ontology (GO) analysis (Figure 7.3e**,** supplemental data 2) to identify any highly conserved protein communities. Clusters 1 and 2 demonstrate concerted dysregulation of proteins involved in cellular mitosis. Specifically, our data point to disruption of kinetochore assembly, as evidenced by suppression of AURKB, INCENP, KNL1, KNSTRN, NUF2, SPC24, and SPDL1 among others (supplemental data 2). Dysregulation of kinetochore activity is further capitulated by lowered expression of BUB1, CDC20, UBE2C, and UBE2S that are required to form the anaphase promoting complex, as well as CEP55, CEP78, KIF20A, and KIFC1 involved in microtubule association with the kinetochore. We also noted

RHO GPTase family members and interactive components such as ANLN, PKN1, PPP2R5D, and PRC1, suggesting both kinetochore disruption and participation in cytoskeletal reorganization.

Previous studies have established the importance of kinetochore activity in cancer[35], as correct association and participation of these components is critical for correct chromosome migration. Review of investigations into the role of the kinetochore highlight that overexpression of constituent proteins may contribute to cancer progress by driving aneuploidy[35]. Given that our data suggests progressive downregulation of kinetochore-associated proteins, we hypothesize our data is reflective of prolonged mitosis and cellular senescence. Literature supports this hypothesis as senescent tumor cell have been shown to contribute to invasion by providing an immunoprotective environment for non-senescent cells[36]; this theme is further confirmed by our data that show suppressed immunity in malignant cell lines (see information on cluster 4 and RNA-protein comparison).

In addition to these possible drivers of PCa progression, our serve to bring attention to the Ndc80 complex. This complex is known to participate in all aspects of kinetochore function and is chromosome segregation in all eukaryotes[37]. The Ndc80 complex is comprised of four subunits: NDC80, NUF2, SPC24, and SPC25. NUF2 and PSC24 were both shown to be downregulated only in the metastatic MT10 cell line, as highlighted by their position in cluster 1 (figure 3d). We searched for the remaining two subunits in our dataset and reveal that both NDC80 and SPC25 are likewise downregulated only in MT10, though they were just shy of the 1-fold change threshold (NDC80 = -0.92, SPC25= -0.809). The stark contrast of NDC80 regulation in malignant cell lines suggests the depletion or silencing of NDC80 contributes to or accompanies metastasis. We propose functional analyses of metastatic phenotypes within the BCaP model (M1 and MT10, Figure 7.1) may be used to further elucidate metastatic potential NDC80.

Diagnostic cluster 3 presented lesser conservation in biological processes but did display enrichment of epidermis development. Among the small collection of proteins in this cluster, several species such as CASP1[38], CAV1[39], CLDN1[40], COL17A1[41], IL18[42], KRT5[43], PPL[44], and SFN[45] have demonstrated connection to various cancers. Prior to investigation, we anticipated various processes such as epidermal-mesenchymal transition (EMT) or cornified envelope formation to be dysregulated as PCa progresses, meaning the identification of these proteins is not surprising. However, these proteins are marked by higher expression in the NT1 line, suppressed in T10, and then partially corrected in MT10, which subverts our expectation that these species may be disrupted longitudinally. We caution drawing conclusion from a small protein population, though two hypotheses may be constructed. First, the greatest disruption of 'normal' epidermal development may be seen in early PCa phenotypes with some activity trending towards baseline in advanced stages. Further analyses of early BCaP phenotypes (i.e., NT1 and T1) may elucidate or reject this reality. Alternatively, we hypothesize the true aberration in cluster 3 is actually the high prevalence of proteins in NT1. Given that cluster 3 also highlights proteins involved in monocarboxylic acid and that their regulation trajectories resemble that of clusters 6 and 7, we consider this hypothesis quite possible and will be investigated in future studies.

Cluster 4, which is comprised of proteins downregulated in malignant T10 and MT10 phenotypes, was both largest in size and most highly conserved in biological pathway participation. Obviously, cytokine and interferon signaling were the most enriched pathways, highlighting the participation of these process in tumor progression[46-48]. Due to significance of the proteins within these pathways, and due to the observations found in our comparisons of transcript-level data, presentation and discussion of proteins within these processes are largely confined to later sections (see RNA-protein comparison). Beyond cytokine and interferon signaling, actin-

filament processes were shown to be dysregulated, restating the known relationship between cancer and cytoskeletal reorganization[49]. We identified numerous classical actors in these pathways such as actins (ACTG1, ACTN1), keratins (KRT1, KRT10, KRT14, KRT17, KRT80), myosins (MPRIP, MYH14, MYL6, MYO18A, MYO1C), and tropomyosins (TPM1, TPM4), as well as GSN in cluster 2. These data are more in-line with the expected epidermal and cytoskeletal dysregulation associated with PCa, reaffirm the loss of adherence to neighboring cells and tissue contribute to invasion[50, 51]. Given recent support, we propose expansion and validation of this protein community may provide relevant targets for anti-cancer therapy and biomarkers of malignancy.

One significant observation worth noting from cluster 4 is the downregulation of DDX60, DDX60L, and DDX3Y. These three proteins are DEAD-box (DDX) helicase superfamily members, each known to participate in RIG-1-dependent type 1 interferon (IFN) production[52]. There is no consensus on the relationship of DDX60 and DDX60L expression and cancer prognoses, with reports suggesting both up-[53, 54] and downregulation[55, 56] are clinical markers. Regardless, our data reveal a significant downregulation of these proteins and downstream partners, suggesting a compromised immune response in malignant cell lines. DDX3X and DDX3Y, relatives of DDX60, were also quantified in our dataset with proteotypic support. Notably, DDX3Y was significantly downregulated with DDX3X showing normal expression. This observation stood out due to the commentary on the activity of these proteins in PCa. Recent literature points to loss-of-Y (LOY)[57, 58] as a marker of cancer susceptibility, though some opinions prior to these reports suggest caution in this hypothesis[59]. Suppression of Y-chromosome-coded DDX3Y in our dataset suggests general agreement with LOY studies, though we do not attempt to draw a firm connection. More importantly, the redundancy between DDX3Y and DDX3X are

often associated with rescue of function[60-62], implying suppression of DDX3Y leads to higher DDX3X activity in the present study. Increased activity of DDX3X would, however, confirm previous transcriptomic analyses of the BCaP cell lines, which have suggested DDX3 as a marker of metastases[63] and regulator of androgen receptor[64]. Given the importance and recent focus on DDX3, we posit further investigations are mandatory to both confirm the presence and downregulation of DDX3Y presented here, as well as further elucidate the unique role of DDX3X.

Cluster 5 presented the greatest breadth and lowest conservation of enriched pathways. The enrichment of proteins involved in cholesterol synthesis reinforce the known energy and metabolic demands of various cancers[65, 66]. However, we observe a similar phenomenon as that seen in cluster 3 where direction of dysregulation is reversed when moving from T10 to MT10. As well, the proteins within this cluster are shown to be related to microtubule binding and mitosis, which harken back to clusters 1, 2, and 4. This provides further support to our earlier conjecture that the aberration in this cluster stems from altered expression in NT1 and that the changes observed in T10 and MT10 are mimicking those seen in other clusters. For this reason, we speculate the proteins confined to clusters 3 and 5 may demonstrate potential utility in assigning non-tumorigenic phenotypes.

The dysregulation pattern observed in cluster 6 indicates these constituent proteins are those progressively upregulated as PCa progresses. This cluster is arguably the most congruent in biological processes as we observe a clear association between the protein members, metabolic transport, and metabolic flux. Very specifically, we pinpoint glutathione (GSH) production as progressively upregulated in accordance with cancer progression (Figure 7.4). Our data display membrane antiporters SLC7A11/SLCA32 and SLC1A3 are upregulated, providing greater influx of cysteine and aspartate, respectively[67]. Cysteine is a critical component in GSH biosynthesis,

with cysteine-glutamate pathway members GCLC, GCLM, and GSS all being upregulated. SLC1A3 further contributes to GSH biosynthesis by providing aspartate that is converted to glutamate in the TCA cycle by way of GOT1, which is also upregulated. GSH is a key mediator used to eliminate reactive oxidative species (ROS) through glutathione peroxidases (GPX) and our data demonstrate upregulation of GPX1 and upregulation of GPX4, though it did not meet the 1-fold change threshold. Based on this, we posit the increased production of GSH and normal/elevated elimination of ROS inhibits ferroptosis and confers cell survival. To further this point, recent literature has noted the contribution of the glutamine transporter SLSC1A5 to drive GSH production and confer chemotherapy resistance[68]. Beyond these transporters, increased SLC7A11/SLCA32 is also shown to promote invasion and metastases through activation of the PIK3-AKT-mTOR pathway[67]; PIK3 is shown to be upregulated in our data with AKT and mTOR showing normal regulation (Figure 7.4). Given the conservation of activity and increasing attention in translational studies, the proteins grouped into this cluster point towards clinical relevance of GSH production and suggest potential utility of SLC-focused therapies.

Finally, proteins constitutively upregulated in malignant phenotypes are grouped into cluster 7. As expected, these proteins are largely associated with glycogen and glucose metabolism (G6PD, PFKM, PGM1), amino acid biosynthesis (BCAT1, FAH, MTRR, SLC1A3, PSAT1, PSPH) as well as restating glutathione production mentioned above (PYCR1, PYCR3. GSS, GCLM, GSTM4), among others (supplemental data 2). There is abundant evidence implicating these processes within cancer[69-72] as they are critical for energy production and survival. In addition, this cluster also revealed proteins involved in in EGFR signaling (supplemental data 2), implicating this pathway in invasion and metastases, as noted before[73]. Notably, our data indicate higher expression of STAT6, an emerging target of interest in cancer therapies due to its role in

mediate immune respose[74]. This upregulation of STAT6 is more compelling, considering the suppression of other STAT family members discussed below (see RNA-protein comparison). Given the topical nature of proteins in this pathway, we further articulate the potential use for these proteins in targeted therapies or as markers of malignancy. Likewise, all 7 clusters presented here may be considered individually or in concert in the effort to form effective, protein-based strategies for disease assessment. As we have so far demonstrated the ability of these clusters to provide relevant, actionable insight into PCa progression, it is our position these clusters should be further expanded and validated through repeat measurements and then explored longitudinally to reveal meaningful prognostic markers.

Transcript-Protein Inconsistency Reveals Suppressed Pathways

Transcriptomic analyses have long been the standard primary measurement in the effort to uncover specific biomolecular alterations specific to human disease. While the sensitivity and efficiency of these methodologies has been widely reported[20-22], a fundamental limitation is the lack of a quantitative relationship between RNA and the proteins for which they code[23-25]. Because a core argument in this work is that protein identification and quantitation enable a more representative view of PCa progression, we sought to evaluate the agreement between transcriptomic and proteomic observations and illuminate potential dysregulation events that may be hidden at the transcript level.

A preconstructed RNA microarray was used to screen NT1, T10, and MT10 cell lines (see methods), with ratios calculated against the control NT1 group and log2 transformed to determine fold change. Coding genes were mapped to all proteins identified in the proteomics dataset and joined to the microarray data. We assume a 1-fold change threshold for microarray data, though it is important to consider the differences in measurement between RNA and protein datasets;

contrasts discussed here are merely qualitative. Overall comparison between these datasets revealed a weak positive correlation for both the T10 (R^2=0.48) and MT10 (R^2=0.58) groups, reflecting the general agreement between transcript evidence and protein expression, though the most obvious similarities are seen at the extremes. Using MT10 as a representative (Figure 7.5, supplemental data 3), numerous proteins are shown to be downregulated at both the transcript and protein levels. Among these, we observe downregulation of the tumor suppressors BCAM[75], FSTL1[76], DPS[77], DSC3[78], and ST14[79] rearticulating their known roles pinpointing relevant targets contributing to invasion and metastases. However, we also note the suppression of some genes/proteins that have previously been presented as poor prognosis markers, such as KRT10[80], KRT14[81], CPA4[82], and CALD1[83]. The suppression of these markers will be confirmed in later studies, though our findings may simply highlight cancer heterogeneity and non-analogous presentation of prognostic markers.

Turning to those proteins upregulated at both the transcript and protein levels, we observe increased levels of known cancer-related actors such as BCAT1[84], AGR2[85] and INA[86], each of which has been suggested as a putative therapeutic target. Opposing these, STOM is suggested as a tumor suppressor[87], indicating the upregulation seen in our datasets may suggest feedback activation or a mechanism promoting survival. In addition to these proteins, we also observe overexpression of hypoxia upregulated 1 (HYOU1), which brings into focus two areas of consideration. First, HYOU1 is known to promote proliferation, invasion and metastases through the PI3K-AKT-mTOR pathway[88], a signaling cascade also promoted through SLC7A11 activity. Our proteomic analyses revealed significant upregulation of SLC7A11 and PI3K, while AKT, mTOR, and RPTOR expression was normal (Figure 7.4, supplemental data 1). The conserved upregulation of this protein community highlights a putative mechanism driving PCa progression

in this BCaP cell model. However, STOM expression has been found to be negatively correlated with AKT activation[87], allowing us to hypothesize that either STOM upregulation or PI3K-AKT activation is inspired through feedback mechanisms.

The second, larger observation gleaned through HYOU1 upregulation is that malignant lines in the BCaP model adopt a permanent hypoxic phenotype. Though it is obvious that oxygen starvation leads to increased expression of HYOU1, we can further confirm hypoxia through the significant downregulation of IVL, an oxygen regulated product[89]. HYOU1 is known to inhibit apoptosis though downregulation of IFN-α/β[88], causing us to question if disruption of this pathway could be observed in our analyses. Though IFN-α/β were not quantified in our analyses, our search did reveal the suppression IFIT1, IFIT2, and IFIT3. Most strikingly, these proteins were only shown to be dysregulated at the protein level, whereas transcriptional analysis indicate no perturbation in expression (Figure 7.5). Intrigued, we searched for all related proteins and interacting partners, revealing nearly ubiquitous downregulation of proteins within the JAK/STAT pathway (Figure 7.6). Both STAT1 and STAT2 were shown to be downregulated, suggesting lowered activity of their activated complex and subsequent reduced activation of interferon response element, ISRE. Examining the multitude of downstream IFN-induced genes, we observe unambiguous downregulation of ubiquitin modifiers (ISG15, ISG20), oligoadenylate synthases (OAS3, OASL), PSMB8, IFIT proteins (IFIT1, IFIT2, IFIT3, IFIT5, IFITM3, IFI35), and HLA class I (HLA-A, HLA-B, HLA-C, HLA-E, HLA-F).

Diagnostic cluster 4, mentioned above (Figure 7.3d), displayed the greatest enrichment of any single biological processes, with cytokine and interferon signaling being the most significant terms (Figure 7.3e). The conserved downregulation of the JAK/STAT proteins is the primary driver of that enrichment and highlights the potentially significant roles of this signaling pathway

in PCa progression. Existing literature[90] has noted that hypoxia drives JAK/STAT suppression through reduced degradation of HIF1a, which may be stabilized through interaction with HSP90[91] and proceeds to drive JAK/STAT inhibition[92]. However, recent reports have suggested the silencing of JAK/STAT is hypoxia specific and only partially inspired through HIF1a mechanisms[93]. HIF1a was not identified or quantified in our proteomics dataset and we therefore cannot confirm either claim. However, since we cannot ascertain any accumulation of HIF1a in the advanced, malignant cell lines, our data suggests agreement with the conjecture that JAK/STAT silencing is mediated by hypoxia alone rather than by HIF1a. These claims must be investigated separately. These biological questions notwithstanding, our data reveal a demonstrable contribution to PCa invasion and metastases from oxygen regulated signaling pathways. We provide evidence that both interferon signaling and antigen presentation are silenced in malignant phenotypes, indicating either solitary or joint contribution to progression. We suggest these are critical drivers of PCa development and advocate for dedicated investigation. The BCaP cell model that revealed these truths makes for a compelling system in which to investigate the effects, allowing not only facile knock in/out of relevant gene targets, but also a logical comparison of the physiological traits inspired after targeted therapy.

Discussion

Library-free data-independent acquisition mass spectrometry is a powerful tool in the effort to uncover biomolecular changes in health and human disease. Here we present the ability to provide reproducible peptide and protein identification and rigorous protein-level abundance estimates. Nevertheless, it may be argued that direct analysis of proteolytic mixtures may never replace the profiling depth enabled found when utilizing pre-fractionation and building an empirical spectral library specific. Certainly, the utilization of newer DIA-based technologies[94] or

the use of instrumentation with higher acquisition speed[95] may have improved the extent of proteomic information obtained within this study. However, quantification of 6,614 proteins across 9 biological samples brings merit and rigor to this investigation, providing a meaningful collection of biomolecules useful for phenotypic comparison and follow-up study. We believe development of this work should focus on two primary areas. First, the expansion of these proteomic comparisons both in depth of coverage and the BCaP phenotypes compared. These investigations will provide support or greater context for the observations made here and enable a more nuanced understanding of PCa physiology. The second area of development is the inclusion or targeted assessment of post-translational modifications (PTMs). As PTMs are a critical component of disease progression, they are worthwhile target in future investigation and may provide more reflective protein quantitation when considered alongside the unmodified peptides shown here. Furthermore, our analyses highlighted JAK/STAT silencing and reaffirmed its relevance to cancer progression. Protein phosphorylation is the underlying currency of this pathway and must be analyzed independently to better understand dysregulation of JAK/STAT in prostate cancer.

A core tenant of this work was to provide distinct biomolecular fingerprints of PCa progression. Our analyses of three BCaP cell lines, which mimic the genetic, phenotypic, and molecular characteristics of human cancer, allowed us to identify 1,242 dysregulated proteins that highlight molecular alterations as PCa advances. These significant proteins could easily distinguish all three tested phenotypes from one another and correctly identified the expected phenotypic relationships. More than this, because various protein communities display a high degree of relation to one another in their direction and magnitude of dysregulation, we easily grouped these proteins into 7 diagnostic clusters. Viewing these clusters longitudinally allows us to understand which protein cohorts are useful as progressive markers and those useful to assign

malignancy (Figure 7.3d). Clusters 4 and 7 contain those proteins consistently up- or downregulated in malignant phenotypes and are useful for severity assignment. However, because we cannot determine at what point during PCa progression these proteins become dysregulated, inclusion of intermediate BCaP phenotypes becomes necessary. Though clusters 4 and 7 present the most logical oncogenic or suppressor targets, clusters 2 and 6 may be considered more immediately interesting. We of course anticipate the number of proteins within these clusters to grow in subsequent, higher-throughput analyses, but this initial report already suggests the potential significance of understanding mitotic disruption and glutathione production in cancer progression.

Finally, our work points towards critical considerations in biomolecular investigation. Though decades of innovation have bolstered the sensitivity and throughput of RNA-based analyses, total reliance on transcript-level information may prove limiting. We agree, on principle, that transcription analyses provide a high-level understanding of protein presence. However, our data) reinforces the understanding that protein abundance cannot be directly inferred at the transcript level, mandating proteomic analyses. For this reason, we fervently suggest that both the entire BCaP cell model, and other future cancer models must employ comprehensive RNA and proteomic analysis. In doing so we will provide more comprehensive, rational presentation of any significant biomolecular alterations observed and provide insight into meaningful experimental design for emerging research.

Summary

In summary, we provide a preliminary investigation of the proteomic perturbations related to and responsible for prostate cancer progression. The novel, progressive BCaP cell model is an invaluable tool for gleaning insight into this disease as it not only provides an avenue for

biomolecular discovery, but it also allows the confident assignment of biomolecular changes that accompany discrete PCa phenotypes. Here we demonstrate the unique proteomic fingerprints of non-tumorigenic, aggressive tumorigenic, and aggressive metastatic, tumorigenic cell lines are immediately applicable for disease stratification. As well, these proteomic fingerprints can be dissected and represented longitudinally to understand those proteins and protein communities that present potential for biomarker validation, malignancy assignment, and severity assessment. In addition to kinetochore disruption and cytoskeletal reorganization highlighted in our findings, our data reveal the silencing of the JAK/STAT signaling cascade is a marker of PCa malignancy, possibly due to disruption of immune activity and silenced interferon signaling. As well, we highlight PCa progression is associated with a possible overproduction of GSH that fails to inspire ferroptosis. Further investigation may reveal the clinical significance of these pathways and suggest potential therapeutic targets. Regardless of outcome, this work validates a suitable methodology for investigating PCa progression that will be expanded to provide more comprehensive understanding of relevant biomolecular changes visible in human cancer.

Acknowledgements

Support for this research is provided in part by the National Institutes of Health (NIH) grants RF1 AG052324 (LL), R01 DK071801 (LL), and U54DK104310 (WAR, LL). The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. LL acknowledges the National Science Foundation funding support (CHE- 2108223 and IOS-2010789), NIH grant support R21AG065728, S10OD028473, and S10OD025084 as well as a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Figures

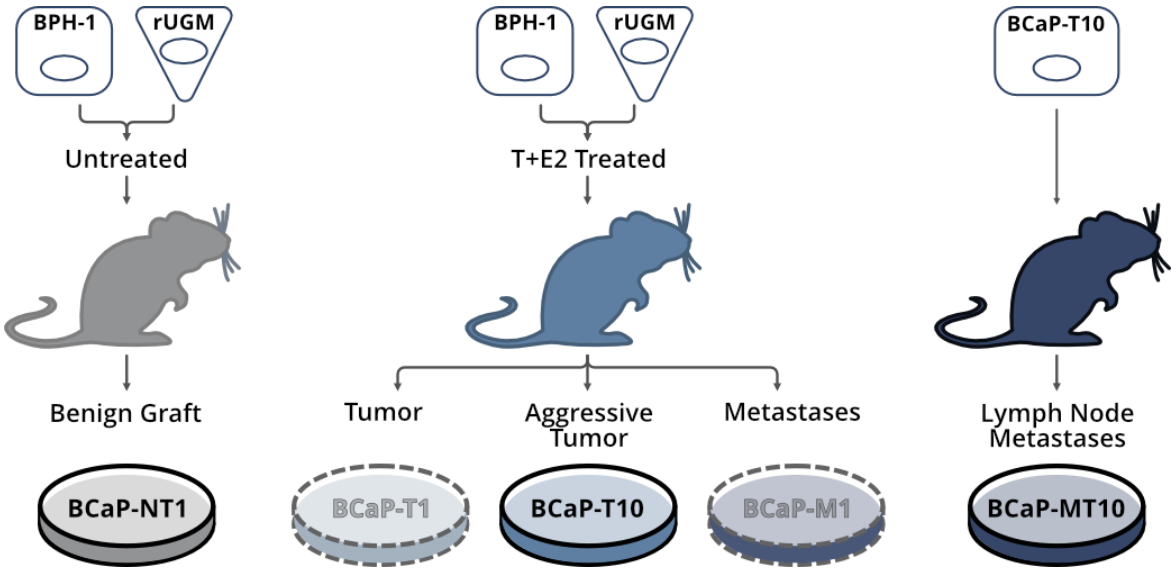


Figure 7.1 Graphical representation of the BCaP cell model.

BPH-1 and rUGM cells were grafted into the renal capsule of adult male athymic mice. Left untreated, these grafts resulted in a non-tumorigenic (NT1) phenotype. After treatment with normal circulating levels of testosterone and estrodiol grafts were found to be tumorigenic (T1) at 2 months, aggressive metastatic at 4 months (T10) and metastasize at 4 months (M1). Metastatic cells were regrafted as before and left untreated to form an aggressive metastatic tumorigenic (MT10) line. T1 and M1 lines are transparent because they are not used or discussed in this work.

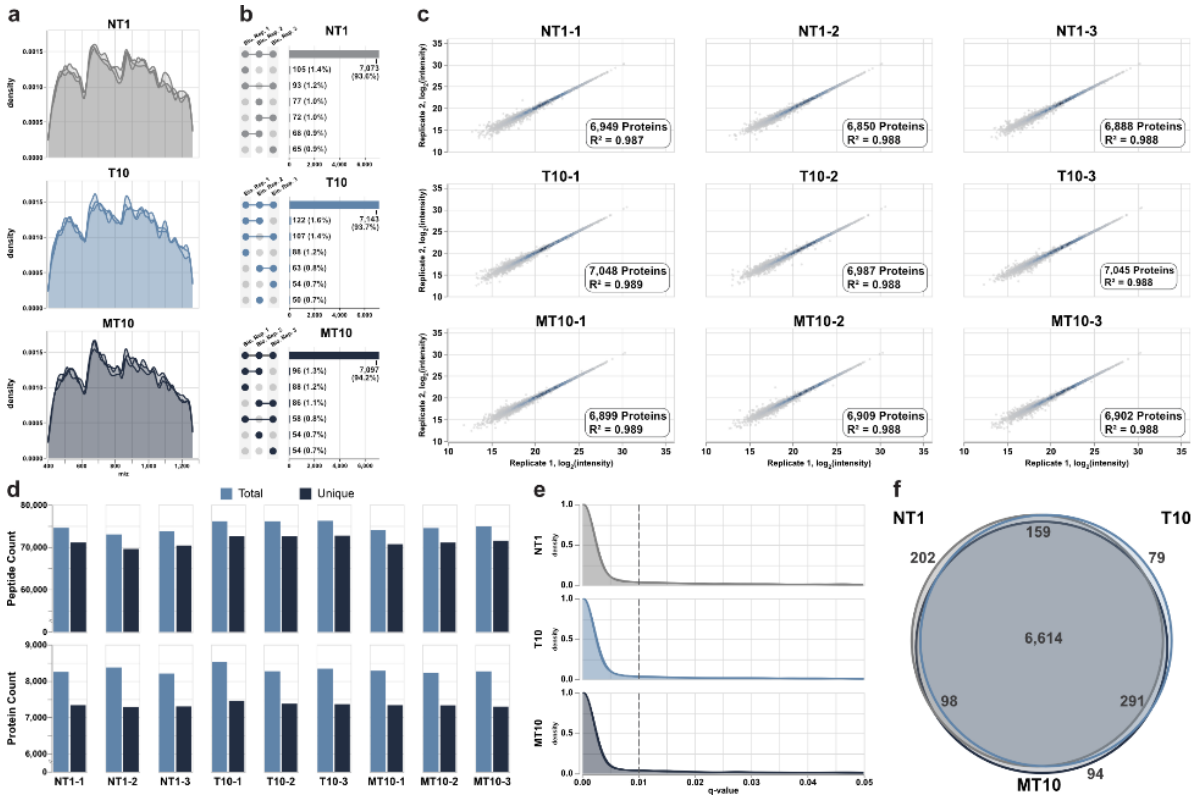


Figure 7.2 Reproducibility, accuracy, and quality of library-free DIA-NN results.

a) Density plots representing the conserved distribution of peptide m/z values. Each biological replicate is a unique line on a given plot. b) Upset plot displaying the overlap of protein identifications in each BCaP line. Nodes represent a single biological replicate; bars represent proteins found in the highlighted nodes. c) Scatter plot comparing MaxLFQ protein abundances between technical replicates. Dot color represents kernel density estimate. d) Quantity of peptides (top) and proteins (bottom) identified in each sample. Unique peptides are non-redundant modified sequences; unique proteins are proteins with proteotypic evidence. e) density plot displaying the protein-level FDR, as estimated via DIA-NN. f) Overlap of final proteins eligible for quantitative comparison; only the 6,614 shared proteins are used for final discussion.

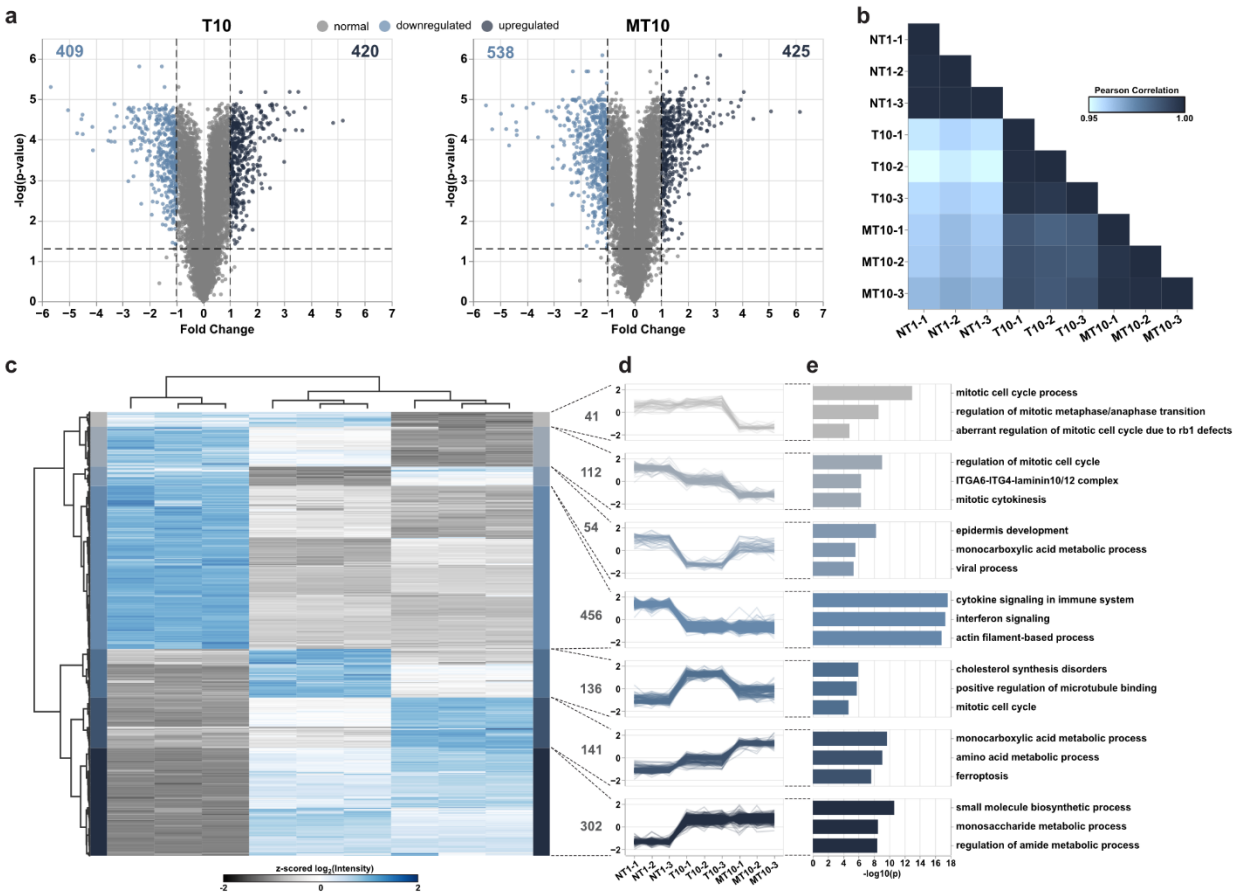


Figure 7.3 Proteomic fingerprints obtained through quantitative DIA-MS analyses.

a) Volcano plots displaying proteins shown to be dysregulated beyond statistical significance (P≤0.05, Benjamani-Hochberg correction applied). b) Pearson correlation of protein abundances between all tested samples. c) Column- and row-wise hierarchical clustering of 1,242 dysregulated proteins. Intensities have been normalized using z-score. d) 6 diagnostic patterns revealed when proteins are grouped together according to magnitude and direction of dysregulation. e) GO analyses of proteins contained within their respective diagnostic cluster.

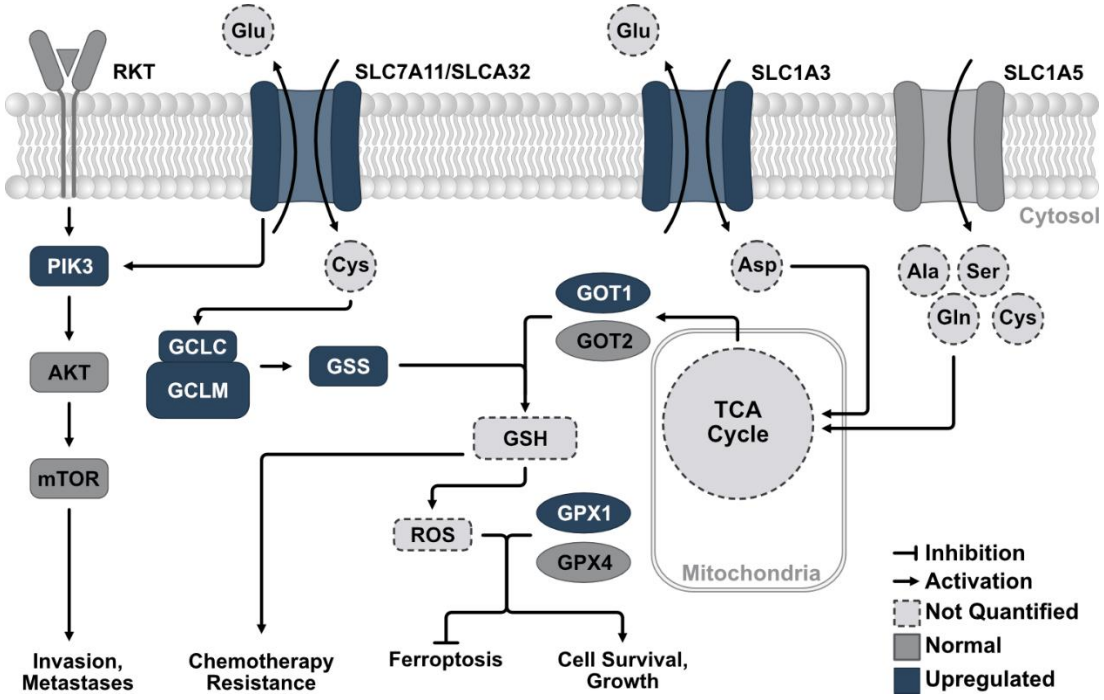


Figure 7.4 Dysregulation of GSH production pathway.

Proteins and small molecules not quantified in this study are represented with light grey backgrounds and dashed outlines. Proteins not significantly dysregulated in this study are dark grey and proteins upregulated are dark blue. Downstream effects of dysregulation (i.e. invasion, metastases) are discussed elsewhere[67].

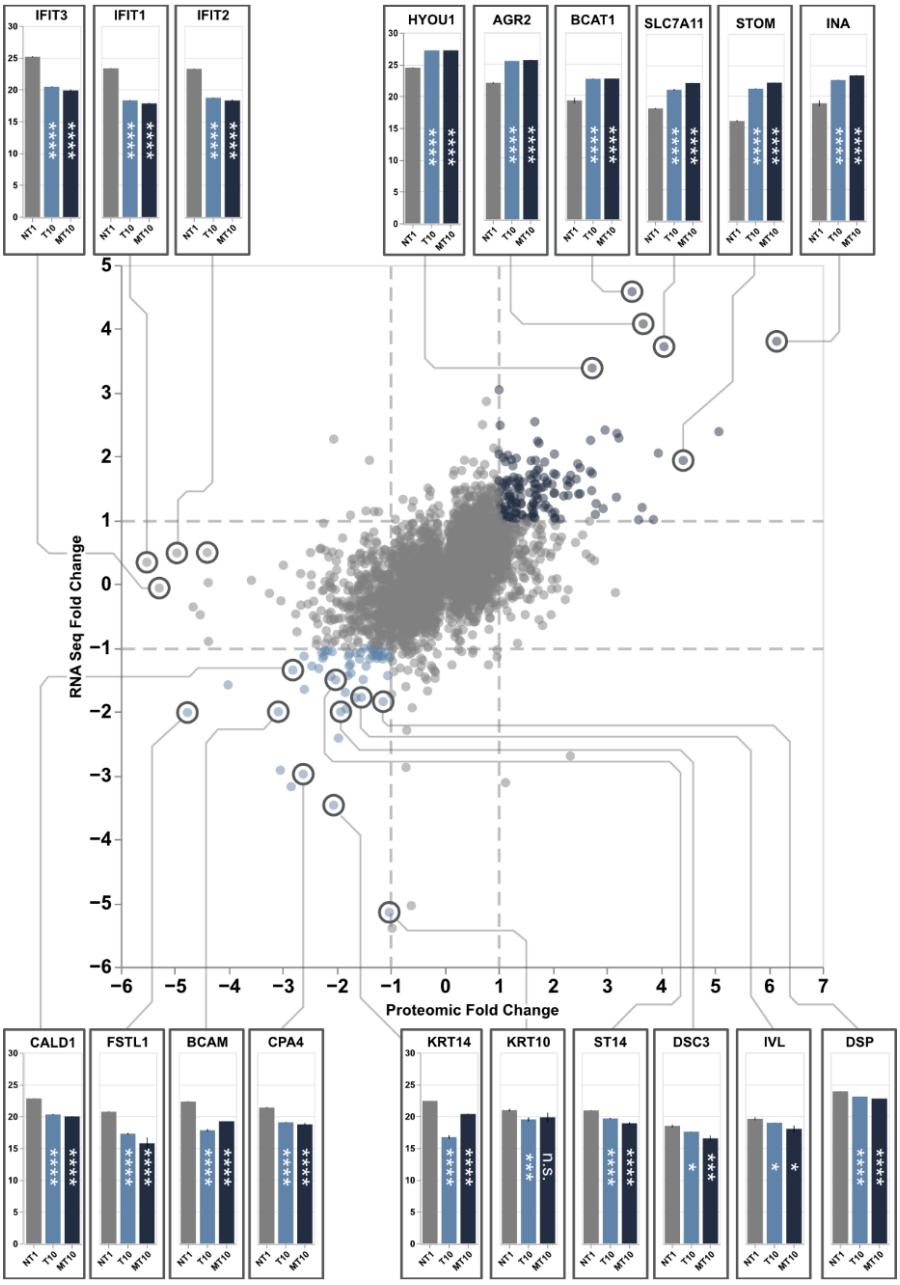


Figure 7.5 Comparison of MT10 RNA and protein biomolecular profiles.

Scatter plot presents comparison of dysregulation determined through microarray (y-axis) and proteomic analyses (x-axis). Bar charts are called out for individual proteins and display the $\log_2$ abundances for NT1, T10, and MT10 cell lines. *P≤0.05; **P≤0.01; ***P≤0.005; ****P≤0.001. All P-values have been corrected via Benjamini-Hochberg correction.

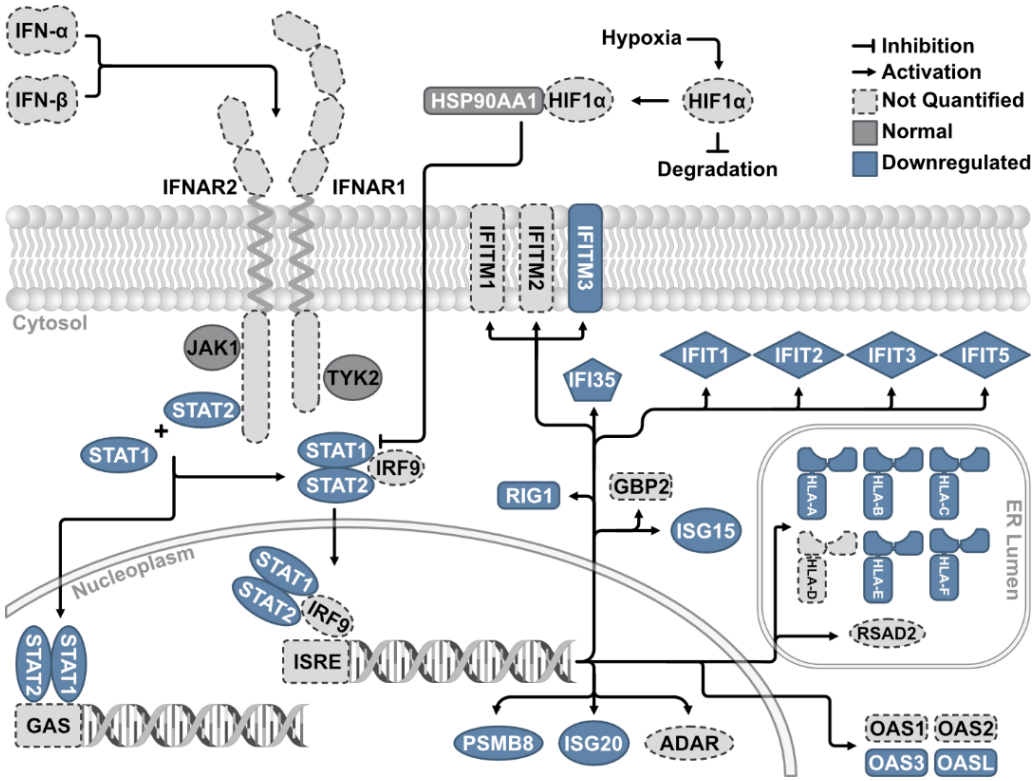


Figure 7.6 Hypoxia-driven suppression of the JAK/STAT pathway.

Proteins not quantified in this study are represented by light grey background and dashed outlines. Proteins not dysregulated are dark grey and proteins downregulated are light blue.

References

1. Berenguer, C. V.; Pereira, F.; Câmara, J. S.; Pereira, J. A. M., Underlying Features of Prostate Cancer—Statistics, Risk Factors, and Emerging Methods for Its Diagnosis. *Current Oncology* **2023,** *30* (2), 2300-2321.

2. Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F., Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* **2021,** *71* (3), 209-249.

3. Bleyer, A.; Spreafico, F.; Barr, R., Prostate cancer in young men: An emerging young adult and older adolescent challenge. *Cancer* **2020,** *126* (1), 46-57.

4. Soares, S. C. M.; de Camargo Cancela, M.; Migowski, A.; de Souza, D. L. B., Digital rectal examination and its associated factors in the early detection of prostate cancer: a cross-sectional population-based study. *BMC Public Health* **2019,** *19* (1), 1573.

5. BEEMSTERBOER, P. M. M.; KONING, H. J. d.; KRANSE, R.; TRIENEKENS, P. H.; MAAS, P. J. v. d.; SCHRÖDER, F. H., PROSTATE SPECIFIC ANTIGEN TESTING AND DIGITAL RECTAL EXAMINATION BEFORE AND DURING A RANDOMIZED TRIAL OF SCREENING FOR PROSTATE CANCER: EUROPEAN RANDOMIZED STUDY OF SCREENING FOR PROSTATE CANCER, ROTTERDAM. *Journal of Urology* **2000,** *164* (4), 1216-1220.

6. Gosselaar, C.; Kranse, R.; Roobol, M. J.; Roemeling, S.; Schröder, F. H., The interobserver variability of digital rectal examination in a large randomized trial for the screening of prostate cancer. *The Prostate* **2008,** *68* (9), 985-993.

7. Borkenhagen, J. F.; Eastwood, D.; Kilari, D.; See, W. A.; Van Wickle, J. D.; Lawton, C. A.; Hall, W. A., Digital Rectal Examination Remains a Key Prognostic Tool for Prostate Cancer: A National Cancer Database Review. *Journal of the National Comprehensive Cancer Network J Natl Compr Canc Netw* **2019,** *17* (7), 829-837.

8. Wang, W.; Guo, Y.; Zhang, D.; Tian, Y.; Zhang, X., The prevalence of benign prostatic hyperplasia in mainland China: evidence from epidemiological surveys. *Scientific Reports* **2015,** *5* (1), 13546.

9. Awedew, A. F.; Han, H.; Abbasi, B.; Abbasi-Kangevari, M.; Ahmed, M. B.; Almidani, O.; Amini, E.; Arabloo, J.; Argaw, A. M.; Athari, S. S.; Atlaw, D.; Banach, M.; Barrow, A.; Bhagavathula, A. S.; Bhojaraja, V. S.; Bikbov, B.; Bodicha, B. B. A.; Butt, N. S.; Caetano dos Santos, F. L.; Dadras, O.; Dai, X.; Doan, L. P.; Eftekharzadeh, S.; Fatehizadeh, A.; Garg, T.; Gebremeskel, T. G.; Getachew, M. E.; Ghamari, S.-H.; Gilani, S. A.; Golechha, M.; Gupta, V. B.; Gupta, V. K.; Hay, S. I.; Hosseini, M.-S.; Hosseinzadeh, M.; Humayun, A.; Ilic, I. M.; Ilic, M. D.; Ismail, N. E.; Jakovljevic, M.;

Jayaram, S.; Jazayeri, S. B.; Jema, A. T.; Kabir, A.; Karaye, I. M.; Khader, Y. S.; Khan, E. A.; Landires, I.; Lee, S.-w.; Lee, S. W. H.; Lim, S. S.; Lobo, S. W.; Majeed, A.; Malekpour, M.-R.; Malih, N.; Malik, A. A.; Mehrabi Nasab, E.; Mestrovic, T.; Michalek, I. M.; Mihrtie, G. N.; Mirza-Aghazadeh-Attari, M.; Misganaw, A. T.; Mokdad, A. H.; Molokhia, M.; Murray, C. J. L.; Narasimha Swamy, S.; Nguyen, S. H.; Nowroozi, A.; Nuñez-Samudio, V.; Owolabi, M. O.; Pawar, S.; Perico, N.; Rawaf, D. L.; Rawaf, S.; Rawassizadeh, R.; Remuzzi, G.; Sahebkar, A.; Sampath, C.; Shetty, J. K.; Sibhat, M. M.; Singh, J. A.; Tan, K.-K.; Temesgen, G.; Tolani, M. A.; Tovani-Palone, M. R.; Valadan Tahbaz, S.; Valizadeh, R.; Vo, B.; Vu, L. G.; Yang, L.; Yazdanpanah, F.; Yigit, A.; Yiğit, V.; Yunusa, I.; Zahir, M.; Vos, T.; Dirac, M. A., The global, regional, and national burden of benign prostatic hyperplasia in 204 countries and territories from 2000 to 2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Healthy Longevity* **2022,** *3* (11), e754-e776.

10. Devlin, C. M.; Simms, M. S.; Maitland, N. J., Benign prostatic hyperplasia – what do we know? *BJU International* **2021,** *127* (4), 389-399.

11. Cheng, L.; Montironi, R.; Bostwick, D. G.; Lopez-Beltran, A.; Berney, D. M., Staging of prostate cancer. *Histopathology* **2012,** *60* (1), 87-117.

12. Nagpal, K.; Foote, D.; Liu, Y.; Chen, P.-H. C.; Wulczyn, E.; Tan, F.; Olson, N.; Smith, J. L.; Mohtashamian, A.; Wren, J. H.; Corrado, G. S.; MacDonald, R.; Peng, L. H.; Amin, M. B.; Evans, A. J.; Sangoi, A. R.; Mermel, C. H.; Hipp, J. D.; Stumpe, M. C., Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *npj Digital Medicine* **2019,** *2* (1), 48.

13. Pérez-Ibave, D. C.; Burciaga-Flores, C. H.; Elizondo-Riojas, M., Prostate-specific antigen (PSA) as a possible biomarker in non-prostatic cancer: A review. *Cancer Epidemiol* **2018,** *54*, 48-55.

14. Saini, S., PSA and beyond: alternative prostate cancer biomarkers. *Cellular Oncology* **2016,** *39* (2), 97-106.

15. Merriel, S. W. D.; Pocock, L.; Gilbert, E.; Creavin, S.; Walter, F. M.; Spencer, A.; Hamilton, W., Systematic review and meta-analysis of the diagnostic accuracy of prostate-specific antigen (PSA) for the detection of prostate cancer in symptomatic patients. *BMC Medicine* **2022,** *20* (1), 54.

16. Hoffman, R. M.; Gilliland, F. D.; Adams-Cameron, M.; Hunt, W. C.; Key, C. R., Prostate-specific antigen testing accuracy in community practice. *BMC Family Practice* **2002,** *3* (1), 19.

17. Lilja, H.; Ulmert, D.; Vickers, A. J., Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. *Nature Reviews Cancer* **2008,** *8* (4), 268-278.

18. Wu, X.; Gong, S.; Roy-Burman, P.; Lee, P.; Culig, Z., Current mouse and cell models in prostate cancer research. *Endocr Relat Cancer* **2013,** *20* (4), R155-70.

19. Liu, T. T.; Ewald, J. A.; Ricke, E. A.; Bell, R.; Collins, C.; Ricke, W. A., Modeling human prostate cancer progression in vitro. *Carcinogenesis* **2018,** *40* (7), 893-902.

20. Bassiouni, R.; Gibbs, L. D.; Craig, D. W.; Carpten, J. D.; McEachron, T. A., Applicability of spatial transcriptional profiling to cancer research. *Molecular Cell* **2021,** *81* (8), 1631-1639.

21. Stark, R.; Grzelak, M.; Hadfield, J., RNA sequencing: the teenage years. *Nature Reviews Genetics* **2019,** *20* (11), 631-656.

22. Zappia, L.; Theis, F. J., Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape. *Genome Biology* **2021,** *22* (1), 301.

23. Bauernfeind, A. L.; Babbitt, C. C., The predictive nature of transcript expression levels on protein expression in adult human brain. *BMC Genomics* **2017,** *18* (1), 322.

24. Liu, Y.; Beyer, A.; Aebersold, R., On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **2016,** *165* (3), 535-550.

25. Koussounadis, A.; Langdon, S. P.; Um, I. H.; Harrison, D. J.; Smith, V. A., Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Scientific Reports* **2015,** *5* (1), 10775.

26. Miles, H. N.; Delafield, D. G.; Li, L., Recent developments and applications of quantitative proteomics strategies for high-throughput biomolecular analyses in cancer research. *RSC Chemical Biology* **2021**.

27. Barkovits, K.; Pacharra, S.; Pfeiffer, K.; Steinbach, S.; Eisenacher, M.; Marcus, K.; Uszkoreit, J., Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition. *Mol Cell Proteomics* **2020,** *19* (1), 181-197.

28. Fröhlich, K.; Brombacher, E.; Fahrner, M.; Vogele, D.; Kook, L.; Pinter, N.; Bronsert, P.; Timme-Bronsert, S.; Schmidt, A.; Bärenfaller, K.; Kreutz, C.; Schilling, O., Benchmarking of analysis strategies for data-independent acquisition proteomics using a large-scale dataset comprising inter-patient heterogeneity. *Nature Communications* **2022,** *13* (1), 2622.

29. Sajic, T.; Liu, Y.; Aebersold, R., Using data-independent, high-resolution mass spectrometry in protein biomarker research: Perspectives and clinical applications. *PROTEOMICS – Clinical Applications* **2015,** *9* (3-4), 307-321.

30. Boys, E. L.; Liu, J.; Robinson, P. J.; Reddel, R. R., Clinical applications of mass spectrometry-based proteomics in cancer: Where are we? *PROTEOMICS n/a* (n/a), 2200238.

31. Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M., DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods* **2020,** *17* (1), 41-44.

32. Cox, J.; Hein, M. Y.; Luber, C. A.; Paron, I.; Nagaraj, N.; Mann, M., Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **2014,** *13* (9), 2513-26.

33. Liu, T. T.; Ewald, J. A.; Ricke, E. A.; Bell, R.; Collins, C.; Ricke, W. A., Modeling human prostate cancer progression in vitro. *Carcinogenesis* **2019,** *40* (7), 893-902.

34. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A. H.; Tanaseichuk, O.; Benner, C.; Chanda, S. K., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* **2019,** *10* (1), 1523.

35. Yuen, K. W. Y.; Montpetit, B.; Hieter, P., The kinetochore and cancer: what's the connection? *Current Opinion in Cell Biology* **2005,** *17* (6), 576-582.

36. Park, S. S.; Choi, Y. W.; Kim, J.-H.; Kim, H. S.; Park, T. J., Senescent tumor cells: an overlooked adversary in the battle against cancer. *Experimental & Molecular Medicine* **2021,** *53* (12), 1834-1841.

37. Tooley, J.; Stukenberg, P. T., The Ndc80 complex: integrating the kinetochore's many movements. *Chromosome Res* **2011,** *19* (3), 377-91.

38. Sun, Y.; Guo, Y., Expression of Caspase-1 in breast cancer tissues and its effects on cell proliferation, apoptosis and invasion. *Oncol Lett* **2018,** *15* (5), 6431-6435.

39. Qian, X. L.; Pan, Y. H.; Huang, Q. Y.; Shi, Y. B.; Huang, Q. Y.; Hu, Z. Z.; Xiong, L. X., Caveolin-1: a multifaceted driver of breast cancer progression and its application in clinical treatment. *Onco Targets Ther* **2019,** *12*, 1539-1552.

40. Bhat, A. A.; Syed, N.; Therachiyil, L.; Nisar, S.; Hashem, S.; Macha, M. A.; Yadav, S. K.; Krishnankutty, R.; Muralitharan, S.; Al-Naemi, H.; Bagga, P.; Reddy, R.; Dhawan, P.; Akobeng, A.; Uddin, S.; Frenneaux, M. P.; El-Rifai, W.; Haris, M., Claudin-1, A Double-Edged Sword in Cancer. *Int J Mol Sci* **2020,** *21* (2).

41. Jones, V. A.; Patel, P. M.; Gibson, F. T.; Cordova, A.; Amber, K. T., The Role of Collagen XVII in Cancer: Squamous Cell Carcinoma and Beyond. *Front Oncol* **2020,** *10*, 352.

42. Feng, X.; Zhang, Z.; Sun, P.; Song, G.; Wang, L.; Sun, Z.; Yuan, N.; Wang, Q.; Lun, L., Interleukin-18 Is a Prognostic Marker and Plays a Tumor Suppressive Role in Colon Cancer. *Dis Markers* **2020,** *2020*, 6439614.

43. Ricciardelli, C.; Lokman, N. A.; Pyragius, C. E.; Ween, M. P.; Macpherson, A. M.; Ruszkiewicz, A.; Hoffmann, P.; Oehler, M. K., Keratin 5 overexpression is associated with serous ovarian cancer recurrence and chemotherapy resistance. *Oncotarget* **2017,** *8* (11), 17819-17832.

44. Li, X.; Zhang, G.; Wang, Y.; Elgehama, A.; Sun, Y.; Li, L.; Gu, Y.; Guo, W.; Xu, Q., Loss of periplakin expression is associated with the tumorigenesis of colorectal carcinoma. *Biomed Pharmacother* **2017,** *87*, 366-374.

45. Fan, X.; Cui, L.; Zeng, Y.; Song, W.; Gaur, U.; Yang, M., 14-3-3 Proteins Are on the Crossroads of Cancer, Aging, and Age-Related Neurodegenerative Disease. *Int J Mol Sci* **2019,** *20* (14).

46. Lee, M.; Rhee, I., Cytokine Signaling in Tumor Progression. *Immune Netw* **2017,** *17* (4), 214-227.

47. Kartikasari, A. E. R.; Huertas, C. S.; Mitchell, A.; Plebanski, M., Tumor-Induced Inflammatory Cytokines and the Emerging Diagnostic Devices for Cancer Detection and Prognosis. *Front Oncol* **2021,** *11*.

48. Dranoff, G., Cytokines in cancer pathogenesis and cancer therapy. *Nature Reviews Cancer* **2004,** *4* (1), 11-22.

49. Aseervatham, J., Cytoskeletal Remodeling in Cancer. *Biology (Basel)* **2020,** *9* (11).

50. Martin, S. K.; Kamelgarn, M.; Kyprianou, N., Cytoskeleton targeting value in prostate cancer treatment. *Am J Clin Exp Urol* **2014,** *2* (1), 15-26.

51. Datta, A.; Deng, S.; Gopal, V.; Yap, K. C.-H.; Halim, C. E.; Lye, M. L.; Ong, M. S.; Tan, T. Z.; Sethi, G.; Hooi, S. C.; Kumar, A. P.; Yap, C. T., Cytoskeletal Dynamics in Epithelial-Mesenchymal Transition: Insights into Therapeutic Targets for Cancer Metastasis. *Cancers* **2021,** *13* (8), 1882.

52. Oshiumi, H.; Miyashita, M.; Okamoto, M.; Morioka, Y.; Okabe, M.; Matsumoto, M.; Seya, T., DDX60 Is Involved in RIG-I-Dependent and Independent Antiviral Responses, and Its Function Is Attenuated by Virus-Induced EGFR Activation. *Cell Reports* **2015,** *11* (8), 1193-1207.

53. Zhang, J.; Fu, M.; Zhang, M.; Zhang, J.; Du, Z.; Zhang, H.; Hua, W.; Mao, Y., DDX60 Is Associated With Glioma Malignancy and Serves as a Potential Immunotherapy Biomarker. *Front Oncol* **2021,** *11*.

54. Wu, H.; Tian, W.; Tai, X.; Li, X.; Li, Z.; Shui, J.; Yu, J.; Wang, Z.; Zhu, X., Identification and functional analysis of novel oncogene DDX60L in pancreatic ductal adenocarcinoma. *BMC Genomics* **2021,** *22* (1), 833.

55. Ye, Z.; Zhang, X.; Zhang, Y.; Liu, L.; Xuan, Z.; Huang, P., Associations of DDX60L With the Clinical Features and Prognosis of Hepatocellular Carcinoma. *Front Oncol* **2022,** *12*, 761021.

56. He, L.; Liu, Y.; Lai, W.; Tian, H.; Chen, L.; Xie, L.; Liu, Z., DNA sensors, crucial receptors to resist pathogens, are deregulated in colorectal cancer and associated with initiation and progression of the disease. *J Cancer* **2020,** *11* (4), 893-905.

57. Willis-Owen, S. A. G.; Domingo-Sabugo, C.; Starren, E.; Liang, L.; Freidin, M. B.; Arseneault, M.; Zhang, Y.; Lu, S. K.; Popat, S.; Lim, E.; Nicholson, A. G.; Riazalhosseini, Y.; Lathrop, M.; Cookson, W. O. C.; Moffatt, M. F., Y disruption, autosomal hypomethylation and poor male lung cancer survival. *Scientific Reports* **2021,** *11* (1), 12453.

58. Cáceres, A.; Jene, A.; Esko, T.; Pérez-Jurado, L. A.; González, J. R., Extreme Downregulation of Chromosome Y and Cancer Risk in Men. *JNCI: Journal of the National Cancer Institute* **2020,** *112* (9), 913-920.

59. Prensner, J. R.; Feng, F. Y., “Lincing” the Y Chromosome to Prostate Cancer: TTTY15 Takes Center Stage. *European Urology* **2019,** *76* (3), 327-328.

60. Sekiguchi, T.; Iida, H.; Fukumura, J.; Nishimoto, T., Human DDX3Y, the Y-encoded isoform of RNA helicase DDX3, rescues a hamster temperature-sensitive ET24 mutant cell line with a DDX3X mutation. *Experimental Cell Research* **2004,** *300* (1), 213-222.

61. Venkataramanan, S.; Gadek, M.; Calviello, L.; Wilkins, K.; Floor, S. N., DDX3X and DDX3Y are redundant in protein synthesis. *RNA* **2021,** *27* (12), 1577-1588.

62. Köferle, A.; Schlattl, A.; Hörmann, A.; Thatikonda, V.; Popa, A.; Spreitzer, F.; Ravichandran, M. C.; Supper, V.; Oberndorfer, S.; Puchner, T.; Wieshofer, C.; Corcokovic, M.; Reiser, C.; Wöhrle, S.; Popow, J.; Pearson, M.; Martinez, J.; Weitzer, S.; Mair, B.; Neumüller, R. A., Interrogation of cancer gene dependencies reveals paralog interactions of autosome and sex chromosome-encoded genes. *Cell Reports* **2022,** *39* (2), 110636.

63. Vellky, J. E.; Ricke, E. A.; Huang, W.; Ricke, W. A., Expression and Localization of DDX3 in Prostate Cancer Progression and Metastasis. *The American Journal of Pathology* **2019,** *189* (6), 1256-1267.

64. Vellky, J. E.; McSweeney, S. T.; Ricke, E. A.; Ricke, W. A., RNA-binding protein DDX3 mediates posttranscriptional regulation of androgen receptor: A mechanism of castration resistance. *Proceedings of the National Academy of Sciences* **2020,** *117* (45), 28092-28101.

65. Huang, B.; Song, B.-l.; Xu, C., Cholesterol metabolism in cancer: mechanisms and therapeutic opportunities. *Nature Metabolism* **2020,** *2* (2), 132-141.

66. Giacomini, I.; Gianfanti, F.; Desbats, M. A.; Orso, G.; Berretta, M.; Prayer-Galetti, T.; Ragazzi, E.; Cocetta, V., Cholesterol Metabolic Reprogramming in Cancer and Its Pharmacological Modulation as Therapeutic Strategy. *Front Oncol* **2021,** *11*.

67. Wu, Z.; Xu, J.; Liang, C.; Meng, Q.; Hua, J.; Wang, W.; Zhang, B.; Liu, J.; Yu, X.; Shi, S., Emerging roles of the solute carrier family in pancreatic cancer. *Clin Transl Med* **2021,** *11* (3), e356.

68. Yoo, H. C.; Park, S. J.; Nam, M.; Kang, J.; Kim, K.; Yeo, J. H.; Kim, J.-K.; Heo, Y.; Lee, H. S.; Lee, M. Y.; Lee, C. W.; Kang, J. S.; Kim, Y.-H.; Lee, J.; Choi, J.; Hwang, G.-S.; Bang, S.; Han, J. M., A Variant of SLC1A5 Is a Mitochondrial Glutamine Transporter for Metabolic Reprogramming in Cancer Cells. *Cell Metabolism* **2020,** *31* (2), 267-283.e12.

69. Khan, T.; Sullivan, M. A.; Gunter, J. H.; Kryza, T.; Lyons, N.; He, Y.; Hooper, J. D., Revisiting Glycogen in Cancer: A Conspicuous and Targetable Enabler of Malignant Transformation. *Front Oncol* **2020,** *10*.

70. Zois, C. E.; Harris, A. L., Glycogen metabolism has a key role in the cancer microenvironment and provides new targets for cancer therapy. *J Mol Med (Berl)* **2016,** *94* (2), 137-54.

71. Lin, X.; Xiao, Z.; Chen, T.; Liang, S. H.; Guo, H., Glucose Metabolism on Tumor Plasticity, Diagnosis, and Treatment. *Front Oncol* **2020,** *10*, 317.

72. Bose, S.; Zhang, C.; Le, A., Glucose Metabolism in Cancer: The Warburg Effect and Beyond. In *The Heterogeneity of Cancer Metabolism*, Le, A., Ed. Springer International Publishing: Cham, 2021; pp 3-15.

73. Uribe, M. L.; Marrocco, I.; Yarden, Y., EGFR in Cancer: Signaling Mechanisms, Drugs, and Acquired Resistance. *Cancers (Basel)* **2021,** *13* (11).

74. Rugeles, A. C. R.; Delgado-Ramirez, Y.; Terrazas, L. I.; Leon-Cabrera, S., Chapter 9 - Targeting the STAT6 signaling pathway as a therapy against colon cancer. In *Immunotherapy in Resistant Cancer: From the Lab Bench Work to Its Clinical Perspectives*, Morales-Montor, J.; Segovia-Mendoza, M., Eds. Academic Press: 2021; Vol. 2, pp 149-172.

75. Akiyama, H.; Iwahana, Y.; Suda, M.; Yoshimura, A.; Kogai, H.; Nagashima, A.; Ohtsuka, H.; Komiya, Y.; Tashiro, F., The FBI1/Akirin2 Target Gene, BCAM, Acts as a Suppressive Oncogene. *PLOS ONE* **2013,** *8* (11), e78716.

76. Seachrist, D. D.; Sizemore, S. T.; Johnson, E.; Abdul-Karim, F. W.; Weber Bonk, K. L.; Keri, R. A., Follistatin is a metastasis suppressor in a mouse model of HER2-positive breast cancer. *Breast Cancer Research* **2017,** *19* (1), 66.

77. Yang, L.; Chen, Y.; Cui, T.; Knösel, T.; Zhang, Q.; Albring, K. F.; Huber, O.; Petersen, I., Desmoplakin acts as a tumor suppressor by inhibition of the Wnt/β-catenin signaling pathway in human lung cancer. *Carcinogenesis* **2012,** *33* (10), 1863-70.

78. Cui, T.; Yang, L.; Ma, Y.; Petersen, I.; Chen, Y., Desmocollin 3 has a tumor suppressive activity through inhibition of AKT pathway in colorectal cancer. *Experimental Cell Research* **2019,** *378* (2), 124-130.

79. Dai, Y.-H.; Wang, Y.-F.; Shen, P.-C.; Lo, C.-H.; Yang, J.-F.; Lin, C.-S.; Chao, H.-L.; Huang, W.-Y., Gene-associated methylation status of ST14 as a predictor of survival and hormone receptor positivity in breast Cancer. *BMC Cancer* **2021,** *21* (1), 945.

80. Kim, J.; Villadsen, R., The Expression Pattern of Epidermal Differentiation Marker Keratin 10 in the Normal Human Breast and Breast Cancer Cells. *Journal of Histochemistry & Cytochemistry* **2020,** *68* (8), 561-570.

81. Han, W.; Hu, C.; Fan, Z.-J.; Shen, G.-L., Transcript levels of keratin 1/5/6/14/15/16/17 as potential prognostic indicators in melanoma patients. *Scientific Reports* **2021,** *11* (1), 1023.

82. Wei, C.; Zhou, Y.; Xiong, Q.; Xiong, M.; Hou, Y.; Yang, X.; Chen, Z., Comprehensive Analysis of CPA4 as a Poor Prognostic Biomarker Correlated with Immune Cells Infiltration in Bladder Cancer. *Biology (Basel)* **2021,** *10* (11).

83. Du, Y.; Jiang, X.; Wang, B.; Cao, J.; Wang, Y.; Yu, J.; Wang, X.; Liu, H., The cancer-associated fibroblasts related gene CALD1 is a prognostic biomarker and correlated with immune infiltration in bladder cancer. *Cancer Cell International* **2021,** *21* (1), 283.

84. Sivanand, S.; Vander Heiden, M. G., Emerging Roles for Branched-Chain Amino Acid Metabolism in Cancer. *Cancer Cell* **2020,** *37* (2), 147-156.

85. Liu, A. Y.; Kanan, A. D.; Radon, T. P.; Shah, S.; Weeks, M. E.; Foster, J. M.; Sosabowski, J. K.; Dumartin, L.; Crnogorac-Jurcevic, T., AGR2, a unique tumor-associated antigen, is a promising candidate for antibody targeting. *Oncotarget* **2019,** *10* (42), 4276-4289.

86. Wang, Y.; Chen, Y.; Li, X.; Hu, W.; Zhang, Y.; Chen, L.; Chen, M.; Chen, J., Loss of expression and prognosis value of alpha-internexin in gastroenteropancreatic neuroendocrine neoplasm. *BMC Cancer* **2018,** *18* (1), 691.

87. Rahman, N. I. A.; Sato, A.; Tsevelnorov, K.; Shimizu, A.; Komeno, M.; Ahmat Amin, M. K. B.; Molla, M. R.; Soh, J. E. C.; Nguyen, L. K. C.; Wada, A.; Kawauchi, A.; Ogita, H., Stomatin-Mediated Inhibition of the Akt Signaling Axis Suppresses Tumor Growth. *Cancer Res* **2021,** *81* (9), 2318-2331.

88. Rao, S.; Oyang, L.; Liang, J.; Yi, P.; Han, Y.; Luo, X.; Xia, L.; Lin, J.; Tan, S.; Hu, J.; Wang, H.; Tang, L.; Pan, Q.; Tang, Y.; Zhou, Y.; Liao, Q., Biological Function of HYOU1 in Tumors and Other Diseases. *Onco Targets Ther* **2021,** *14*, 1727-1735.

89. Chou, S. C.; Azuma, Y.; Varia, M. A.; Raleigh, J. A., Evidence that involucrin, a marker for differentiation, is oxygen regulated in human squamous cell carcinomas. *British Journal of Cancer* **2004,** *90* (3), 728-735.

90. Arnaiz, E.; Harris, A. L., Role of Hypoxia in the Interferon Response. *Frontiers in Immunology* **2022,** *13*.

91. Kataria, N.; Martinez, C. A.; Kerr, B.; Zaiter, S. S.; Morgan, M.; McAlpine, S. R.; Cook, K. M., C-Terminal HSP90 Inhibitors Block the HIF-1 Hypoxic Response by Degrading HIF-1α through the Oxygen-Dependent Degradation Pathway. *Cell Physiol Biochem* **2019,** *53* (3), 480-495.

92. Ivanov, S. V.; Salnikow, K.; Ivanova, A. V.; Bai, L.; Lerman, M. I., Hypoxic repression of STAT1 and its downstream genes by a pVHL/HIF-1 target DEC1/STRA13. *Oncogene* **2007,** *26* (6), 802-812.

93. Miar, A.; Arnaiz, E.; Bridges, E.; Beedie, S.; Cribbs, A. P.; Downes, D. J.; Beagrie, R. A.; Rehwinkel, J.; Harris, A. L., Hypoxia Induces Transcriptional and Translational Downregulation of the Type I IFN Pathway in Multiple Cancer Cell Types. *Cancer Research* **2020,** *80* (23), 5245-5256.

94. Skowronek, P.; Meier, F., High-Throughput Mass Spectrometry-Based Proteomics with dia-PASEF. *Methods Mol Biol* **2022,** *2456*, 15-27.

95. Delafield, D. G.; Lu, G.; Kaminsky, C. J.; Li, L., High-end ion mobility mass spectrometry: A current review of analytical capacity in omics applications and structural investigations. *TrAC Trends in Analytical Chemistry* **2022,** *157*, 116761.

Supplementary Methods

Column Preparation

25cm of capillary (360µm o.d., 75µm i.d.) were taken for column packing. Using a flame, a small 3cm portion of the capillary coating was removed, approximately 5 cm from one end. An emitter was pulled using a Sutter P2000 laser puller. Column shells were flushed with MeOH to ensure adequate spray. RPLC columns were packed with 3um packing material at 100psi for 5 seconds followed by packing with 1.7µm packing material at 1500psi. Columns were packed until 15cm of the capillary was filled. Packing material for RPLC columns was suspended in chloroform. For PGC columns, this process is identical, save for the packing step. Packing material, 3µm, was suspended in MeOH and packed at 1500psi until a final length of 15cm was reached. Capillary was trimmed to a final length of ~17cm and used in-line with the Ultimate 3000 nano-flow LC system.

Supplemental Tables

For brevity, supplementary data tables have not been included here. The tables are included as part of the submitted manuscript and may be found online or may be obtained directly through directly contact the authors of this chapter.

Supplementary Figures

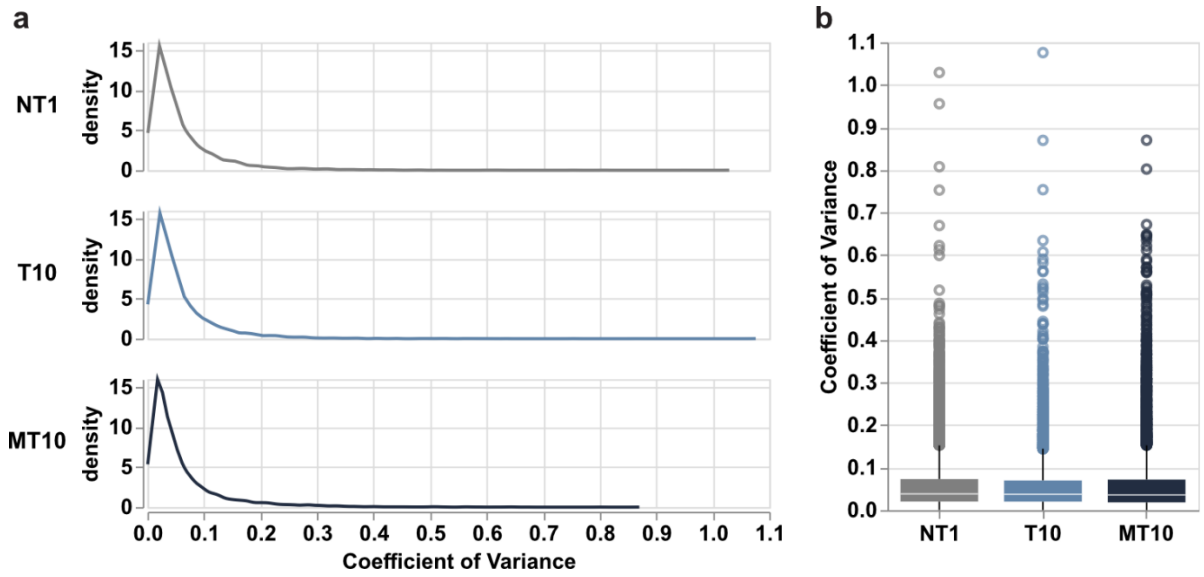


Figure S7.1 Coefficient of variance calculated for all quantified proteins.

Variance is plotted both as density plots (a) and boxplots (b) to visualize the congregation of proteins below 10% CV.

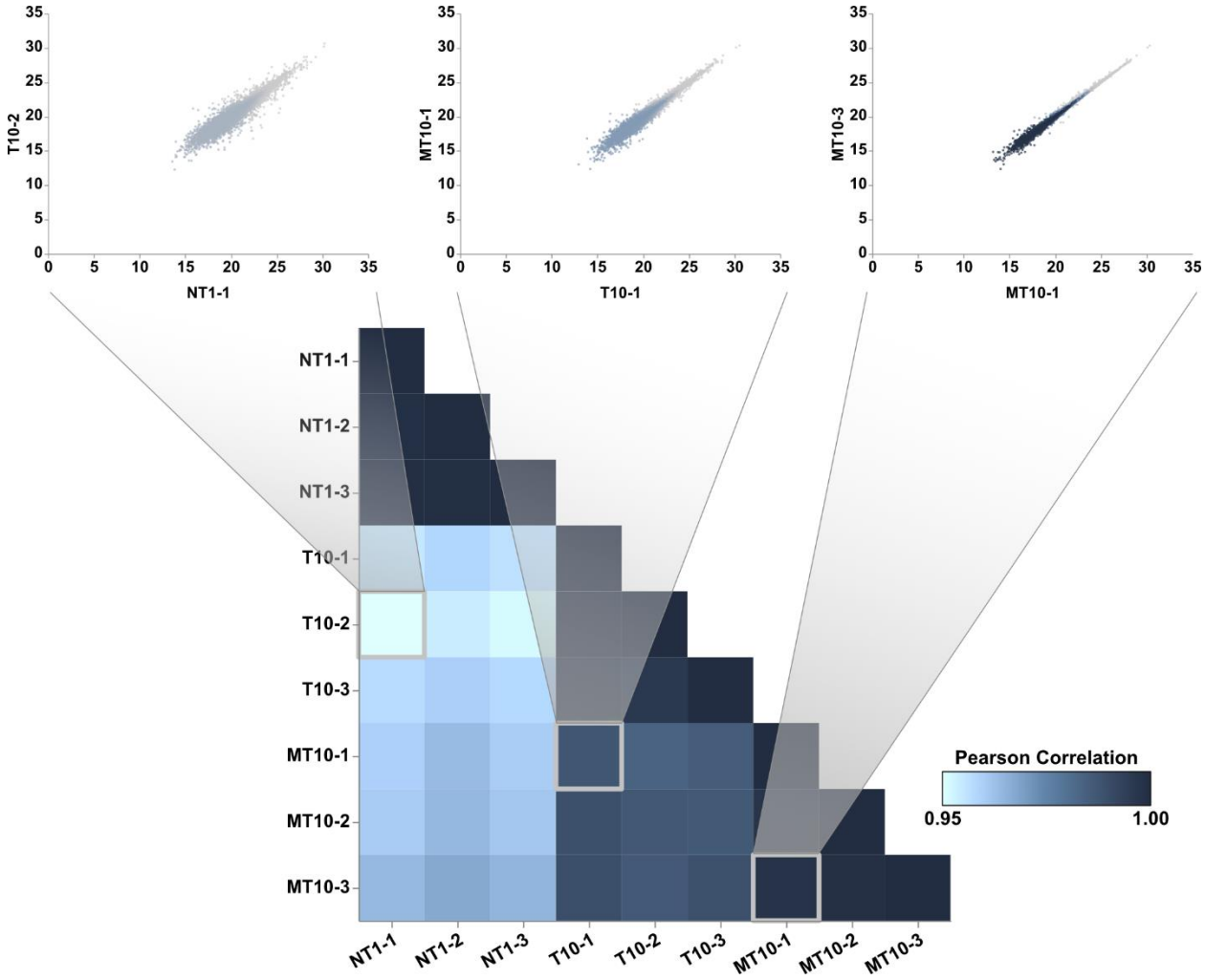


Figure S7.2 Intersample correlation.

Pearson correlation heatmap is shown at the bottom – the same as figure 3b in the main text – and is overlayed with specific sample comparisons. Scatter plots above represent the correlation of protein expression between the two represented samples and are colored to show kernel density estimate.

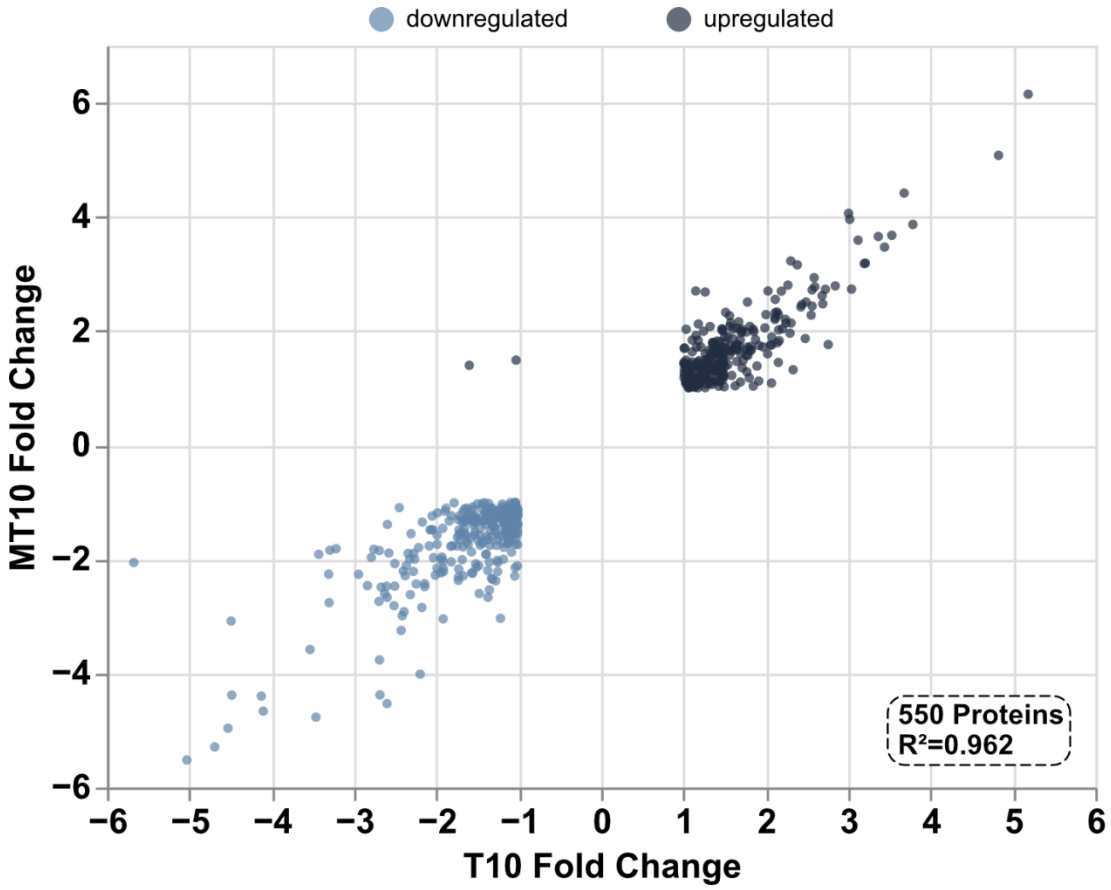


Figure S7.3 Pearson correlation of proteins up- and downregulated in malignant cell lines. Scatter plot points are colored according to their direction of dysregulation and plotted to show the consistency between T10 and MT10 cell lines.

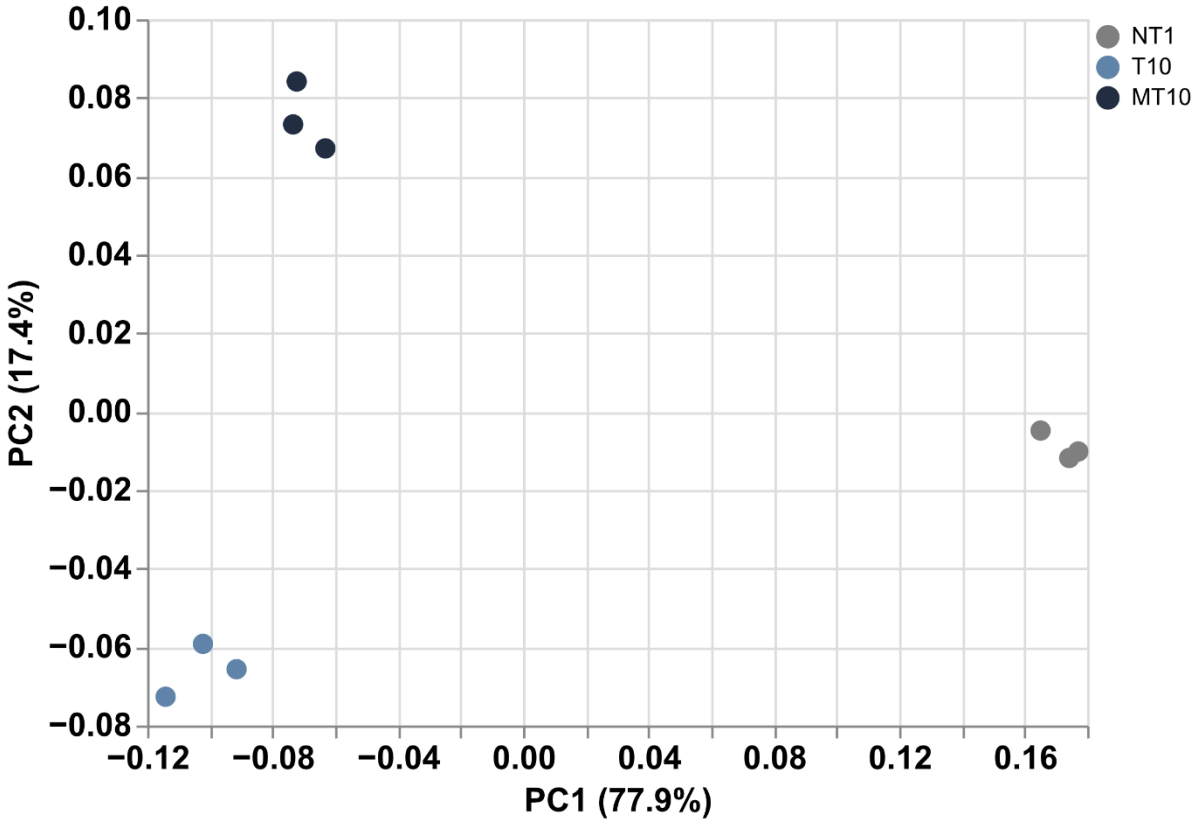


Figure S7.4 Dimensional reduction of all tested samples.

Using all 6,614 quantified proteins, NT1, T10, and MT10 cell lines are immediately separated, and all biological replicates are clustered together.

Chapter 8: Sample Agnostic Spectral Libraries Enable Quantitation of >9,300 Cerebrospinal Fluid Proteins Across Neurodegenerative Disease Patient Cohorts

Delafield, D. G.[†], Zhong, X. [†], Sauer, C. S., Yu, Q., Zetterberg, H., Li, L. Sample Agnostic Spectral Libraries Enable Quantitation of >9,300 Cerebrospinal Fluid Proteins Across Neurodegenerative Disease Patient Cohorts. In submission. 2023.

[†] These authors contributed equally to this work.

Abstract

Proteomic analyses of mild cognitive impairment (MCI) and its frequent successor, Alzheimer's Disease (AD), are necessary to elucidate means of early detection and putative therapeutic targets. Cerebrospinal fluid (CSF) offers the most direct observation of neurological protein perturbation but suffers from low sample availability and high dynamic range of protein expression. Here we present a modular framework to generate and employ sample agnostic spectral libraries to enhance profiling and quantitative depth. Utilizing an open-source machine learning approach to 'calibrate' comprehensive libraries to new experimental conditions, we quantified 9,313 protein groups in CSF through data-independent acquisition (DIA) mass spectrometry, nearly a 14-fold increase compared to a traditional DIA workflow that uses a data-dependent acquisition (DDA) spectral library. Revealing 1,642 significantly dysregulated protein groups against healthy controls, this study not only validates a flexible approach towards comprehensive sample profiling but also provides understanding of protein targets useful for disease stratification and treatment.

Introduction

Among the more than 50 million worldwide diagnoses of dementia, Alzheimer's Disease (AD) is the most prevalent and widely recognized with multiple reports forecasting a significant increase in confirmed diagnoses over the coming years[1]. The classical presentation of AD pathology is the formation of neurofibrillary tangles and amyloid plaques that inspire inflammation and initiate a cascade of adverse neuropathologic changes[2], which in turn can lead to memory deficits, cognitive decline, and a deteriorated quality of life[1]. Though developed nations in North America and Western Europe currently lead in number of dementia diagnoses[3], these figures point

both towards the rising rates of incidence as well as global disparities in detection, stratification, and correct diagnosis of these neurological disorders. Such adverse realities highlight the urgent need to develop and validate rapid, facile strategies for disease identification and assignment of severity.

Because brain tissue may only be studied postmortem, researchers have turned to biofluids as a primary source to study AD-related proteomic perturbations. Cerebrospinal fluid (CSF) is considered the gold standard in the effort to extract proteomic information related to neurological disorders due to its confinement to the central nervous system and contact with the brain. Mass spectrometry (MS)-based proteomic analysis has been a method-of-choice to uncover biomolecular candidates to serve as effective biomarkers or therapeutic targets and has seen regular employment in the analysis of CSF[4, 5]. However, the extreme dynamic range of protein expression within CSF coupled with stochastic precursor selection in typical shotgun proteomics experiments severely hinders the profiling depth that may be achieved. Data-independent acquisition mass spectrometry (DIA-MS) contrasts these traditional methodologies by isolating and fragmenting all precursor ions within a pre-defined mass range, offering substantially higher throughput than routine data-dependent acquisition (DDA) analyses. In turn, DIA removes sampling bias, increases profiling depth, and bolsters experimental reproducibility, making it a favorable strategy in dynamic, sample-limited scenarios[6].

Though the benefits of DIA-MS are widely reported, a principal concern is the pursuit of time efficient experimental design and avenues for adept data deconvolution[7]. As DIA data analysis commonly depends on the use of spectral libraries – empirical collections of precursor masses, retention times and tandem mass spectra – constructing these libraries is considered a necessary component of experimental design, increasing sample requirement and analysis time.

Further, the information gleaned from DIA analyses is effectively limited to the depth of the library, mandating exceptional quality. Several tools have been validated for their ability to perform 'library-free' analysis[8-10], generating *in silico* tandem mass spectra, and predicting an expected retention time based on experimental data or internal calibrants. While successful, these freeware and proprietary tools do not offer a modular approach that provides the flexibility expected in custom experimental design, confining users to a predetermined machine learning architecture and prohibiting the use of empirical datasets to build spectral libraries.

Advances in DIA-MS-based human health research hinge on flexibility in the approaches employed to predict retention times, the chromatography paradigm employed during analysis, and portion of the proteome being targeted. Only with this modularity will researchers be able to rapidly adapt new technologies to a chosen analytical platform and be empowered to reemploy comprehensive spectral libraries tailored to a given proteomic objective. To provide this flexibility and modularity in approach, we validate here the ability to accurately predict peptide retention time using an open-source machine learning model. This trained model was then re-deployed, forecasting the expected retention times of all peptide sequences within an external, third-party spectral library. Forecasting these retention times effectively calibrates this spectral library to our experiment, aligning empirical spectra with their anticipated elution times and allowing downstream software to identify peptide precursors that would otherwise be missed. Verifying success of this approach would outline an open framework easily adapted to new investigations as each component of this workflow - the chromatographic separation, the machine learning architecture, the external spectral library, and the processing software – may be exchanged according to user needs.

We therefore provide empirical validation of this methodology through retention time calibration of the MASSIVE-KB[11] human proteome spectral library and deploying this library in the analysis of CSF from neurodegenerative disease cohorts. Coupling these untargeted analyses with label-free quantitation, we successfully quantified >9,300 protein groups with 1,642 shown to be significantly dysregulated against healthy controls. In addition to the significant improvements in peptide and protein identifications seen through this approach, our analyses rearticulate the biological significance of numerous protein species within Alzheimer's Disease and further reveal areas of proteome reorganization that provide insight into therapeutic targets and diagnostic markers.

Results

Experimental Design & Validation of Accurate Retention Time Prediction

To establish the advantages of sample agnostic libraries over traditional libraries constructed from sample fractionation, our experimental design compared these two methodologies in parallel, with traditional DDA analyses included to serve as baseline performance. Control, MCI and AD cohort samples were prepared as described with individual samples analyzed via DDA and SWATH-like DIA (see methods). Meanwhile, all the cerebrospinal fluid (CSF) samples were pooled and split amongst 15 high pH fractions and analyzed via DDA to construct a traditional spectral library (referred to as the 'sample specific' library). The outputs of all DDA analyses (both fractionated and pooled samples) were used to compile the machine learning dataset and the resulting model was used to predict retention time of sequences within the MassIVE-KB spectral library. This final, retention time-corrected spectral library (the 'sample agnostic' spectral library) was used for precursor identification (Figure 8.1).

Peptide identifications from DDA and fractionated samples (see methods) were filtered to accommodate a peptide-level 1% FDR cutoff, leaving 164,269 total identifications (25,284 unique sequences) to use for model training and validation. Duplicate identifications (i.e., from replicate DDA analyses, etc.) were retained in this collection to account for expected run-to-run variance and ensure model generalization. The dataset (Supplemental Data 1) was split 70:30 for training and testing, respectively. Though numerous machine learning-based retention time prediction architectures have been previously reported, we utilized DeepRT+[12], an open-source capsule network implementation. As capsule networks demonstrate the ability to preserve spatial information during convolution, they are well-suited to recognize relationships between amino acids and their effect on retention time. After 20 rounds of training, the predictive model (referred to as the 'global' model) showed excellent agreement between predicted and experimental retention times (Figure 8.2a, Supplemental Data 2). Within the testing dataset, predicted retention times showed an average difference of 0.8 minutes compared to experimental values with 94.7% of all predicted retention times showing deviation ≤5 minutes (Figure 8.2b).

We further ensured the reliability of our chosen capsule network retention time prediction workflow through 10-fold cross validation. Using the same collection of peptide identifications, data were partitioned 90:10 for training and testing and subjected to the same training procedure as before. This process was repeated 10 times to ensure all precursors were used in a testing set at least once. All ten cross validation iterations demonstrated high reproducibility in retention time prediction, all but two showing linear regression fit >0.99 (Figure 8.2c-d, Supplemental Data 3). With each iteration rearticulating the consistency of retention time prediction seen in the global model (Figure 8.2c-e, Figure S8.1), these data establish low out-of-sample error and highlight the suitability of the global model to predict retention time of previously unseen peptide sequences.

Though demonstrated effective in our approach, this machine learning architecture is only one component of the entire workflow. Future applications seeking to emulate this approach may exchange this architecture for newly conceived approaches or a model that more suitable to a given experiment.

Assessing Reproducibility, Reliability, and Data Quality of a Sample Agnostic Spectral Library

To establish stringent filtering criteria when employing the sample agnostic library, the first consideration was parameters that provide high intrasample reproducibility. These criteria were established using one of the three healthy control samples as a representative, arbitrary test case. Utilizing precursor isotope dot product – a Skyline[13] measure of confidence based on the similarity between extracted precursor isotopic masses – to construct progressive thresholds, all data demonstrate high reproducibility and coalesce along the expected correlational trend (Figure 8.3a). The intrasample reproducibility and depth of quality data provide the ability to set strict cutoff thresholds, with evidence of 10,011 proteins remaining when the isotope dot product threshold is ≥0.8. Critically evaluating this threshold against precursors identified in DDA analyses and the sample specific library determined that a threshold of ≥0.7 to be more rigorous, retaining approximately 50% of the sample specific library but only 13.47% of the agnostic library (Figure 8.3b). We confirmed the suitability of this lower threshold by manually inspecting the MS^1 and MS^2 information of peptide precursors identified only in the agnostic library (Figure 8.3c). To provide the strictest overall filtering criteria, this dot product threshold quality it utilized in tandem with the requirement that a given peptide precursor must be identified in every analyzed biological replicate.

Having determined data quality thresholds, extracted precursors from all experiments were filtered for redundancy and averaged across replicates. Peptide abundances were aggregated by protein group with the average abundance representing protein quantity. Only precursors that were identified in every sample were considered and no more than the top 6 precursors were averaged for a given protein (see methods); proteins remaining after this step are considered quantifiable proteins. In all, DDA analysis of individual CSF samples resulted in only 500 quantifiable protein groups mapping to 585 potential protein matches – expectedly modest performance in this acquisition paradigm. The sample specific library provided 655 quantifiable protein groups (769 putative proteins), representing 42% of all proteins available in the specific library. The sample agnostic library significantly outpaced these results, revealing 9,313 quantifiable protein groups mapping to 9,702 protein sequences, a ~14-fold increase in protein identifications (Figure 8.3d, Supplemental Data 4-5). The majority of these proteins were commonly quantified between two of the three identification strategies, though DDA and the sample specific analyses were able to quantify proteins not found in the agnostic library (Figure 8.3e) due to the positive identification of unique, variably modified peptide sequences, which is discussed below.

To ensure quantitative reliability, we compared protein-level quantitation between the sample specific and sample agnostic libraries. Of the 681 proteins commonly quantified between these two groups, 528 proteins were quantified using the same peptide precursors and could be directly compared between the two methods. The consistency of protein quantitation between the two libraries (Figure 8.3f) further demonstrates the reliability of the agnostic library compared to traditional spectral libraries. Discrepancies in protein quantitation arise largely due to differences in retention time (predicted vs empirical), the allowable tolerances during data extraction (see methods), and accurate bounding of precursors during data processing. We further investigated the

source of unique peptide identifications that led to quantifiable proteins not seen in the agnostic library. Across all samples, 530,984 unique peptide sequences were identified only using the sample agnostic library with an additional 12,026 sequence shared between the agnostic library and DDA analyses (fractionated and individual samples, Supplemental Data 6). Of the 9,858 peptide sequences not identified using the agnostic library, 7,249 sequences (73.5%) were due to variable modifications only allowed during DDA database searches (Figure 8.3g). 1,630 sequences (16.5%) were not available in the agnostic library due to filtering criteria or were identified as part of a missed cleavage, leaving 929 sequences as truly unidentified. However, when accounting for low quality precursors, we only expanded our search to those sequences with a dot product ≥0.5 due to computational constraints; the final 10% of unidentified sequences would likely be found in the agnostic library if our search was expanded. It should be noted that newer versions of the MassIVE-KB (released after the time of analysis) have better support for common variable modifications, which may assist future iterations of this workflow.

Quantitative Investigations of Cerebrospinal Fluid

Having established the reproducibility and reliability of a sample agnostic spectral library, we applied this approach to label free quantitative investigation of cerebrospinal fluid (CSF) between healthy controls, mild cognitive impairment (MCI) and Alzheimer's Disease (AD) patient cohorts (see methods). With all protein abundances compared against control to determine fold change and statistical significance determined via Student's t-test, our data reveal 1,642 significantly dysregulated proteins across the two disease cohorts (fold change≥2, P≤0.05, Supplemental Data 5, 7). While both MCI and AD cohorts revealed similar quantities of significantly upregulated protein groups – 307 and 336, respectively (Figure 8.4a) – the MCI

cohorts displayed a striking difference in the number of downregulated protein groups, 993, nearly doubling the quantity downregulated in AD patients.

We further dissected these dysregulated proteins by employing clustering analysis to correlate protein expression levels with disease state, identifying three protein subpopulations bearing notable resemblance. The two groups most immediately distinguished are those significantly up- (160 protein groups) and downregulated (356 protein groups) in both MCI and AD patient cohorts (Figure 8.4b). While our global analysis required a fold change ≥2 to be considered significantly dysregulated, these protein populations far exceeded this threshold, averaging 5.08±0.21-fold higher and 4.21±0.32-fold lower than the same proteins in control groups. Further, the expression profiles of these proteins were relatively consistent across MCI and AD cohorts, evidenced by the low standard deviation and high correlation (Pearson r=0.98, Figure 8.4c, Figure S8.2) between cohorts.

Given the significant difference in expression between healthy and disease samples and the remarkable similarity across MCI and AD cohorts, we questioned whether these proteins had any previously established relationship to neurodegenerative disease. From among the 160 commonly upregulated protein groups (Supplemental Data 7), we identified several conserved protein groups with known relationship to Alzheimer's Disease and neurological disorders such as cadherins and protocadherins[14], growth factors, metalloproteinases[15, 16], among others (Figure S8.3). Beyond this, several interesting targets emerged such as P-selectin glycoprotein ligand 1 (SELPLG), which mediates inflammatory response[17], is shown to be upregulated in AD cohorts[18] and is a known substrate for the APP-cleaving protease BACE1[19]. Translin (TSN) was also shown to be upregulated, suggesting increased neuronal trafficking of brain-derived neurotrophic factor[20]. The tubulin-modifying protein monoglycylase TTLL8 (TTLL8)[21, 22] was conserved across disease

groups, as well as cell surface hyaluronidase (CEMIP2), pointing towards the antagonistic activity of hyaluronan in AD[23]. Our data also reveal upregulation of spermine synthase (SMS), highlighting production of free radical-scavenging spermine[24]. Neuropilin tolloid-like 1 (NETO1), which associates with the amyloid precursor/ n-methyl-d-aspartate complex[25], is also significantly upregulated.

The 356 consistently downregulated protein groups rearticulate the dysregulation of similar protein clusters as noted above, whereas this collection of proteins was also found to be rich in modifying enzymes such as kinases, phosphatases, transferases, and others (Supplemental Data 7). Among the individual targets of interest, synaptodpodin (SYNPO) has been previously validated as a downregulated protein in AD[26], though knock-out studies have shown amelioration of AD symptoms[27]. Myeoblastin (PRTN3) has also previously shown a high negative correlation with amyloid[28] and is conjectured to have a sign of dementia and synaptic loss[29]. Serpins B7 and B9 were shown to be downregulated in disease groups, the latter of which is a regulator of granzyme B and interleukin-1β[30]. Interestingly, our data show a consistent downregulation of spermidine synthase (SRM), contrasting the upregulation of SMS and highlighting the recently suggested therapeutic effects of spermidine supplementation in AD[31]. Neurturin (NRTN), a neurotrophic factor[32] suggested as therapeutic treatment for neurodegenerative diseases[33, 34], as well as the neuroprotective[35] thrombospondin (TSP) that has is a known downregulated protein in AD[36], are commonly dysregulated and are also highlighted in latter analyses (see below). Degradation of microtubule-associated protein 1A (MAP1A) is a known consequence of AB-induced neurodegeneration[37] and is likewise dysregulated in our data. Others include tetraspanin 5 (TSPAN5), which regulates the cleavage of amyloid precursor (APP) by ADAM10[38], and EphrinB3 that impacts axonal growth, spatial learning, and memory[39]. While these significantly

up- and downregulated proteins point towards a broad, evidence-based protein panel that can distinguish disease from healthy control, the conserved regulation between MCI and AD render them incapable of deciphering disease cohorts from one another.

The final subpopulation emerging in our analyses were those 637 protein groups found to be significantly downregulated in MCI cohorts that but were not downregulated in AD (Supplemental Data 5). When accounting for these proteins, all three patient cohorts were immediately distinguished from one another in all downstream analyses. This is exemplified through column-wise hierarchical clustering, which revealed both the stratification of disease state and the correlation of similarly expressed proteins (Figure 8.4d). The capacity to distinguish control, MCI, and AD patients is further confirmed through dimensional reduction analysis, revealing concise distribution within sample cohorts and no ambiguity between disease states (Figure 8.4e). Despite the heterogeneity of this protein population, we identified participants in several relevant processes such as tubulins (TUBB1, TUBB6, TUBAL3, TUBA1C, and others) involved in axonal guidance[40], shootin 1 (SHTN1) involved in neuronal migration and polarity[41], as well as neurofibromin 1[42] (NF1) and syntaphilin[43] (SNPH) involved in brain development.

Overall, these data serve as evidence as to the specificity and utility of the proposed method. Having established the significant improvements in profiling depth and ensuring quantitative reliability, closely examining those proteins shown to be significantly dysregulated reveals numerous proteins related to dementia and AD. Further, the correlation of these disease-specific dysregulated proteins between AD and MCI that starkly contrasts the healthy controls demonstrates the data extracted and kept for comparison is not stochastic but is instead topically relevant to the system being observed. Finally, the proteomic profiles extracted during analysis highlight the accuracy with which disease cohorts may be distinguished. Though commonly

dysregulated proteins were sufficient in distinguishing healthy controls from disease cohorts, the unique expression profiles that are conserved between unique MCI and AD biological replicates allow for immediate discrimination of disease groups. This evidence further suggests a flexible approach for building sample agnostic spectral libraries will decrease experimental burden and facilitate rapid, accurate patient profiling useful for disease detection and stratification.

Proteomic Reorganization Associated with Neurodegenerative Disease

Having established the capacity for the sample agnostic spectral library to quantify proteomic perturbations useful in distinguishing healthy controls from disease cohorts, as well as stratify MCI and AD disease states, we further scrutinized those dysregulated proteins to characterize proteomic pathways and processes significantly impacted within our disease cohorts, as well as illuminate proteins known to be of interest in AD research and treatment. As a cursory analysis, upregulated and downregulated proteins were separated and subjected to Gene Ontology enrichment (see methods). Examining these analyses, our data reveal the dysregulation of several central processes and pathways known to be active participants in Alzheimer's Disease.

Proteins shown to be significantly downregulated were notably involved in adaptive immune response (Figure 8.5a, Figure S8.4), reaffirming the emerging evidence that immunodeficiency contributes to AD pathogenesis[44], a compelling direction for future proteomic analyses. Noting the activation of immune response, we expected to observe an increase in cytokine production that promote the hallmark inflammatory response to AD. Indeed, our ontology results did indicate upregulation of the genes related to Interleukin-1 (IL-1) signaling (Figure S8.5-6) as well as genes related to activation of nuclear factor kappa B (NF-KB), the upstream regulator of IL-1 production. However, dissecting the proteins assigned to these gene clusters, upregulation

was largely confined to the proteasomal processing proteins, whereas other proteins dysregulated in these pathways seem to indicate inhibition of the overall processes. Of the proteins related to NF-KB activation, for example, we demonstrate significant downregulation of the catalytical IKKa subunit (CHUK) and the NF-KB essential modulator (NEMO/IKBKG) while NF-KB p105 (NFKB1) – a preprocessed subunit – displayed significant upregulation. Suggesting overall inhibition or slowing of NF-KB activation, this hypothesis is further complemented by the downregulation of protein ubiquitination (Figure 8.5b., Figure S8.5) necessary for NF-KB processing, as well as the downregulation of IL-18 and IL-37, two closely related downstream products of activated NF-KB. As NF-KB-related processes are routinely implicated in AD, our observations may serve to indicate feedback inhibition of NF-KB activation in our disease cohorts, though further analyses are required for confirmation.

In addition to dysregulation of protein ubiquitination – a critical driver of protein regulation and degradation[45] – our data also reveal downregulation of genes controlling phosphorylation (Fig 5b). This observation is expected giving the known importance of protein phosphorylation in AD[46], namely the hyperphosphorylation of Tau[47]. Phosphoproteomic analysis in AD is an area of interest and is amenable to the machine learning architecture we employed here[12] but was not a goal of the present study. Among the numerous processes of interest shown to be dysregulated, our analyses repeatedly indicate negative regulation of cell cycle (Fig 5a-b), a known actor in Alzheimer's and other neurodegenerative diseases[48]. This observation in tandem with downregulated mitotic cell cycle processes and DNA repair (Figure 8.5a-b), as well as an upregulation of BRAF (Figure 8.5c), rearticulate previous reports associating prolonged mitosis, cellular senescence, and age-related deterioration in AD[49]. In addition, our data reveal downregulation of programmed cell death and apoptosis, which often coincides with chronic oxidative stress in Alzheimer's Disease[50]. Finally,

we also observed downregulation of genes pertaining to exocytosis and diminished synaptic transmission in AD[51, 52].

Given the quantitative depth of this experiment and the number of disease-related proteins already observed, we expected common AD hallmarks to show similar expression to previous reports. One limitation of this experiment is that three common AD markers, Tau (MAPT), amyloid precursor (APP), and apolipoprotein E (APOE) were not shown to be significantly dysregulated, bearing similar intensity between healthy and disease groups (Figure 8.5c). For MAPT, this limitation is largely due to the low number of peptides identified in all samples (Figure S8.7). The peptides used for quantitation of APP and APOE were not useful in determining quantitative differences, but still facilitate some degree of distinction between healthy and disease groups when clustering analysis is used (Figure S8.7). We speculate these observations are a result of pooling numerous patient samples (n=5 for each pool) prior to analysis; this may be overcome in latter experiments. Despite this limitation, manual inspection of known AD-related proteins revealed significant changes in protein abundance between healthy and AD cohorts, some proteins changing by >2 orders of magnitude (Figure 8.5c). Of those not previously discussed, we identified numerous mitochondrial microenvironment proteins (NDUFAB1, NDUFA4, UQCRFS1, UQCRC1, among others) that have been recently connected to AD[53-55]. We also reveal downregulation of nicastrin (NCSTN), suggesting disruption of AB processing that can lead to AB accumulation[56]. Kinesin light chain (KLC1), downregulated in this study, has been previously reported in brains of Alzheimer's patients[57], but brain acid soluble protein (BASP1) is shown to be upregulated, suggesting a move towards neuroprotection[58].

Among our analysis, we also noticed significant dysregulation of the semaphorin family of proteins, as well as their plexin and neuropilin receptors (Figure 8.5c). Semaphorins are long-

known participants in nervous system development that have recently been discussed for their roles in adult neuronal plasticity[59]. Of the semaphorins dysregulated in our study, SEMA6D is the least evidenced in connection to AD, though it has been proposed to be associated to AD pathogenesis[60] through the known TREM2-PLXNA1-SEMA6D interaction[61]. SEMA3D[62], SEMA3F[63], and SEMA4C[64], are critical components of the adult brain with dysregulation of plexin and neuropilins known to be linked to neurological disorders[59]. Though the role of these proteins in AD cannot be ignored, their known association with the extracellular matrix (ECM) caused us to question whether or not our disease cohorts exhibited some of ECM reorganization in response to AD. Referencing proteins known to be expressed in the ECM[65], we identified 55 known ECM components with significant dysregulation in at least one disease cohort (Figure 8.5d). Transglutaminases (TGM2, TGM3, TGM5), metalloproteinases (ADAMTS15, ADAM9, ADAM21, MMP15), serpins (SERPINH1, SERPINA10, SERPINB9, SERPINB7), S100 proteins (S100A4, S100A6, S100A8), and thrombospondins (THSB1, THSB3), among others, were shown to be dysregulated against healthy controls, rearticulating previous reports that suggest the multi-faceted role of ECM within neurological disorders[66]. Taken together, these data rearticulate the known proteomic alterations associated with AD and further illuminate the complex network of aberrations contributing to, or stemming from, AD pathogenesis. As the importance of these biomolecular alterations becomes clear and the clinical relevance of comprehensive proteomic profiling is established, rapid, sample-efficient strategies such as that presented here are sure to be of value in disease diagnoses and monitoring.

Discussion

Mass spectrometry-based proteomic profiling is fundamentally limited by the inherent complexity of biological mixtures and the direct correlation between profiling depth, sample requirement, and analysis time. DIA analyses offer a direct remedy to these ailments but often constrain users to a few rigid solutions for data deconvolution. This sample agnostic spectral library approach utilizes an open-source machine learning architecture that may be changed, optimized, or substituted without penalty to downstream analysis. The modularity presented within seeks to enable researchers to implement new technologies and freely optimize analytical components (chromatography, instrumentation, etc.), an aspect paramount to the changing landscape of mass spectrometry. The approach presented within does come with the cost of being computationally expensive; however, this limitation may be balanced by reducing the proteomic search space to those proteins of greatest relevance or those evidenced in survey DDA analyses.

Given the ubiquity of library-based approaches in DIA analyses, this framework for generating and employing agnostic libraries is presented to directly enhance those more traditional approaches typically employed. Utilizing a sample specific library – constructed through DDA analysis of fractionated pooled samples – significantly underperformed this agnostic library approach in terms of identification depth and quantitation; though the sample specific library did outperform strict DDA analysis of individual samples. However, we do acknowledge the earlier generation of mass spectrometer used within this experiment. With the rise of instrumental platforms that provide substantially improved acquisition speed and profiling depth in routine DDA experiments, later evaluations may determine that a sample specific library is more competitive than that presented here. Nevertheless, as a means to eliminate the need for extensive

fractionation, reduce sample requirement, and shorten analysis time, this approach remains worthwhile.

Having applied this methodology to analysis of CSF, the content of which is dominated by albumin and globulin, the capacity to identify and quantify >9,300 protein groups highlights the potential utility of this method in unique proteomic investigations. Knowing the concerted effort behind utilizing non-invasive sampling and easily collected biofluids, as well as the ever-present need for extensive sample cleanup prior to analysis, this analytical method represents an avenue towards ascertaining extensive proteomic information without the need for dedicated enrichments or depletion of abundant analytes. Though extensive trials are needed to assess the efficacy of this agnostic library approach in different biofluids, this initial report indicates it is a competitive approach that will substantially reduce sample handling and improve analytical throughput.

Finally, the utilization of publicly available data to construct the agnostic spectral library is a principal strength. As the approach demonstrated here utilizes the MassIVE-KB human HCD spectral library, constructed from data provided by the proteomics community and is reported to contain ‘best-evidence’ spectra for each constituent peptide, it is reasonable to conclude that a similar community-drive approach can, and potentially should, be employed for unique proteomic objectives. Whether it is proteomic analysis from specific tissue types, organ systems, organisms, or post-translational modification profiling, each of these unique directions may benefit from the construction of evidence-based spectral libraries that can be used for DIA analysis. In this way, researchers may achieve deep, rapid proteomic profiling within their chosen system without the need for laborious sample fractionation and long DDA instrument runs. This possibility will require intentional efforts from the research community to both compile these spectral libraries, as well as agreeing on criteria for reliable precursor identification. DIA MS analyses will never

replace DDA, especially in the pursuit of novel discovery, but more intentional efforts to blend the success of these two paradigms are critical to rapid development in clinical MS applications.

Methods

CSF

CSF samples were obtained from Wisconsin Alzheimer's Disease Research Center and all study procedures have been approved by the University of Wisconsin Institutional Review Board. Disease state was assigned via a comprehensive neuropsychological test battery, Mini-Mental State Examination, positron emission tomography (PET) imaging of β-amyloid, and ^{18}F Fluoro-2-deoxy-glucose (FDG)-PET of hypometabolism. 45 total patients were used within this study, evenly split across healthy control, MCI and AD cohorts. Each disease cohort of 15 patients was further divided into 3 subgroups to serve as biological replicates; each patient sample was used only once. Sex distribution and age information of all study subjects was provided in Supplemental Data 8.

Sample Preparation

CSF protein concentrations were estimated via bicinchoninic acid (BCA) protein assay (Thermo Pierce, Rockford, IL). 10 µg protein was taken from each patient and pooled according to the description above. Using 50 mM tris-HCl (pH=8.0) as the stabilizing buffer, proteins were denatured in 8M urea at room temperature for 5 minutes. Disulfide bonds were reduced with 100mM DTT for 1 hour at room temperature followed by alkylation with 200mM IAA for 30 minutes at room temperature. Urea content was reduced to 0.9M before adding mass spec grade

trypsin/Lys-C mix 1:25 (*w:w*). Protein digestion was incubated overnight at 37 ^{o}C, quenched by 10% TFA to pH 3, and desalted with Bond Elut OMIX C18 pipette tips (Agilent, Santa Clara, CA).

High pH reversed-phase fractionation.

To build a sample spectral library, 2 µL CSF sample from each patient was mixed, followed by separation on a Phenomenex Kinetex 5 µm C18 100Å column (L=150mm i.d.=2.1mm). Fractionation was facilitated by a Waters e2659 separation module with a Waters 2489 UV/Visible detector. Mobile phase A was 10mM NH_4HCO_2 (pH 10) and mobile phase B was 10mM NH_4HCO_2 in 90% ACN (pH 10). Samples were loaded and fractionated using the following gradient: 1% buffer B from minute 0-3, 35%B at minute 50, 60%B at minute 54, 70%B at minute 60, 100%B minutes 61-74, 1%B from minutes 74.5-94. Flow rate was held constant at 0.2mL/min. The fractions of the first 60 minutes were collected with 2-min interval and recombined into a total of 15 final fractions based on UV chromatogram. All the fractions were concentrated under vacuum. This experimental protocol was performed in duplicates and termed as Mix1 and Mix2 in Supplemental data 4.

Mass Spectrometry Data Acquisition

All separations were performed on NanoUltimate 3000 UPLC chromatography stack using a 15cm self-packed capillary C18 reverse phase column. Mobile phases A and B were 0.1% FA and 0.1% FA/100% ACN, respectively. The CSF samples from CTRL, MCI, or AD were reconstituted in 50 µL mobile phase A and analyzed in technical replicates using both DDA and DIA methods. The high pH fractions were resuspended in 20 µL mobile phase A and analyzed only once with DDA method. Samples were loaded and trapped on the column at 3% B for 16 minutes. The gradient ramped from 3% to 35% B over 94 min, followed by a 10-minute elution at

75% B for 10 min, then held at 95% B for 10 minutes and a 10-minute equilibration at 3% B. The gradient was consistent between all DDA and DIA experiments.

Data were collected on a ThermoFisher Scientific Q-Exactive HF mass spectrometer. In DDA, a full MS scan over a *m*/*z* range of 300-1600 was acquired at a resolving power of 60,000, an AGC target of $1e^6$, and a maximum ion injection time of 50ms. Tandem MS analysis was performed in a data-dependent top 15 manner for precursors at charge states 2-6 using normalized higher energy collisional dissociation (HCD) of 30%. Dynamic exclusion was set to 45 seconds with an exclusion width of 10 ppm. In DIA, a full MS scan at a resolving power of 60,000 was acquired over a *m*/*z* range of 400-1200 with an AGC target of $1e^6$ and a maximum injection time of 50ms. The MS/MS was collected by using a 36 m/z isolation window over 400-1200 *m*/*z* (a total of 24 mass windows) at a normalized HCD of 30%. The control, MCI and AD cohort samples were technically analyzed twice in both DDA and DIA. Throughout this study, biological replicates are denoted by suffixes of sample names (e.g., CTRL1 and CTRL2 are biological replicates, Supplemental Data 8) while technical replicates are denoted separately as “rep 1” or “rep 2.”

Database Searching

DDA .raw files were searched using PEAKS Xpro. Precursors were matched using a 20ppm tolerance with fragments allowed a 0.02Da mass error. PEAKS was instructed to consider only strict trypsin digestion with up to 2 missed cleavage sites. All cysteines were considered to have a fixed carbamidomethyl modification while n-terminal acetylation, Asn and Glu deamidation, and Met oxidation were allowed as variable modifications. De novo peptides were matched against the UniProt reviewed human proteome (downloaded October 30, 2021) and all results were filtered to a peptide-level 1% FDR cutoff.

Machine Learning

Machine learning was facilitated using the previously published and validated DeepRT+[12] capsule network machine learning model, the authors of which receive full credit for implementation and utility. Peptide sequences from PEAKS database searching were split 70:30 training and testing and subjected to a 20-epoch training cycle, as noted in the original publication[12]. The final model (the 20th epoch) was used as the 'global model' throughout this report.

Spectral Library Generation.

The Human HCD Spectral Library was downloaded as a .mgf file from the MassIVe-KB website (https://massive.ucsd.edu). As a note, the library used in this report is now labeled as a 'v1' release and newer 'v2' releases are available since the time of analysis. Peptide sequences were extracted from the .mgf file and subjected to retention time prediction using the global ML model described above. The resulting retention times were then reinserted to the .mgf file. The final .mgf was submitted to Peaks for database matching, using the same parameters listed above; results were exported as .mzid. The final .mzid file was submitted to Skyline (version 21.2.0.425) and the resulting spectral library was used for all subsequent analyses.

Gene Ontology

Gene ontology was performed using Metascape[67]. Dysregulated proteins were submitted as UniProt accession numbers and 'express analysis' was used for all comparisons. Gene ontology bar charts and protein networks were downloaded directly from the results page. Protein network colors and position were edited in Cytoscape[68].

Precursor Assignment and Extraction

Peptide searches were imported into Skyline using the spectral library described above using a DIA workflow. Only modifications recognized in the .mgf file were retained during searching. Transition settings were set as follows: precursor charges, 2-7; ion charges, 1-2; ion types, y, b, p; product ions from, ion 3; product ions to, last ion; min m/z, 50; max m/z, 2,000; ion match tolerance, 0.05; pick 6 product ions; minimum product ions, 3. MS1 filtering was set as follows: isotope peaks included, count; precursor mass analyzer, Orbitrap; peaks, 3; resolving power, 60,000 at m/z 400. MS/MS filtering was set as follows: acquisition method, DIA; product mass analyzer, Orbitrap; isolation scheme was set to the 24 mass windows used during analysis; resolving power, 30,000 at m/z 400. Retention time filtering was allowed to use any scans within 5 minutes of the MS/MS IDs. Data were searched against the UniProt reviewed human proteome (downloaded October 30, 2021) with up to 2 missed cleavages. 1 reverse sequence decoy was generated for every peptide. mProphet models were trained within Skyline to generate FDR estimations for identified precursors.

Protein Level Quantitation.

All quantitation was performed at the MS^1 level. All precursors were compiled and filtered to remove decoys and low-scoring peptides. Further filtering was applied to remove any redundant peptides attributed to the same protein in the same sample, keeping the most intense occurrence. Valid precursors were then confined to those peptides that were found in every sample (i.e. at least one technical duplicate of each biological replicate). The final number of peptides was then reduced to consider no more than the top 6 peptides from each protein. All peptides used for quantitation were found in every sample. Quantitation was performed by $\log_2$ transforming the precursor area and averaging all precursors for a given protein. Repeating this process for all

biological replicates provided N=9 datapoints for quantitation. Protein intensities were averaged across biological replicates and fold change was calculated as $intensity_{disease} - intensity_{control}$. Statistical significance was calculated via student's independent t-test.

Data and Code Availability.

All .raw mass spectrometry data files and retention time-corrected spectral libraries have been uploaded to the MassIVE repository with the primary accession code MSV000091165 (https://doi.org/doi:10.25345/C5S46HG56). All code and files needed to recreate the figures and analyses within the manuscript can be accessed at https://github.com/lingjunli-research/csf_dia_eval_lfq. Additional summary tables containing all retention times, peptide identities, protein quantity estimations, protein descriptions, and patient information have been provided as Supplemental Data 1-8.

Figures

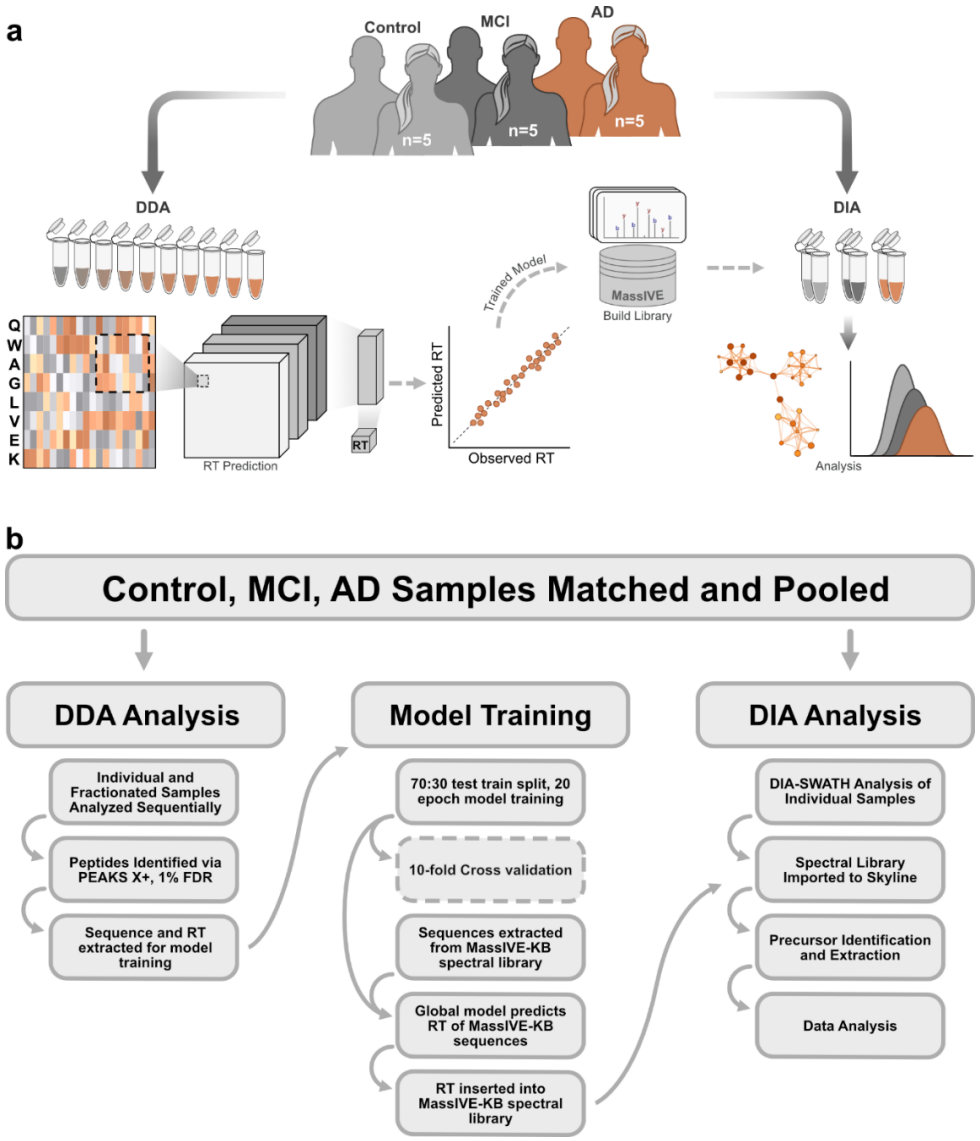


Figure 8.1 Workflow for sample agnostic spectral library construction.

a) Outline of parallel DDA and DIA analyses. Pooled cohorts were fractionated via high pH fractionated and analyzed via DDA. The resulting peptide sequences and retention times were used to train a capsule network model to predict retention times. The trained model then inferred retention times of all MassIVE-KB sequences. The final spectral library was used to inform downstream analysis of DIA experiments. b) step-by-step outline of the sample agnostic spectral library approach.

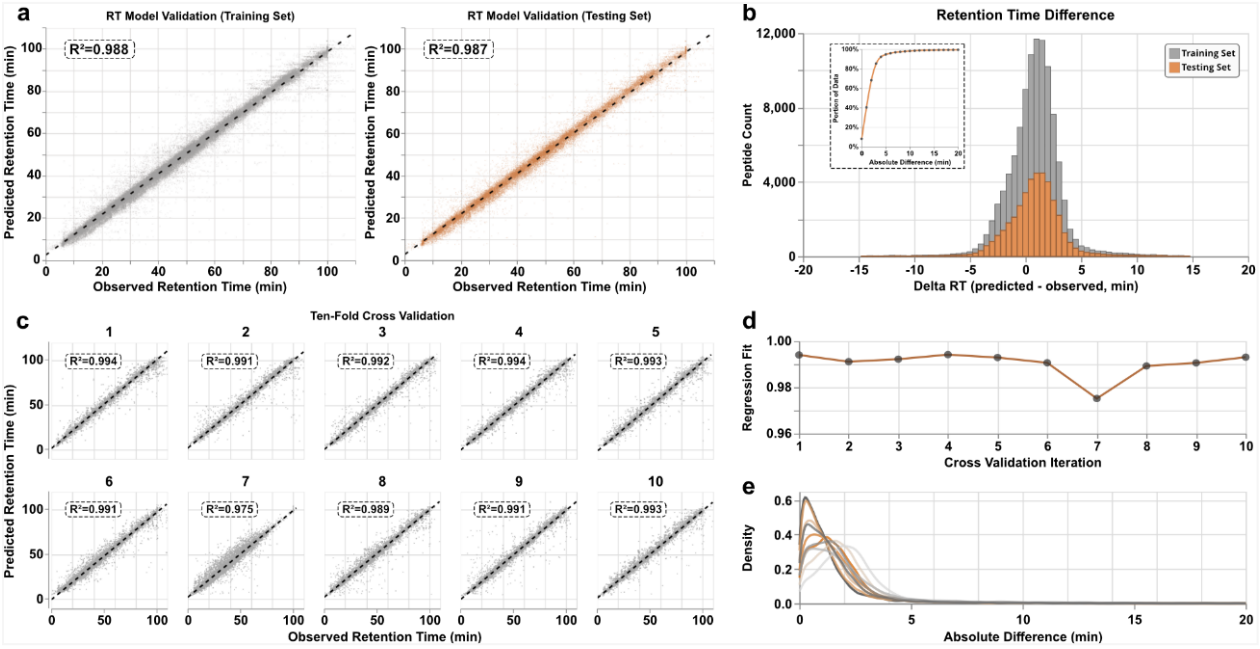


Figure 8.2 Validation of capsule network approach for retention time prediction.

a) Comparison of predicted and empirical retention times for all peptide sequences used for the testing set (left) and training set (right). b) Binned histogram denoting the difference between retention time predictions and their empirical value. (inset) Portion of all testing set predictions falling within a set time difference threshold. c) All iterations of ten-fold cross validation used to assess out-of-sample error and general applicability of method. Though some iterations demonstrate greater reproducibility than the global model, this is likely attributed to a larger training test set. d) Sum of least squares regression fit for all ten cross validation iterations. e) Density plots of all ten cross fold validation iterations demonstrating that all models predict ~95% of all retention times to within 5 minutes of the empirical value.

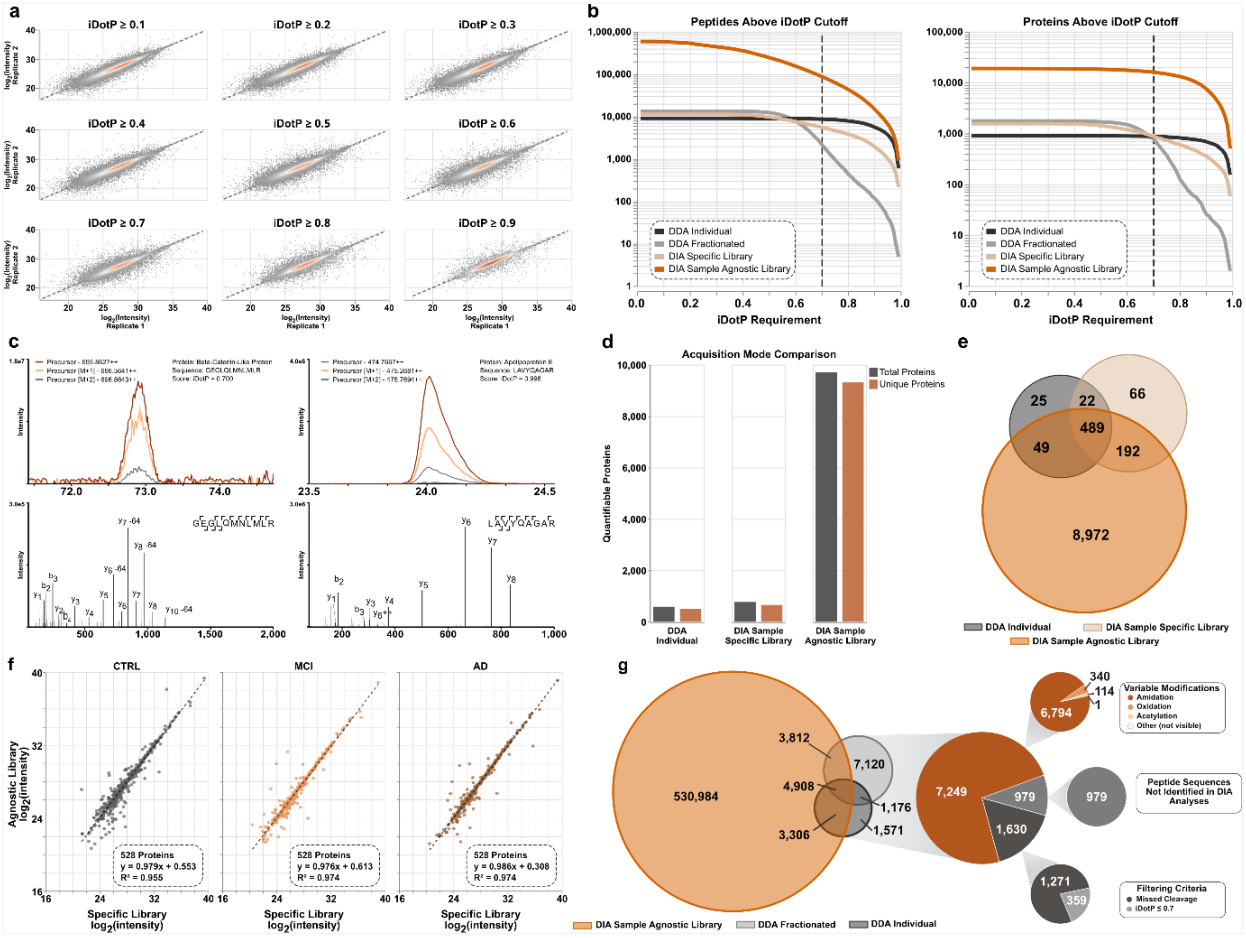


Figure 8.3 Investigation of sample agnostic library reliability and quality.

a) Intra-sample reproducibility of MS^1 protein quantitation. Each subplot displays all proteins that can be quantified when peptide precursors have an isotopic dot product greater than or equal to the shown value. b) Comparison of all datasets displaying peptide precursors and proteins above a given isotopic dot product. c) Representative MS^1 and MS^2 spectra of peptide precursors only identified through the sample agnostic spectral library. The quality of data at both the low and high end of our chosen cutoff reinforces the reliability of chosen metrics. d) Comparison of total and unique proteins identified through each analysis paradigm. 'Total proteins' represents all proteins potentially identified and 'unique proteins' are the number of protein groups evidenced with at least one proteotypic peptide. e) Overlap of protein groups identified in each analysis paradigm. f)

Correlation of MS^1 level protein quantitation when using the sample agnostic spectral library and the sample specific library. g) Overlap of unique peptide sequences identified through each analytical paradigm. Those sequences not found in the agnostic library are further broken down to address discrepancy.

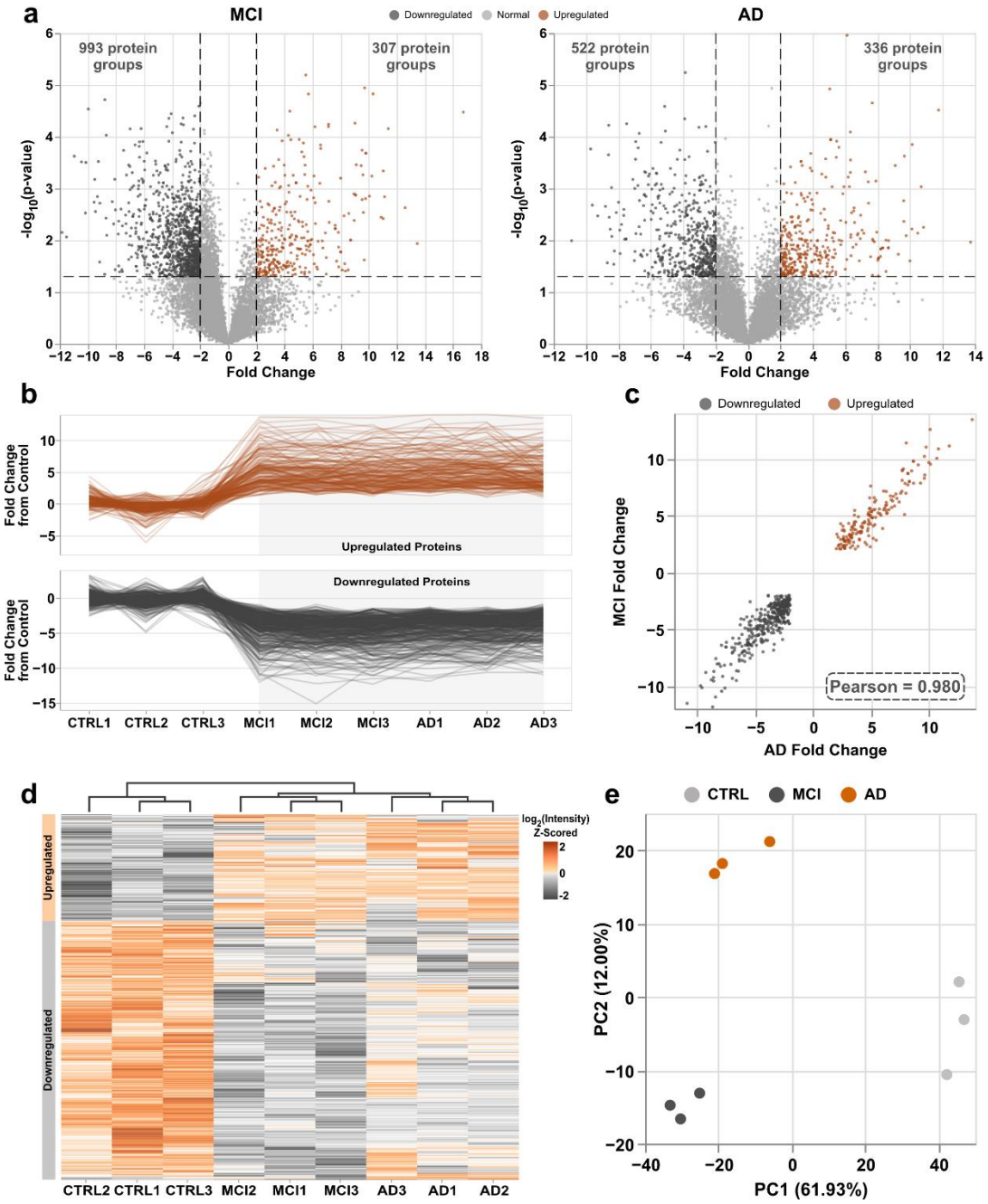


Figure 8.4 Quantitative evaluation of CSF facilitates disease state stratification.

a) Volcano plots representing fold change of protein expression and the -$\log_{10}$(statistical significance) determined through student's t-test. Only proteins with fold change ≥2 and P≤0.05 (-$\log_{10}$(P)≥1.3) are considered 'significantly dysregulated.' b) K-means clustering analysis highlighting the distinct dysregulation of proteins across disease states. c) Scatter plot displaying

difference in protein expression across disease states. The high Pearson correlation highlights the propensity for proteins dysregulated in MCI patients to also be dysregulated in AD. d) Column- and row-wise hierarchical clustering of protein-level expression, demonstrating separation of disease states and healthy control. e) Principal component analysis (PCA) of all cohort samples, based on protein level expression.

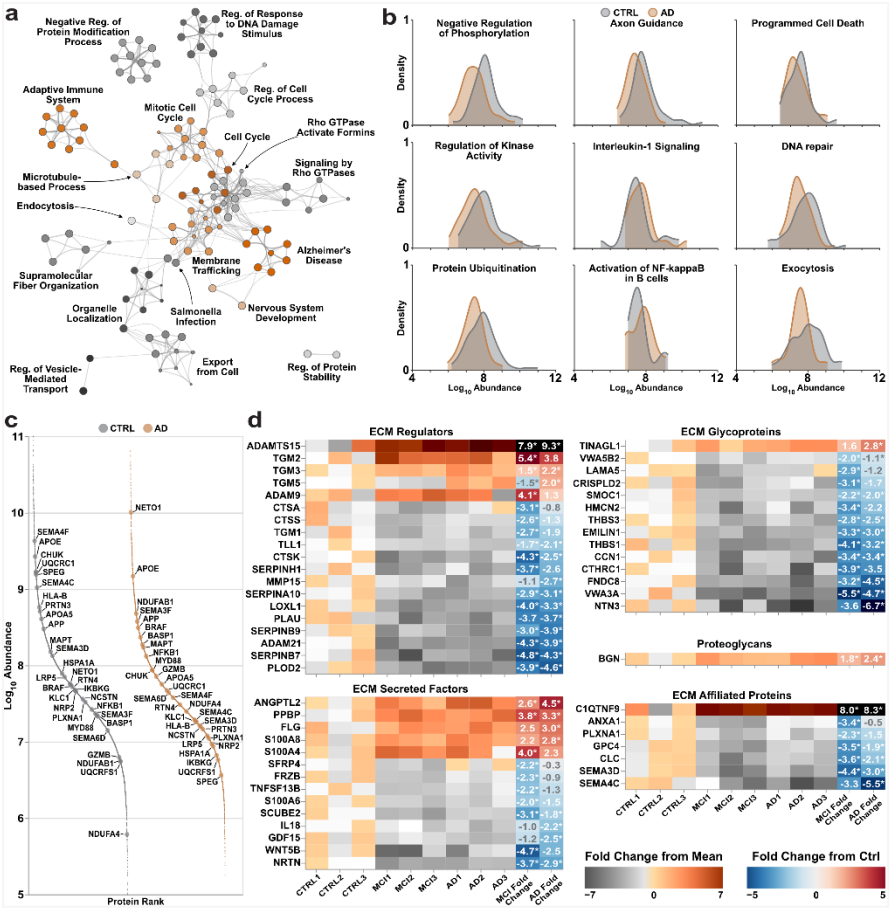


Figure 8.5 Proteome reorganization in neurological disease cohorts.

a) Metascape gene ontology protein network displaying putative pathways significantly downregulated across MCI and AD cohorts. Proteins are clustered in space according to similarity in pathway participation; clusters are named according to the most enriched term. Colors are arbitrary but serve to denote distinct clusters; original Metascape colors are preserved in Figure S8.4for clarity. b) Density plots of protein abundances associated with various biological pathways. AD and control plots are overlaid to display dysregulation. c) Ranked protein abundances for all proteins quantified; only AD and control groups are shown. Proteins with text are various species discussed in the main text or other known proteins of interest. d) Heatmap of extracellular matrix proteins; proteins are grouped according to Matrisome DB[65]. Fold changes are $\log_2$ and statistical significance is denoted by asterisks.

References

1. Breijyeh, Z.; Karaman, R., Comprehensive Review on Alzheimer's Disease: Causes and Treatment. *Molecules* **2020,** *25* (24).

2. DeTure, M. A.; Dickson, D. W., The neuropathological diagnosis of Alzheimer's disease. *Molecular Neurodegeneration* **2019,** *14* (1), 32.

3. Rizzi, L.; Rosset, I.; Roriz-Cruz, M., Global epidemiology of dementia: Alzheimer's and vascular types. *Biomed Res Int* **2014,** *2014*, 908915.

4. Bader, J. M.; Geyer, P. E.; Müller, J. B.; Strauss, M. T.; Koch, M.; Leypoldt, F.; Koertvelyessy, P.; Bittner, D.; Schipke, C. G.; Incesoy, E. I.; Peters, O.; Deigendesch, N.; Simons, M.; Jensen, M. K.; Zetterberg, H.; Mann, M., Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease. *Molecular Systems Biology* **2020,** *16* (6), e9356.

5. Johnson, E. C. B.; Dammer, E. B.; Duong, D. M.; Ping, L.; Zhou, M.; Yin, L.; Higginbotham, L. A.; Guajardo, A.; White, B.; Troncoso, J. C.; Thambisetty, M.; Montine, T. J.; Lee, E. B.; Trojanowski, J. Q.; Beach, T. G.; Reiman, E. M.; Haroutunian, V.; Wang, M.; Schadt, E.; Zhang, B.; Dickson, D. W.; Ertekin-Taner, N.; Golde, T. E.; Petyuk, V. A.; De Jager, P. L.; Bennett, D. A.; Wingo, T. S.; Rangaraju, S.; Hajjar, I.; Shulman, J. M.; Lah, J. J.; Levey, A. I.; Seyfried, N. T., Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nature Medicine* **2020,** *26* (5), 769-780.

6. 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* **2020,** *13*, 564446.

7. Meyer, J. G.; Schilling, B., Clinical applications of quantitative proteomics using targeted and untargeted data-independent acquisition techniques. *Expert Review of Proteomics* **2017,** *14* (5), 419-429.

8. Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M., DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods* **2020,** *17* (1), 41-44.

9. Sinitcyn, P.; Hamzeiy, H.; Salinas Soto, F.; Itzhak, D.; McCarthy, F.; Wichmann, C.; Steger, M.; Ohmayer, U.; Distler, U.; Kaspar-Schoenefeld, S.; Prianichnikov, N.; Yılmaz, Ş.; Rudolph, J. D.; Tenzer, S.; Perez-Riverol, Y.; Nagaraj, N.; Humphrey, S. J.; Cox, J., MaxDIA enables library-based and library-free data-independent acquisition proteomics. *Nature Biotechnology* **2021,** *39* (12), 1563-1573.

10. Tsou, C.-C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A.-C.; Nesvizhskii, A. I., DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nature Methods* **2015,** *12* (3), 258-264.

11. Wang, M.; Wang, J.; Carver, J.; Pullman, B. S.; Cha, S. W.; Bandeira, N., Assembling the Community-Scale Discoverable Human Proteome. *Cell Systems* **2018,** *7* (4), 412-421.e5.

12. Ma, C.; Ren, Y.; Yang, J.; Ren, Z.; Yang, H.; Liu, S., Improved Peptide Retention Time Prediction in Liquid Chromatography through Deep Learning. *Analytical Chemistry* **2018,** *90* (18), 10881-10888.

13. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010,** *26* (7), 966-8.

14. Jia, Z.; Wu, Q., Clustered Protocadherins Emerge as Novel Susceptibility Loci for Mental Disorders. *Front Neurosci* **2020,** *14*, 587819.

15. Wang, X. X.; Tan, M. S.; Yu, J. T.; Tan, L., Matrix metalloproteinases and their multiple roles in Alzheimer's disease. *Biomed Res Int* **2014,** *2014*, 908636.

16. Duits, F. H.; Hernandez-Guillamon, M.; Montaner, J.; Goos, J. D.; Montañola, A.; Wattjes, M. P.; Barkhof, F.; Scheltens, P.; Teunissen, C. E.; van der Flier, W. M., Matrix Metalloproteinases in Alzheimer's Disease and Concurrent Cerebral Microbleeds. *J Alzheimers Dis* **2015,** *48* (3), 711-20.

17. Zhang, X.; Zhu, M.; Jiang, X. L.; Liu, X.; Liu, X.; Liu, P.; Wu, X. X.; Yang, Z. W.; Qin, T., P-selectin glycoprotein ligand 1 deficiency prevents development of acute pancreatitis by attenuating leukocyte infiltration. *World J Gastroenterol* **2020,** *26* (41), 6361-6377.

18. Järemo, P.; Milovanovic, M.; Buller, C.; Nilsson, S.; Winblad, B., P-selectin paradox and dementia of the Alzheimer type: Circulating P-selectin is increased but platelet-bound P-selectin after agonist provocation is compromised. *Scandinavian Journal of Clinical and Laboratory Investigation* **2013,** *73* (2), 170-174.

19. Vassar, R.; Kandalepas, P. C., The β-secretase enzyme BACE1 as a therapeutic target for Alzheimer's disease. *Alzheimer's Research & Therapy* **2011,** *3* (3), 20.

20. Gao, L.; Zhang, Y.; Sterling, K.; Song, W., Brain-derived neurotrophic factor in Alzheimer's disease and its pharmaceutical potential. *Translational Neurodegeneration* **2022,** *11* (1), 4.

21. Vu, H. T.; Akatsu, H.; Hashizume, Y.; Setou, M.; Ikegami, K., Increase in α-tubulin modifications in the neuronal processes of hippocampal neurons in both kainic acid-induced epileptic seizure and Alzheimer's disease. *Sci Rep* **2017,** *7,* 40205.

22. Hausrat, T. J.; Janiesch, P. C.; Breiden, P.; Lutz, D.; Hoffmeister-Ullerich, S.; Hermans-Borgmeyer, I.; Failla, A. V.; Kneussel, M., Disruption of tubulin-alpha4a polyglutamylation prevents aggregation of hyper-phosphorylated tau and microglia activation in mice. *Nature Communications* **2022,** *13* (1), 4192.

23. Reed, M. J.; Damodarasamy, M.; Pathan, J. L.; Chan, C. K.; Spiekerman, C.; Wight, T. N.; Banks, W. A.; Day, A. J.; Vernon, R. B.; Keene, C. D., Increased Hyaluronan and TSG-6 in Association with Neuropathologic Changes of Alzheimer's Disease. *J Alzheimers Dis* **2019,** *67* (1), 91-102.

24. Polis, B.; Karasik, D.; Samson, A. O., Alzheimer's disease as a chronic maladaptive polyamine stress response. *Aging (Albany NY)* **2021,** *13* (7), 10770-10795.

25. Banno, M.; Koide, T.; Aleksic, B.; Yamada, K.; Kikuchi, T.; Kohmura, K.; Adachi, Y.; Kawano, N.; Kushima, I.; Ikeda, M.; Inada, T.; Yoshikawa, T.; Iwata, N.; Ozaki, N., A case control association study and cognitive function analysis of neuropilin and tolloid-like 1 gene and schizophrenia in the Japanese population. *PLoS One* **2011,** *6* (12), e28929.

26. Wingo, A. P.; Dammer, E. B.; Breen, M. S.; Logsdon, B. A.; Duong, D. M.; Troncosco, J. C.; Thambisetty, M.; Beach, T. G.; Serrano, G. E.; Reiman, E. M., Large-scale proteomic analysis of human brain identifies proteins associated with cognitive trajectory in advanced age. *Nature communications* **2019,** *10* (1), 1-14.

27. Aloni, E.; Oni-Biton, E.; Tsoory, M.; Moallem, D. H.; Segal, M., Synaptopodin Deficiency Ameliorates Symptoms in the 3xTg Mouse Model of Alzheimer's Disease. *J Neurosci* **2019,** *39* (20), 3983-3992.

28. Mostafavi, S.; Gaiteri, C.; Sullivan, S. E.; White, C. C.; Tasaki, S.; Xu, J.; Taga, M.; Klein, H.-U.; Patrick, E.; Komashko, V.; McCabe, C.; Smith, R.; Bradshaw, E. M.; Root, D. E.; Regev, A.; Yu, L.; Chibnik, L. B.; Schneider, J. A.; Young-Pearse, T. L.; Bennett, D. A.; De Jager, P. L., A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nature Neuroscience* **2018,** *21* (6), 811-819.

29. McCorkindale, A. N.; Patrick, E.; Duce, J. A.; Guennewig, B.; Sutherland, G. T., The Key Factors Predicting Dementia in Individuals With Alzheimer's Disease-Type Pathology. *Frontiers in Aging Neuroscience* **2022,** *14*.

30. Burgh, R. v. d.; Meeldijk, J.; Jongeneel, L.; Frenkel, J.; Bovenschen, N.; Gijn, M. v.; Boes, M., Reduced serpinB9-mediated caspase-1 inhibition can contribute to autoinflammatory disease. *Oncotarget* **2016,** *7* (15).

31. Lumkwana, D.; Peddie, C.; Kriel, J.; Michie, L. L.; Heathcote, N.; Collinson, L.; Kinnear, C.; Loos, B., Investigating the Role of Spermidine in a Model System of Alzheimer's Disease Using Correlative Microscopy and Super-resolution Techniques. *Frontiers in Cell and Developmental Biology* **2022,** *10*.

32. Sampaio, T. B.; Savall, A. S.; Gutierrez, M. E. Z.; Pinton, S., Neurotrophic factors in Alzheimer's and Parkinson's diseases: implications for pathogenesis and therapy. *Neural Regen Res* **2017,** *12* (4), 549-557.

33. Samiotaki, G.; Acosta, C.; Wang, S.; Konofagou, E. E., Enhanced delivery and bioactivity of the neurturin neurotrophic factor through focused ultrasound-mediated blood--brain barrier opening in vivo. *J Cereb Blood Flow Metab* **2015,** *35* (4), 611-22.

34. Mitra, S.; Behbahani, H.; Eriksdotter, M., Innovative Therapy for Alzheimer's Disease-With Focus on Biodelivery of NGF. *Frontiers in Neuroscience* **2019,** *13*.

35. Kang, S.; Byun, J.; Son, S. M.; Mook-Jung, I., Thrombospondin-1 protects against Aβ-induced mitochondrial fragmentation and dysfunction in hippocampal cells. *Cell Death Discovery* **2018,** *4* (1), 31.

36. Son, S. M.; Nam, D. W.; Cha, M. Y.; Kim, K. H.; Byun, J.; Ryu, H.; Mook-Jung, I., Thrombospondin-1 prevents amyloid beta-mediated synaptic pathology in Alzheimer's disease. *Neurobiol Aging* **2015,** *36* (12), 3214-3227.

37. Clemmensen, C.; Aznar, S.; Knudsen, G. M.; Klein, A. B., The microtubule-associated protein 1A (MAP1A) is an early molecular target of soluble Aβ-peptide. *Cell Mol Neurobiol* **2012,** *32* (4), 561-6.

38. Harrison, N.; Koo, C. Z.; Tomlinson, M. G., Regulation of ADAM10 by the TspanC8 Family of Tetraspanins and Their Therapeutic Potential. *International Journal of Molecular Sciences* **2021,** *22* (13), 6707.

39. Dines, M.; Lamprecht, R., The Role of Ephs and Ephrins in Memory Formation. *Int J Neuropsychopharmacol* **2016,** *19* (4).

40. Santiago-Mujika, E.; Luthi-Carter, R.; Giorgini, F.; Kalaria, R. N.; Mukaetova-Ladinska, E. B., Tubulin and Tubulin Posttranslational Modifications in Alzheimer's Disease and Vascular Dementia. *Frontiers in Aging Neuroscience* **2021,** *13*.

41. Toriyama, M.; Shimada, T.; Kim, K. B.; Mitsuba, M.; Nomura, E.; Katsuta, K.; Sakumura, Y.; Roepstorff, P.; Inagaki, N., Shootin1: A protein involved in the organization of an asymmetric signal for neuronal polarization. *J Cell Biol* **2006,** *175* (1), 147-57.

42. Kallionpää, R. A.; Valtanen, M.; Auranen, K.; Uusitalo, E.; Rinne, J. O.; Peltonen, S.; Peltonen, J., Increased risk for dementia in neurofibromatosis type 1. *Genetics in Medicine* **2021,** *23* (11), 2219-2222.

43. Lin, M.-Y.; Cheng, X.-T.; Tammineni, P.; Xie, Y.; Zhou, B.; Cai, Q.; Sheng, Z.-H., Releasing Syntaphilin Removes Stressed Mitochondria from Axons Independent of Mitophagy under Pathophysiological Conditions. *Neuron* **2017,** *94* (3), 595-610.e6.

44. Bettcher, B. M.; Tansey, M. G.; Dorothée, G.; Heneka, M. T., Peripheral and central immune system crosstalk in Alzheimer disease — a research prospectus. *Nature Reviews Neurology* **2021,** *17* (11), 689-701.

45. Rape, M., Ubiquitylation at the crossroads of development and disease. *Nature Reviews Molecular Cell Biology* **2018,** *19* (1), 59-70.

46. Henriques, A. G.; Müller, T.; Oliveira, J. M.; Cova, M.; da Cruz e Silva, C. B.; da Cruz e Silva, O. A. B., Altered protein phosphorylation as a resource for potential AD biomarkers. *Scientific Reports* **2016,** *6* (1), 30319.

47. Wegmann, S.; Biernat, J.; Mandelkow, E., A current view on Tau protein phosphorylation in Alzheimer's disease. *Current Opinion in Neurobiology* **2021,** *69*, 131-138.

48. Rao, C. V.; Farooqui, M.; Asch, A. S.; Yamada, H. Y., Critical role of mitosis in spontaneous late-onset Alzheimer's disease; from a Shugoshin 1 cohesinopathy mouse model. *Cell Cycle* **2018,** *17* (19-20), 2321-2334.

49. Liu, R. M., Aging, Cellular Senescence, and Alzheimer's Disease. *Int J Mol Sci* **2022,** *23* (4).

50. Tönnies, E.; Trushina, E., Oxidative Stress, Synaptic Dysfunction, and Alzheimer's Disease. *J Alzheimers Dis* **2017,** *57* (4), 1105-1121.

51. Musunuri, S.; Khoonsari, P. E.; Mikus, M.; Wetterhall, M.; Häggmark-Mänberg, A.; Lannfelt, L.; Erlandsson, A.; Bergquist, J.; Ingelsson, M.; Shevchenko, G.; Nilsson, P.; Kultima, K., Increased Levels of Extracellular Microvesicle Markers and Decreased Levels of Endocytic/Exocytic Proteins in the Alzheimer's Disease Brain. *J Alzheimers Dis* **2016,** *54* (4), 1671-1686.

52. Perdigao, C.; Barata, M. A.; Araujo, M. N.; Mirfakhar, F. S.; Castanheira, J.; Guimas Almeida, C., Intracellular Trafficking Mechanisms of Synaptic Dysfunction in Alzheimer's Disease. *Front Cell Neurosci* **2020,** *14*, 72.

53. Giannos, P.; Prokopidis, K.; Raleigh, S. M.; Kelaiditi, E.; Hill, M., Altered mitochondrial microenvironment at the spotlight of musculoskeletal aging and Alzheimer's disease. *Scientific Reports* **2022,** *12* (1), 11290.

54. Chen, F.; Bai, J.; Zhong, S.; Zhang, R.; Zhang, X.; Xu, Y.; Zhao, M.; Zhao, C.; Zhou, Z., Molecular Signatures of Mitochondrial Complexes Involved in Alzheimer's Disease via Oxidative Phosphorylation and Retrograde Endocannabinoid Signaling Pathways. *Oxidative Medicine and Cellular Longevity* **2022,** *2022*, 9565545.

55. Ma, S. L.; Tang, N. L. S.; Lam, L. C. W., Association of gene expression and methylation of UQCRC1 to the predisposition of Alzheimer's disease in a Chinese population. *Journal of Psychiatric Research* **2016,** *76*, 143-147.

56. Bolduc, D. M.; Montagna, D. R.; Gu, Y.; Selkoe, D. J.; Wolfe, M. S., Nicastrin functions to sterically hinder γ-secretase–substrate interactions driven by substrate transmembrane domain. *Proceedings of the National Academy of Sciences* **2016,** *113* (5), E509-E518.

57. Chen, X.-Q.; Das, U.; Park, G.; Mobley, W. C., Normal levels of KIF5 but reduced KLC1 levels in both Alzheimer disease and Alzheimer disease in Down syndrome: evidence suggesting defects in anterograde transport. *Alzheimer's Research & Therapy* **2021,** *13* (1), 59.

58. Chung, D.; Shum, A.; Caraveo, G., GAP-43 and BASP1 in Axon Regeneration: Implications for the Treatment of Neurodegenerative Diseases. *Front Cell Dev Biol* **2020,** *8*, 567537.

59. Carulli, D.; de Winter, F.; Verhaagen, J., Semaphorins in Adult Nervous System Plasticity and Disease. *Front Synaptic Neurosci* **2021,** *13*, 672891.

60. Balabanski, L.; Serbezov, D.; Atanasoska, M.; Karachanak-Yankova, S.; Hadjidekova, S.; Nikolova, D.; Boyanova, O.; Staneva, R.; Vazharova, R.; Mihailova, M.; Damyanova, V.; Nesheva, D.; Belejanska, D.; Mehrabian, S.; Traykov, L.; Toncheva, D., Rare genetic variants prioritize molecular pathways for semaphorin interactions in Alzheimer's disease patients. *Biotechnology & Biotechnological Equipment* **2021,** *35* (1), 1256-1262.

61. Nakanishi, Y.; Kang, S.; Kumanogoh, A., Neural guidance factors as hubs of immunometabolic cross-talk. *International Immunology* **2021,** *33* (12), 749-754.

62. Berndt, J. D.; Halloran, M. C., Semaphorin 3d promotes cell proliferation and neural crest cell development downstream of TCF in the zebrafish hindbrain. *Development* **2006,** *133* (20), 3983-92.

63. Sahay, A.; Kim, C.-H.; Sepkuty, J. P.; Cho, E.; Huganir, R. L.; Ginty, D. D.; Kolodkin, A. L., Secreted Semaphorins Modulate Synaptic Transmission in the Adult Hippocampus. *The Journal of Neuroscience* **2005,** *25* (14), 3613-3620.

64. Simonetti, M.; Paldy, E.; Njoo, C.; Bali, K. K.; Worzfeld, T.; Pitzer, C.; Kuner, T.; Offermanns, S.; Mauceri, D.; Kuner, R., The impact of Semaphorin 4C/Plexin-B2

signaling on fear memory via remodeling of neuronal and synaptic morphology. *Molecular Psychiatry* **2021,** *26* (4), 1376-1398.

65. Shao, X.; Taha, I. N.; Clauser, K. R.; Gao, Y.; Naba, A., MatrisomeDB: the ECM-protein knowledge database. *Nucleic Acids Research* **2019,** *48* (D1), D1136-D1144.

66. Sun, Y.; Xu, S.; Jiang, M.; Liu, X.; Yang, L.; Bai, Z.; Yang, Q., Role of the Extracellular Matrix in Alzheimer's Disease. *Front Aging Neurosci* **2021,** *13*, 707466.

67. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A. H.; Tanaseichuk, O.; Benner, C.; Chanda, S. K., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* **2019,** *10* (1), 1523.

68. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003,** *13* (11), 2498-504.

Supplemental Tables

For brevity, supplementary data tables have not been included here. The tables are included as part of the submitted manuscript and may be found online or may be obtained directly through directly contact the authors of this chapter.

Supplemental Figures

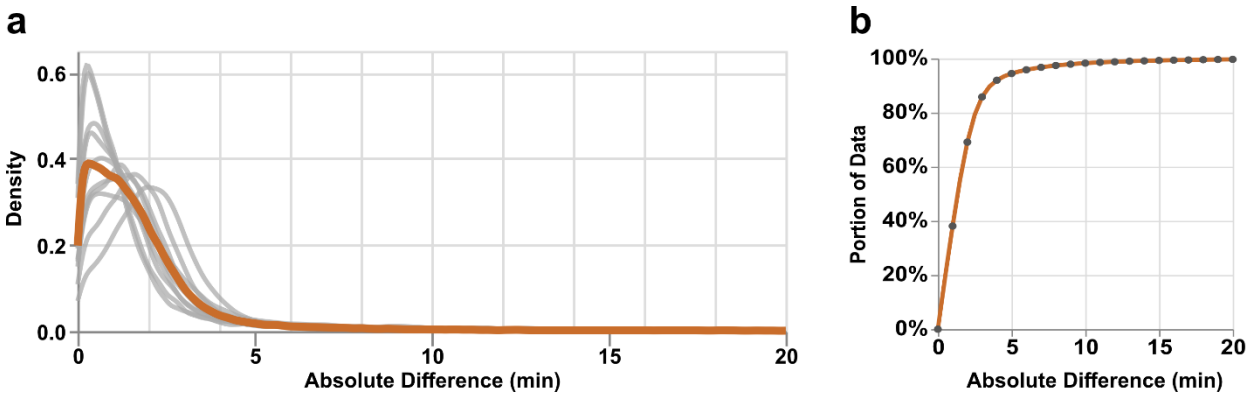


Figure S8.1 Average retention time predictions across all cross validations.

a) Density plots of all ten cross validation iterations (gray) displaying the distribution of retention time prediction error. Considering all predictions from all cross validations, the average density is shown in orange. b) Plot displaying the portion of all retention time predictions from all cross validations that fall within a given error window.

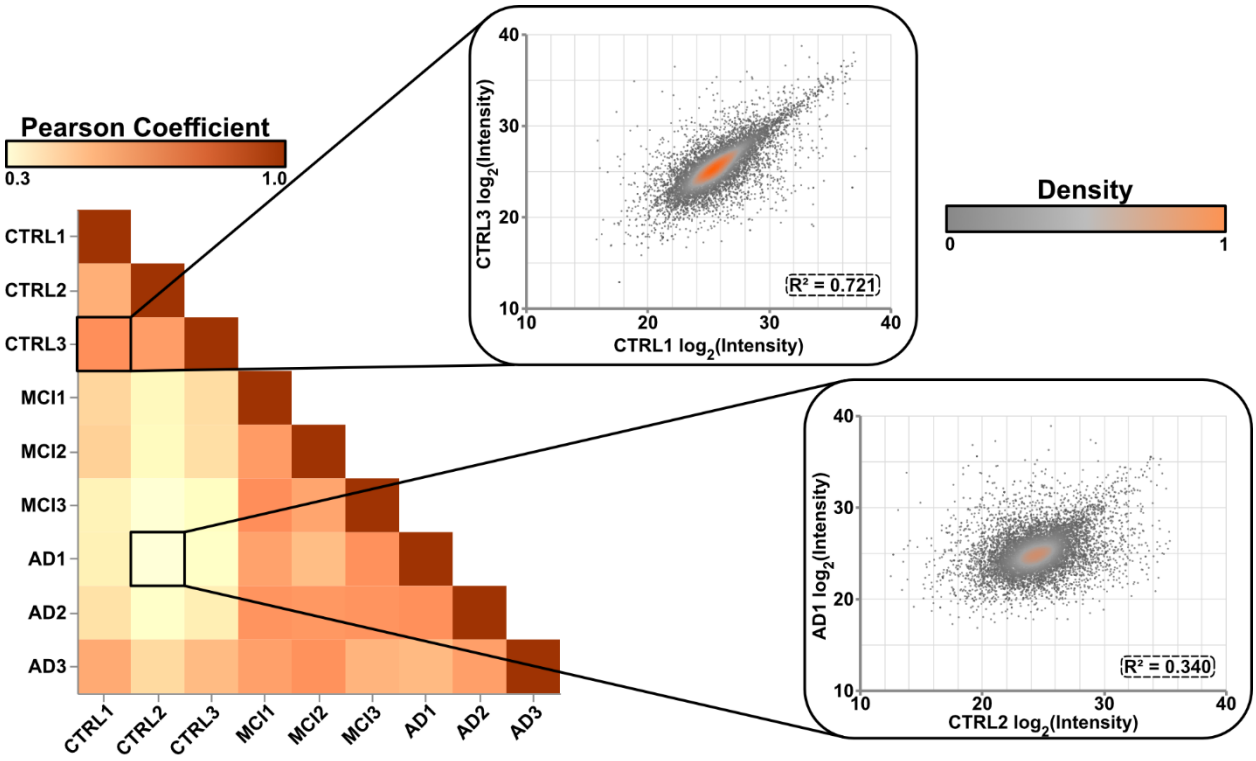


Figure S8.2 Correlation within and across patient cohorts.

Pearson correlation between all samples analyzed within the experiment, demonstrating the conservation of expression profiles between healthy and disease state cohorts. The color coding of each box within the heatmap corresponds to the Pearson coefficient between the row- and column-indexed samples, calculated based on the protein-level expression. The two called out density plots represent the protein intensity correlation across the compared samples; Pearson coefficients are denoted within each density plot.

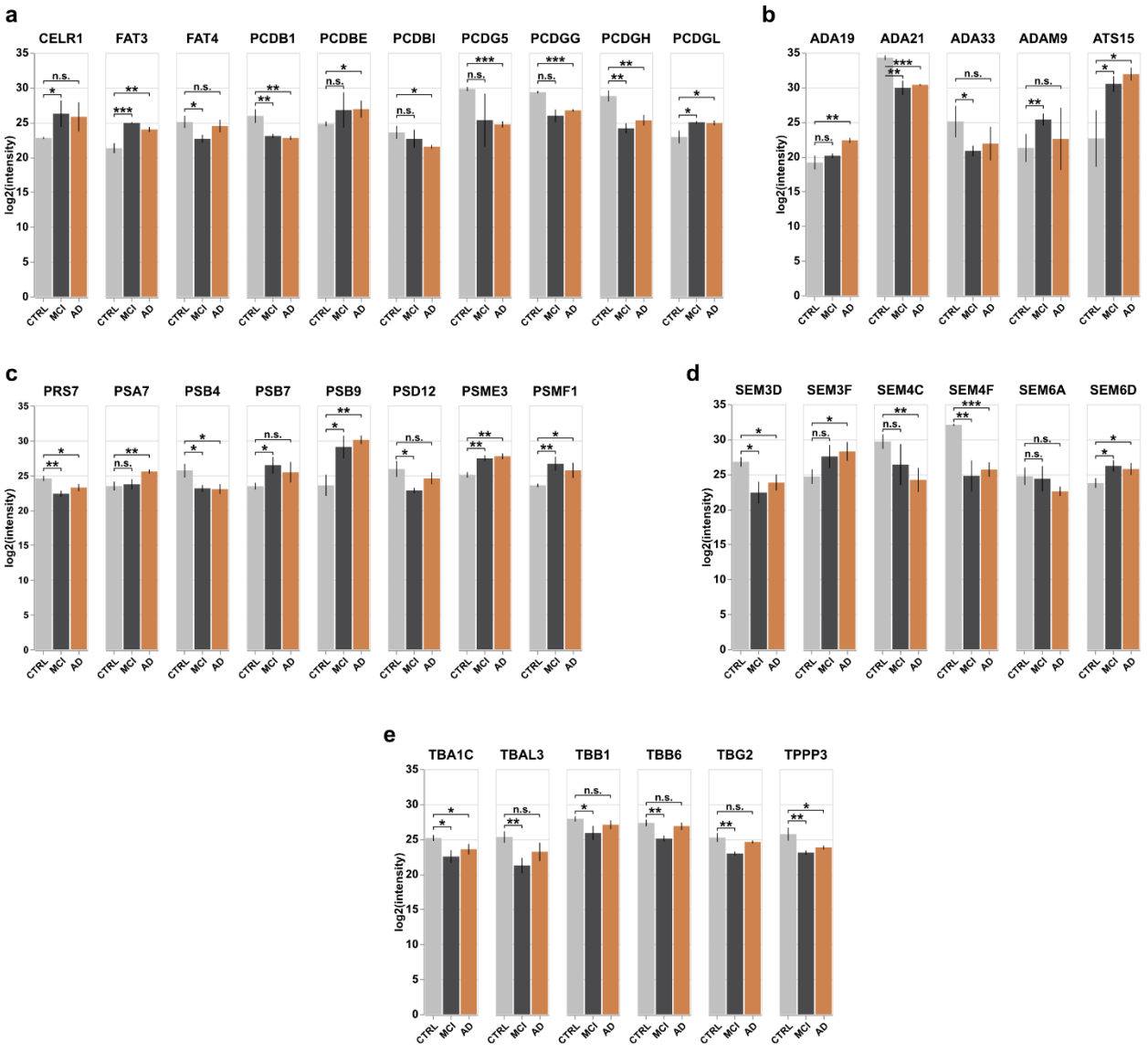


Figure S8.3 Conserved dysregulation across protein families.

Bar plots displaying the average log2 intensity of protein expression across all three patient cohorts; error bars represent the standard deviation. Among all dysregulated species, significant populations of the cadherin and protocadherin (a), disentegrin and metalloproteinase (b), proteasomal component/subunit (c), and semaphorin (d) families were consistently dysregulated. Whereas expression of these four protein groups across MCI and AD cohorts seems relatively independent of disease state, tubulins (e) show consistent dysregulation across cohorts; tubulins are notably downregulation in MCI cohorts and show slight recovery in AD.

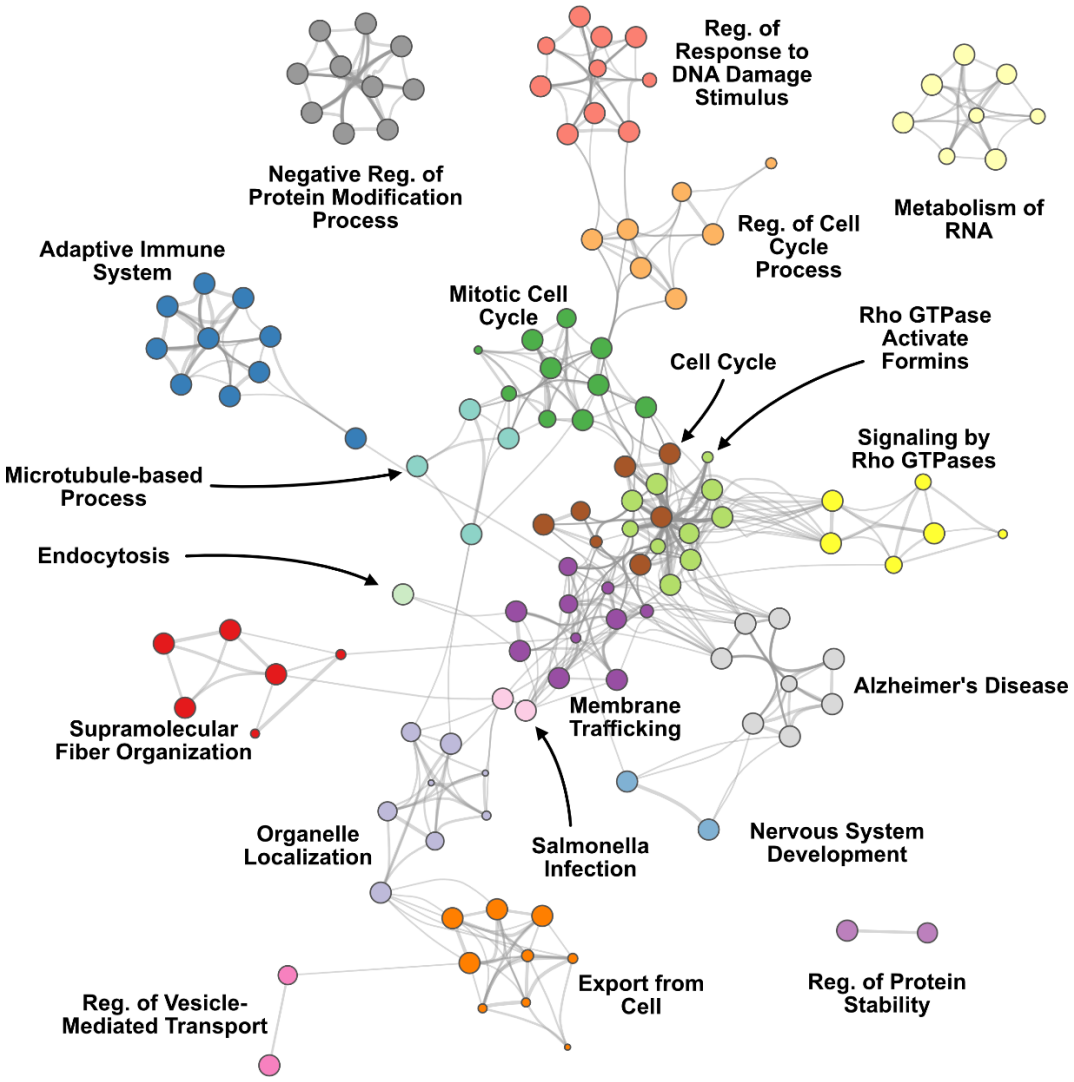


Figure S8.4 Metascape gene ontology protein network.

Network displaying the putative pathways significantly downregulated across MCI and AD samples. This network displays the same information as that shown in Figure 8.5a within the main text; this representation preserves the distinct, unedited cluster colors to provide clarity.

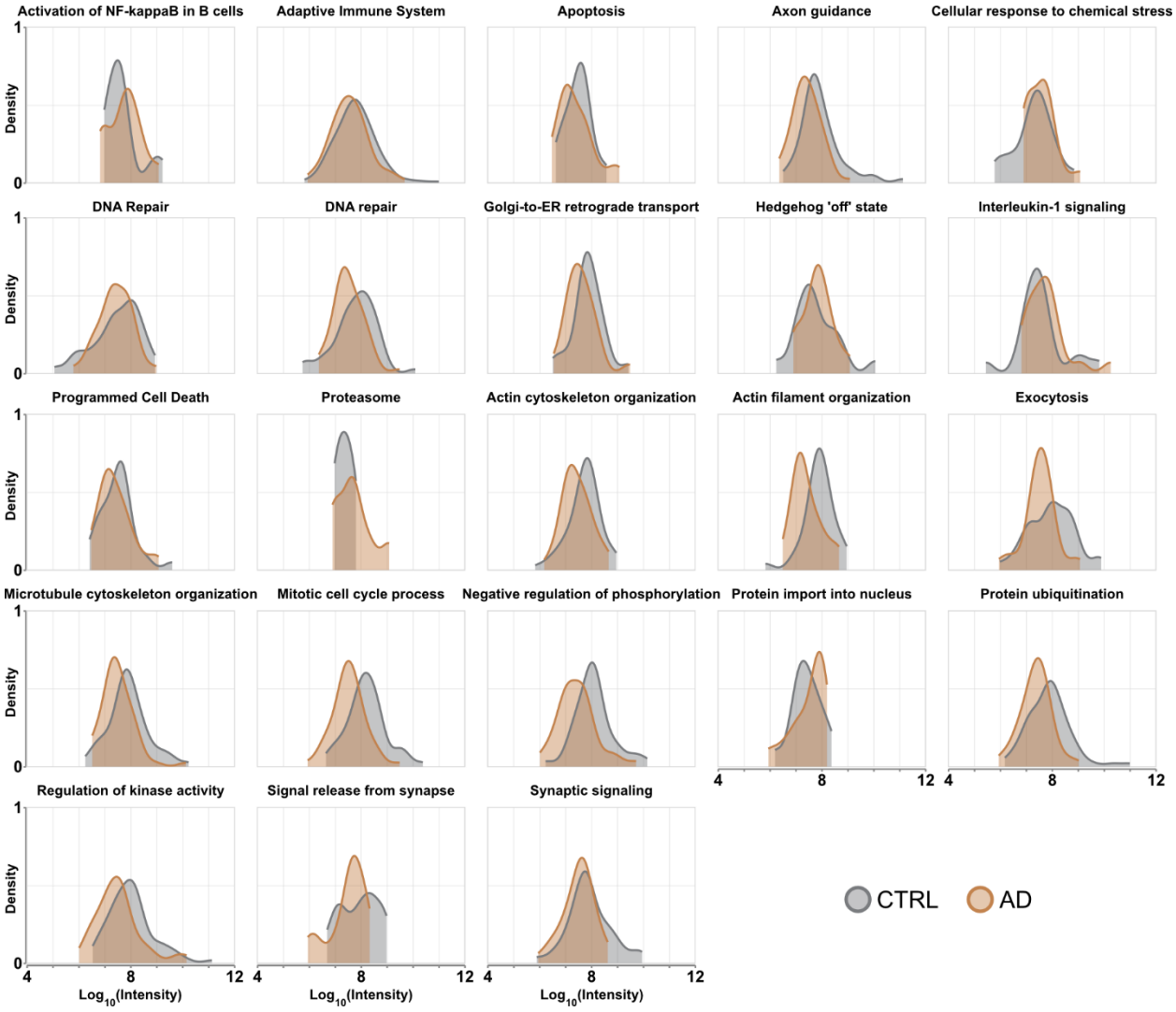


Figure S8.5 Gene ontology expression changes between control and AD cohorts.

Density plots represent the log10 protein-level intensity of all proteins associated with a given biological process (as determined by Metascape). Pathways and processes represented are those discussed directly or tangentially within the main text, as well as other pathways of interest not directly discussed within.

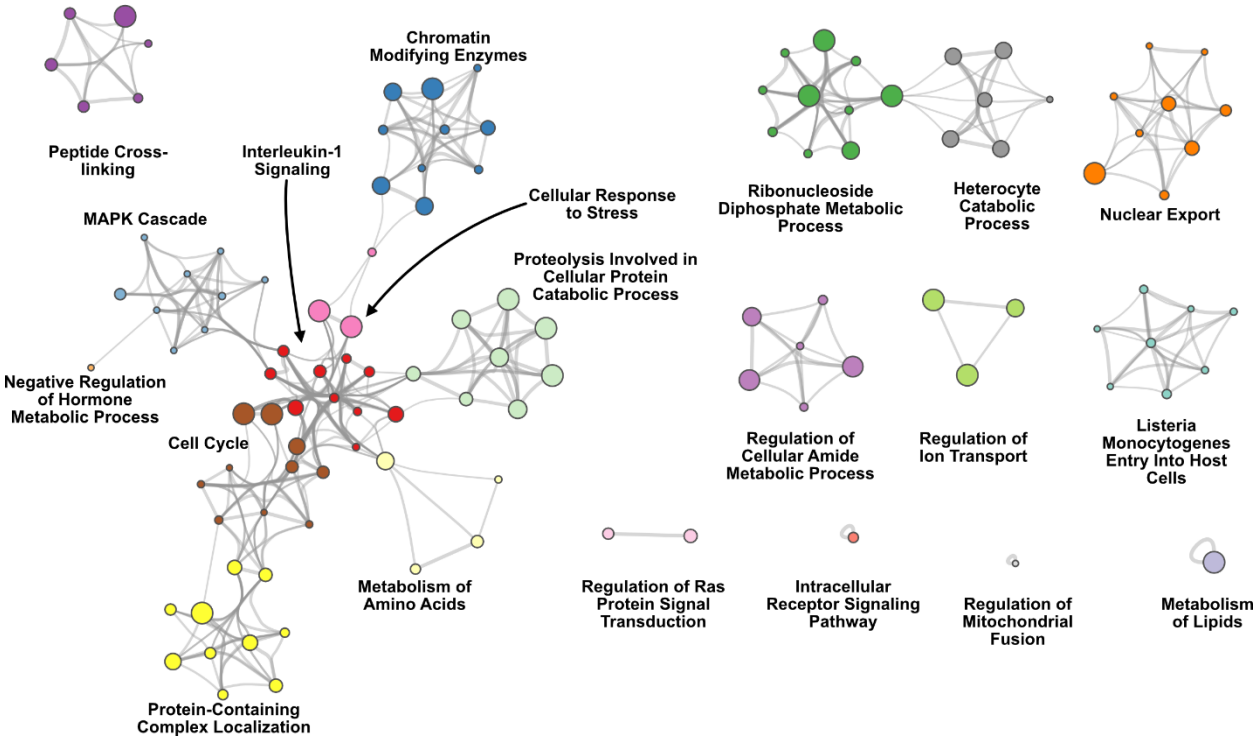


Figure S8.6 Pathways and processes upregulated across MCI and AD cohorts.

Metascape gene ontology protein networks displaying those proteins shown to be significantly upregulated in either MCI or AD patient cohorts. Similar to other protein networks, proteins are clustered in space according to known participation within shared protein groups, grouped according to similarity, and groups are labeled according to the most significantly enriched gene ontology term.

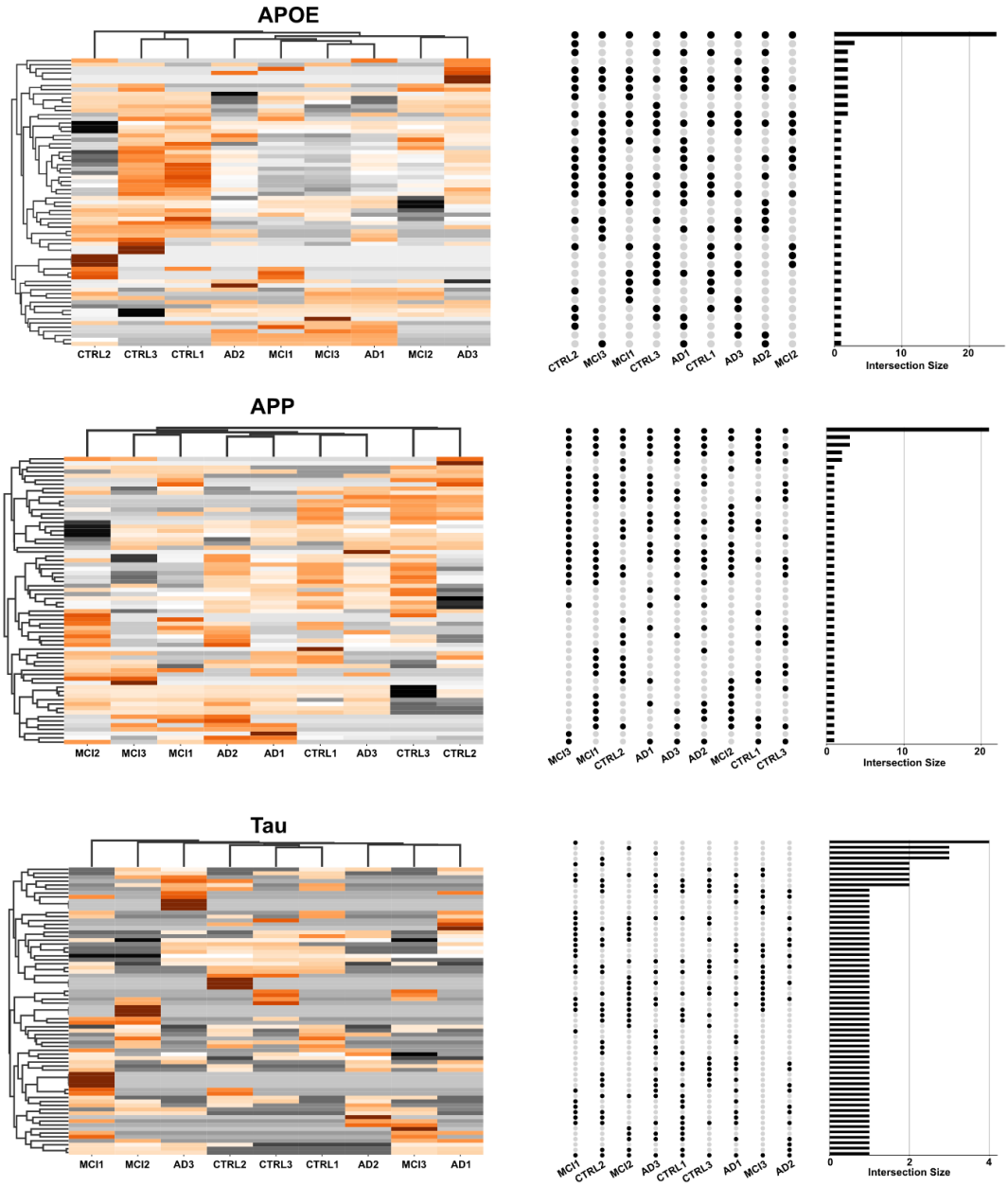


Figure S8.7 Peptide-level expression differences of common AD hallmarks.

Left) Hierarchical clustering heatmaps displaying peptide-level column- and row-wise clustering of patient cohorts, as determined for each hallmark protein. Missing values are zero-filled. While some slight clustering is possible based on peptide expression levels, the prevalence of missing values and uniform expression prevent any quantitative distinction at the protein level. Right) Upset plots displaying the number of peptides identified between groups. Black dots to the left

represent individual groups and bars to the right represent the number of peptides shared in those respective groups.

Chapter 9: High-End Ion Mobility Mass Spectrometry: A Current Review of Analytical Capacity in Omics Applications and Structural Investigations

Delafield, D. G., Lu, G., Kaminsky, C. J., & Li, L. (2022). High-end ion mobility mass spectrometry: A current review of analytical capacity in omics applications and structural investigations. *TrAC Trends in Analytical Chemistry*, *157*, 116761.

Abstract

Mass spectrometry-based biomolecular analyses have become permanent fixtures of academic, industrial, and clinical research settings. The rise in utilization of mass spectrometry has, in turn, spurred on a technological arms race, with every major vendor seeking to provide instrumentation that is more sensitive, higher in resolution, or may otherwise offer fundamental advantages during analysis. Enabling higher sensitivity, increased instrumental duty cycle, reduced analysis time and lower sample requirements, gas phase ion separation techniques now provide a fourth dimension of analysis, enabling rapid structural characterization and high throughput -omics profiling in a single run. Presented here is a current review of the latest iterations and applications of high-end ion-mobility enabled instrumentation, the Agilent 6560 IM-QTOF, Waters Cyclic, Bruker timsTOF, and Thermo FAIMS Pro instrument platforms. Describing their engineering developments and analytical success over the past two decades, we highlight notable advantages and considerations for novice and experienced biomolecular researchers alike.

Introduction

Having recently garnered significant attention and invigorated utilization, the past six decades of instrumental development have given way to a gilded era of Ion Mobility Spectrometry (IMS). With the earliest reports of ion mobility provided over a century ago[1] and the first published instrumental iteration nearing its sixtieth anniversary[2], the analytical capacity and breadth of meaningful applications have long been hindered by the rate of technical development. With Drift Tube (DTIMS), Field Asymmetric Waveform (FAIMS), and Traveling Wave Ion Mobility (TWIMS) arriving nearly twenty years apart – followed soon by Trapped Ion Mobility (TIMS) – these ion separation modalities have long been relegated to niche research focus and the analysis of structural conformation.

Within the past ten years, there has been a fundamental shift in IM-based biomolecular analyses. While the potential improvements in discovery -omics analyses that may be found through the addition of ion separation regimes have long been suggested, it was not until the commercialization of proteomics-specific IM instrumentation in the mid-2010s (Figure 9.1) that these analytical improvements were realized, paving the way for current high-end ion mobility instrumentation. “High-end” ion mobility instrumentation, as discussed below, refers to the latest iterations of four unique IM paradigms that directly augment biomolecular investigation by demonstrating significant improvements in sensitivity and sample coverage or offering unparalleled success in analyte differentiation and structural characterization through high gas-phase resolution (R>200). While this generation of instrumentation may be considered a competitive advantage in the pursuit of biological and biomolecular insight, these flagship instruments have each established their own niche advantages, use cases, and drawbacks. (Table 1) These unique instrumental capabilities, along with the technological innovations that make them

possible, are described within this review. Surveying recent literature for four cutting-edge instrumental platforms – the Agilent 6560 DTIMS, Waters Cyclic IMS, Bruker timsTOF, and Thermo FAIMS Pro – it becomes immediately clear that targeted and untargeted -omics investigations comprise an increasingly large sector of utility in IMS research, with structural investigations continuing to be a paradigm-defining application. Given this reality, we will confine our discussion to these areas to provide broad relevance with those more unique IMS applications being worthy of their own, independent review. Here we seek to provide readers with heuristic guidance in experimental design, as well as highlight analytical strengths of four IMS paradigms that will facilitate future biomolecular analyses.

Drift Tube Ion Mobility Spectrometry (DTIMS)

Among various ion mobility paradigms, Drift Tube Ion Mobility Spectrometry (DTIMS) is often considered fundamental, as it was the earliest developed. The distinctive feature of DTIMS is the uniform electric field applied across the drift tube, which is filled with a neutral buffer gas such as nitrogen or helium. This system, then, can directly measure the amount by which an ion is slowed due to collisions with the carrier gas as it passes through the tube. The low, uniform electric field used in DTIMS is consistent with the classical Mason-Schamp equation[3]. If the experiment parameters are precisely controlled and recorded, DTIMS is the only ion mobility paradigm that can provide precise collisional cross section (CCS) measurement without the need for calibration ions[4]. The long history of DTIMS development has been discussed in detail elsewhere[5, 6] but in this review we will confine our discussion to a cutting-edge commercial DTIMS instrument, the Agilent 6560 Ion Mobility Q-TOF. The successful application of DTIMS in structural investigation and omics-related studies over the past two decades are also discussed below.

Background and Engineering Developments

In 1998, with the advancement of high-speed electrical components, David Clemmer and colleagues built the first drift tube TOF-coupled ion mobility mass spectrometer, which enabled the observation of drift time and *m/z* value for the whole ion system[7]. In the following years, Richard Smith's group integrated the electrodynamic ion funnels at the front and at the rear of the drift tube[8, 9]. The introduction of ion funnels improves ion accumulation efficiency and enhances detection sensitivity. In 2014, the first commercial drift tube IM-MS system (6560 Ion Mobility Q-TOF) was launched by Agilent Technologies[10]. This system has been upgraded further with more precise gas control components[11]. The Agilent 6560 is composed of a 78 cm ring electrode-stacked drift tube, bracketed by the ion funnels. Under this configuration, ions travel through the drift tube under a uniform weak electric field (10~20 $V \cdot cm^{-1}$). Unlike another kind of drift tube IMS-TOF launched by TOFWERK[12], in which the pressure of the drift tube is around 760 Torr, the Agilent 6560 has a relatively low pressure of around 4 Torr. Despite the low pressure of Agilent 6560 that results in fewer collisions between the analyzed ions and neutral buffer gas, which limits the ion mobility resolving power to around 60, the low-pressure system enables higher sensitivity of the detection and is more suitable for complex samples analysis.

The initial design and typical operation of DTIMS utilizes a single pulse, which has an ion accumulation time of around 40-60 ms, forcing all other ions to wait until the previous ion packet has reached the detector. However, this long accumulation time brings with it the potential for space-charge effects that cause loss of low *m/z* (*m/z* < 250) ions[13]. Attempting to reduce the accumulation time in single pulse mode would further reduce duty cycle. To overcome these shortcomings, Agilent unveiled an ion multiplexing methodology. In multiplexed mode, ions are injected in multiple packets at predetermined intervals. Although the first ion packet is still

traveling through the drift tube, the following ion packet can also be injected. As each packet receives a shorter accumulation time, this strategy serves to enhance duty cycle and reduce the negative impacts of the space charge effects[14-17]. However, multiplexed mode results in overlapping ion mobility spectra. To deconvolute the complicated ion mobility spectrum the Hadamard Transformation algorithm was introduced to obtain deconvoluted drift times[18, 19]. Beyond this, high resolution demultiplexing (HRdm)[20] with Hadamard Transformation can also improve the signal-to-noise ratio[21, 22] and lower the limit of detection approximately 10-fold. The narrowed ion mobility peaks resulting from the Hadamard Transformation provides an increase in ion mobility resolving power from around 60 to between 100 and 200[23].

Structural Investigations

Collisional Cross Section (CCS) has emerged as the most ubiquitous and widely used metric in ion mobility-based structural analysis of gas-phase ions. Given the low uniform electric field applied to the drift tube, DTIMS serves as the gold standard for CCS measurement for structural investigations. CCS measurement has been obtained for a broad range of ion species including small organic compounds, carbohydrates, lipids, peptides, denatured proteins, and native-like proteins[10, 24-30], which have been fundamental in enabling the CCS calibration of TWIMS and other high-resolution ion mobility modalities[31]. These previous CCS measurements were foundational to the instrumental success of the Agilent 6560. Evolving from the criterion CCS measurement, DTIMS experiments on the Agilent 6560 platform can also be performed with CCS calibration with the so-called “single field” CCS method, using calibration ions to simplify the CCS measurement. Using this single field mode for CCS calibration, an interlaboratory evaluation showed high reproducibility (an average, absolute bias of 0.54%) of CCS measurement across different ion species can be achieved[32]. Furthermore, based on the comprehensive CCS

database constructed via DTIMS, many machine learning methods have also been developed to predict the CCS values in a short time without the need of complicated modeling calculation[33-37].

Beyond the measurement of CCS values, one of the important applications of DTIMS is the separation of isomers. Typical isomeric species including metabolites[38], lipids[39, 40], carbohydrates[41], and peptides[42, 43] are[42, 43] discussed extensively by other reviews, but this section will focus the discussion on recent applications of the Agilent 6560 platform. For example, organic pollutants per- and polyfluoroalkyl substances (PFAS) contain many isobars and constitutional isomers, which will generate similar fragment ions. Therefore, the traditional LC-MS/MS strategy is not suitable for the analysis of these compounds. By introducing ion mobility separation as an additional dimension, a lower detection limit and higher confidence structural identification can be achieved[44, 45] (Figure 9.2). Other organic molecular isomer separation applications include steroid metabolites[46-49], bile acids[50], peptide conformers/isomers[43, 51], and isobaric/isomeric biomarkers in newborn screening[23]. Ozonolysis, Paternò−Büchi reactions and cuprous ion-induced fragmentation have also been coupled with IM separation to identify double bond position in lipids[52-54]. Using the Agilent 6560 platform, the separation capacity of glycan isomers can be further enhanced by the incorporation of metal ions[55, 56] or derivatization[57]. Meanwhile, in addition to the electrospray ionization source, the Agilent 6560 can also be coupled to other ionization modalities to produce spatial information for mass spectrometry imaging. The Julia Laskin group has successfully used a desorption electrospray ionization source to achieve high-resolution imaging of biomolecular isomers in tissue[40, 58, 59]. Infrared matrix-assisted laser desorption electrospray ionization (IR-MALDESI), developed by David Muddiman and colleagues, can also be coupled to the Agilent 6560 to separate different analyte classes[60].

In addition to the separation of small molecular isomers, another structural investigation application of DTIMS is to study the gas-phase structures of large intact proteins or protein assemblies[61, 62]. In native mass spectrometry, the non-covalent interactions between proteins and ligands or proteins and proteins can be preserved. Further, proteins with different charge states can fold/unfold into different conformations due to the existence of secondary structure elements and Coulombic repulsive forces. Using the Agilent 6560 platform, solvent evaporation conditions and front end voltages can be tuned to best preserve the native state[63]. Beyond this, collision-induced unfolding (CIU) is gradually becoming a useful technique in the field of native ion mobility mass spectrometry to study the conformation and stability of intact proteins or protein complexes[64]. Ruotolo and colleagues employed sulfur hexafluoride gas in the source region in front of the Agilent 6560 drift tube to enhance collision activation efficacy[65]. This modification significantly improves CIU performance, allowing for the comparison of structure and stability between monoclonal antibodies and their biosimilar therapeutics[65-68].

Relevance to -omics Applications

Ion separation in DTIMS resides on the timescale of milliseconds, making it suitable for coupling LC separations that operate on the order of seconds. One advantage for DTIMS related -omics studies is that, in addition to the retention time and accurate mass, $^{DT}CCS_{N2}$ annotation can also be achieved in the same experiment. Since the launch of Agilent 6560, numerous metabolomics investigations have been performed with recorded $^{DT}CCS_{N2}$ annotation[69-74]. The Zheng-Jiang Zhu group collected more than 5,000 empirical metabolite CCS values from literature to predict the CCS for more than 1.6 million small molecules[75]. Beyond LC separation, electrophoretic separation, which exploits a compound's size and charge, operates under a similar mechanism to ion mobility separation. The relationship between the effective mobility and CCS

values was evaluated by coupling capillary zone electrophoresis (CZE) to DTIMS[76]. Lipidomic research has also been widely explored on the Agilent 6560 platform [77-83].

Compared to the nearly 100% duty cycle in TIMS, the poor duty cycle in DTIMS hinders further application in -omics applications. This limitation is largely mitigated when operating in multiplexed mode[19]. The Smith group first used ion multiplexing to identify and quantify liver fibrosis proteins from blood serum[84]. Another successful proteomic application is the identification of host protein signatures to evaluate the treatment effect of pulmonary tuberculosis[85]. More recently, multiplexed DTIMS has also been used in metabolomics. Compounds of emerging concern (CEC) in human urine samples were investigated and a comprehensive CCS database was built using the Agilent 6560 platform[86]. Untargeted metabolomics demultiplexing analysis can also be achieved[16]. To simplify the ion mobility-mass spectrometry-based -omics workflow and the detection of low abundance ions, the Smith group developed PNNL Preprocessor, which can integrate data interpolation, demultiplexing, multidimensional smoothing, and saturation repair functions. The PNNL Preprocessor software is proven to have faster processing speed and yields greater lipid annotation in lipidomics[87] analyses.

Considerations and Future Directions

Considering CCS values have high inter-laboratory reproducibility compared to inconsistent retention times in LC separations, DTIMS, the only direct CCS measurement paradigm, continues to play an important role in the ion mobility field. It is conceivable that in the future each ion will have an accurate CCS value determined by DTIMS as an intrinsic property like the mass-to-charge ratio. However, the current iteration of DTIMS presents some limitations. Poor duty cycle resulting from the tradeoff between ion accumulation time and total time in the drift tube hinders ion utilization efficiency. Advanced ion injection strategies, such as

multiplexing, require further development to overcome this obstacle. As mentioned above, the low-pressure (4 Torr) environment in the drift tube of the Agilent 6560 platform further limits the ion mobility resolving power due to fewer collisions between ions and neutral buffer gas. Further increasing pressure in the drift tube will be an effective way to enhance the ion mobility resolving power. Overall, we foresee that DTIMS will play an increasingly important role in the future of scientific research.

Cyclic Ion Mobility Spectrometry (CIMS)

Cyclic Ion Mobility Spectrometry (CIMS) is a unique entry in the high-resolution IMS family. It is based on Traveling Wave IMS (TWIMS), which was developed in the early 2000s and commercialized soon after by Waters Corporation[88, 89]. In accordance with other IMS modalities, it is readily interfaced between liquid chromatography separations and mass spectrometry. While TWIMS is a commonly employed and successfully commercialized IMS modality, its earlier iterations failed to provide resolving powers greater than 50[90]. Such low resolving powers struggle to differentiate CCS values that differ by less than 1%. TWIMS resolution increases roughly with the square root of path length, so overcoming these diminishing returns necessitates ultra-long path length devices[91]. CIMS has been developed to achieve high resolution separations by extending the traditionally short TWIMS path length and enabling analytes to undergo multiple passes[92] (Figure 9.3), substantially improving resolution. CIMS technology was commercialized in a tandem IMS time-of-flight mass spectrometry system by Waters Corporation in 2019 as the Select Series Cyclic IMS and is one of the only ultra-long path length IMS platforms to be commercialized[93-95]. Since the resolving power of the IMS separation is proportional to the number of passes around the path, resolving power and analysis time are easily tunable by the operator, allowing for convenient optimization[92]. The system does, however,

require some tradeoffs at higher resolving powers. Principally, as higher resolving powers are achieved, a narrower CCS window must be measured, otherwise higher mobility compounds would wrap around those of lower mobility. Due to this complication, the Waters Cyclic IMS provides a very powerful platform when targeted super-high-resolution separations are desired, but it is not generally suitable for discovery-based applications. A brief comparison of Waters' TWIMS-based IM-MS offerings is highlighted in **Table 2**, should this IM paradigm be of further interest.

Operating Principles and Engineering

The CIMS cell is effectively a circular TWIMS cell that has been modified with an ion entry/exit port. A relatively mature IMS archetype, traditional TWIMS separates ions in a drift tube-like fashion. Instead of applying a constant potential, however, ions are subjected to uniform "waves" of potential that travel the length of the cell. Higher-mobility ions are able to "surf" these waves more effectively and are overtaken by them less often than low-mobility ions[91]. TWIMS analysis utilizes pulsed ion injection, much like DTIMS, bringing with it the same shortcomings described above. To overcome this limitation, the Select Series CIMS contains a quadrupole mass filter and an ion trap upstream from the CIMS cell, where ions are accumulated prior to injection and analysis[92]. It should be noted that CCS value measurements in CIMS must be obtained from careful calibration of the instrument and not physical/electronic characteristics alone. Although uncertainty and variation in "true" CCS value is still a limiting factor, relatively recent progress in optimization of TWIMS calibration has made high-confidence calibration much more reliably achievable, an especially important consideration for super-high resolution CIMS[96, 97].

One of the largest drawbacks to CIMS' application in -omics investigations is the necessity to eject ions from the cyclical flight path. Since the CIMS cell contains an ouroboric ion path,

ejection of certain ions will become necessary after a certain number of passes to prevent higher mobility ions from overtaking those with lower mobility. This results in the potential for extremely high-resolution separations at the cost of sample coverage and CCS range. If targeted high-resolution separation is all that is desired, the system can be utilized in "IM Isolation" mode, where all ions outside of a desired mobility range are immediately ejected during the first pass, and the remainder may continue being separated in subsequent passes. The Waters Cyclic is also the only platform on this list capable of true tandem IMS. It can be operated in IMS^n mode, where ions are separated as normal in the CIMS cell, then instead of being sent to the TOF for MS analysis, selected packets will be reintroduced to the pre-array store and may then be reinjected into the CIMS cell under different conditions a theoretically unlimited number of times[92].

In terms of resolving power, the Select Series CIMS has a single pass resolving power of 60-80, and this value should theoretically increase by the square root of the number of passes. Experimentally, this relationship remains true to the theory, with resolving powers of ~750 observed after 100 passes[92]. Giles *et al*. report ion losses of less than 2.5% per pass through the CIMS loop for small relatively stable ions though it is expected that this value will vary greatly between analyte families and robustness[92].

Structural Investigations

Since the introduction of the first commercialized ion mobility mass spectrometer, the Waters Synapt HDMS, TWIMS has been widely used in structural investigation of different ion species such as small organic compounds, glycans, peptides, and proteins. The Waters Cyclic, with high resolving power and unique geometry, further enhances structural investigations.

Complex small molecule mixtures, such as petroleum, contain many isobaric and isomeric compounds, which provides an ideal application scenario for ion mobility differentiation. For

example, the Waters Cyclic can help to identify the CH- and CHS- petroleum species solely on structural difference, whereas traditional mass spectrometry may fail to resolve these species due to the small 3.4 mDa SH4/C3 mass difference[98]. The IM isolation function, which focuses on specified mobility regions, can help to reduce the interferences of isobaric compounds. Notably, gradually increasing the number of passes from 1 to 10 has demonstrated gradual separation of isomeric benzo[*b*]naphtho[2,3-*d*]thiophene and anthra[2,3-*b*]thiophene[98]. Other similar complex mixture separations on Waters Cyclic include lipid isomers[99, 100], crude oil compounds[101, 102], environmental contaminants[103-106], natrual compounds[107], isomeric drugs and related metabolites[108-110]. Beyond this, Gabe Nagy and colleagues also utilized the Waters Cyclic to study the effect of isotopic substitutions in isotopologues and isotopomers on the mobility change[111]. Surprisingly, two deuterated palmitic acid isotopomers with deuterium labeled at different positions show different mobility. This finding challenges the classical Mason−Schamp equation, in which isotopomers should not be resolved given their identical mass and structure. Given this finding, access to this ultra-high-resolution IM technology may prompt reevaluation of tradition IMS theory. Besides, temporal compression was found to be capable to improve the IMS peak intensity in CIMS[112].

Glycan sequencing is challenging in analytical science due to the complexity of the monosaccharide building blocks, which contain several chiral centers. The anomericity and regiochemistry of the linkages between the monosaccharides further complicate structural characterization. The Waters Cyclic offers several advantages for the elucidation of glycan structure. First, high-resolution ion mobility with multiple passes facilitates separation of oligosaccharides[113-115]. For example, mixtures of three pentasaccharides cannot be resolved in 1-pass (R~65). However, after 5-passes the resolving power increases to ~145, allowing facile

separation of pentasaccharides constituents[113]. This application, however, is not immune to the "wrap-up" effect. In this example separation, 7 CIMS passes will result in the IM peak of the low-mobility branched mannopentaose being overtaken by the highest-mobility cellopentaose, which will distort the IM measurement for different ion species[116]. To overcome this issue, IM isolation mode can be used to select a specific ion mobility range, allowing for target molecules to complete a higher number of passes. The other advantage of the Waters Cyclic in glycan analysis is the unique IMS^n function. Not only are the glycan precursor ions isomeric, but also the product ions. The engineering design of the Waters Cyclic allows the selection and dissociation of the precursor ions with a specific ion mobility range and can further separate product ions, an approach similar to MS^n. The isomeric disaccharide and trisaccharide building blocks of the glycans display specific ion mobility fingerprints and diagnostic fragmentation patterns, which enable the sequencing of oligosaccharides[117-119]. As well, the Waters Cyclic also demonstrates potential in assigning exact fucosyl[120] and sulfate[121] positions and elucidating the structure of glycopeptides[122].

Peptide isomers resulting from the stereoisomerism and chemical modification of different residues are difficult to analyze due to their identical mass and possible co-elution in reversed-phase liquid chromatography. Compared to other advanced ion mobility paradigms, the Waters Cyclic has a trap cell, in which precursor ion can be fragmentated by collision-induced dissociation (CID) prior to entry into the CIM separator. Fragment ions of isomerized peptides will display a recognizable arrival time shift, which can be used for the site-specific localization of isomerization[42, 123-125]. As an example, 4 passes with resolving power around 130 in the Waters Cyclic was found to be sufficient to map the racemization or L/D-amino acid substitution site in protein therapeutics[126]. Other successful applications of the Waters Cyclic include the separation of cross-linking peptides[127] and the assignment of disulfide bridge pairing[128].

Finer gas-phase structure of proteins can be provided by multi-pass separation in the Waters Cyclic. Meanwhile, the increased length of IM separation will also increase the time of protein ions spent in the gas phase. Thalassinos and colleagues used the Waters Cyclic to study the gas phase stability of protein ions and found the native protein conformation is stable on the order of hundreds of milliseconds[129]. Further, the IMS^n function of the Waters Cyclic can separately slice specific ion mobility range of proteins ion to perform the collision-induced unfolding (CIU) experiments, which will provide more detailed information about the protein unfolding pathway[129-131]. In addition, the Waters Cyclic can be further retrofitted with an electron capture dissociation (ECD) cell either in front or rear of the cIM separator to enhance top-down protein characterization[132]. Similarly, Vicki Wysocki and colleagues have incorporated a simple surface-induced dissociation (SID) cell into the Waters Cyclic instrument, which enables surface-induced unfolding (SIU) experiments[133, 134]. It should be noted that this is also the first commercialization of SID, which will provide new insight into the analysis of proteins and protein complexes.

Considerations and Future Directions

With increasing CIMS pass numbers, time spent within the CIMS flight path will increase. Across all IM paradigms longer flight times are associated with reduced transmission, leading to a reduction in sensitivity. Additionally, the existence of the "wrap-up" effect resulting from multiple passes may hinder the accurate measurement of *m/z* and arrival time for the whole ion system. The above features limit application of the Waters Cyclic in omics-related investigations. Furthermore, the CCS measurement of target ions is also an issue for Waters Cyclic. The electrical field in the CIMS separator is not uniform, which means the CCS values cannot be calculated from first principles. The current work-around is to use calibration ions; however, the CCS values of

calibrant ions obtained from DTIMS have an experimental uncertainty between 0.5% and 2%[10]. Although single-pass CCS measurements on the Waters Cyclic agree well with the literature, multi-pass CCS measurement cannot be reliably achieved until the validation of higher accuracy CCS standards. Nevertheless, the customizable ion mobility resolving power and unique IMS^n function of the Waters Cyclic promise to bring more exciting IMS applications in the future.

Trapped Ion Mobility Spectrometry (TIMS)

Trapped ion mobility spectrometry (TIMS) is a relatively recent addition to the bioanalytical toolbox, and was patented in the late 2000s by Melvin Park and associates at Bruker Daltonics[135]. Despite the short turnaround time since its commercialization in 2017 as the timsTOF line of instruments, TIMS has quickly matured into a powerful and convenient platform for a wide variety of structural analyses and discovery-based investigations. Specifically, TIMS provides a sensitive and flexible platform that is well suited to add another dimension of separation in between existing chromatographic and mass spectrometry-based workflows[136-138]. In addition to the ESI-based instrumentation that brought TIMS into the spotlight, Bruker has recently unveiled TIMS units with matrix-assisted laser desorption/ionization (MALDI) imaging capability that is consistent with the current gold-standard[139]. An optional MALDI-2 postionization laser and TIMS' unique parallel-accumulation serial fragmentation (PASEF) capabilities round out the modality as an impressive, highly sensitive platform for modern structural biology and -omics investigations[140, 141].

Operating Principles and Engineering

The TIMS separation principle is effectively inverted from traditional ion mobility modalities. In drift tube-style devices, rotationally averaged collisional cross sections (CCS) of gas-phase ions are measured by accelerating them through an environment populated with inert

gas, which impedes ion motion toward the detector in a way that is proportional to the ions' CCS. In TIMS, ions are immobilized in a region filled with moving gas. This trapping is accomplished inside a segmented linear quadrupole ion trap. In this TIMS cell, different plates along the length vary the potential on ions as a function of distance, so that ions with larger CCS values will be pushed further along the cell due to the energy they receive from the carrier gas (Figure 9.4)[142]. The accumulated ions can then be eluted through the TIMS analyzer by sequentially lowering the position-dependent plate potential as a function of time. This sequence of accumulation and elution can be adjusted to optimize for fast scans (tens of ms) or for high resolution separations (hundreds of ms)[143]. Using optimized stepping scan functions can provide IMS resolving powers >300 while reducing overall experiment time, and increasing duty cycle[144]. As detailed below, there are a host of parameters specific to the TIMS cell and accompanying ion optics that may be altered and optimized to meet experiment-specific needs. By elongating the TIMS cell and creating two separate trapping regions within, the first "ion accumulation trap" can collect ions while the second trap analyzes a previously collected batch. This technique, dubbed Parallel Accumulation-Serial Fragmentation (PASEF), provides a duty cycle of up to 100%, albeit usually with a reduction in maximum resolving power[145]. PASEF can be harnessed to successfully increase MS/MS coverage and maintains the sensitivity of the TIMS cell[137]. The extra dimension of separation that TIMS provides, when combined with the potential to create highly reproducible data sets, and high ionization efficiency makes it a very attractive platform for data-independent acquisition (DIA) experiments. Dubbed "diaPASEF," this acquisition mode has been shown to overcome traditional drawbacks to the technique such as low ion utilization and convoluted spectra[124].

Much like other high-resolution IMS techniques, and unlike drift-tube IMS, the CCS of ions cannot be readily calculated from first principles. Instead, instruments must be calibrated with

standards of accurately known CCS values. A properly calibrated TIMS produces CCS values that are reproducible, accurately matching drift-tube values to around 1% and demonstrating high reproducibility[146, 147]. Calibration has even been validated by inserting a drift tube before the TIMS cell, thereby comparing CCS values from both IMS modalities in tandem and with the exact same analytes[148].

Role in Structural Investigations

Due to its high resolving power, easy interfacing between liquid chromatography and mass spectrometry, and high duty cycle, TIMS separations have quickly been adopted for a wide variety of structural characterization. Fast, simple, and effective analysis of chiral compounds can be achieved in the gas phase thanks to the TIMS cell[149]. Impressively, the high resolution capabilities also allow for separation of isobaric lysine propionylation and acroleination, which have CCS values that differ by as low as 1%[150]. TIMS has been utilized effectively in data-dependent acquisition PASEF (DDA-PASEF) mode to determine the mechanisms of SARS-CoV-2 host protein interactions and identify binding motifs[151]. In terms of tertiary and quaternary structure of proteins, TIMS has demonstrated the ability to conveniently separate complex antibody-drug conjugate mixtures prior to MS analysis, allowing for high-throughput structural characterization of multiple attributes in top-down analysis[152].

Cross-linking mass spectrometry is a proteomic technique that involves intensive data analysis to differentiate cross-linked peptides from linear digested ones. TIMS provides a convenient way to easily discriminate linked and non-reacted peptides by CCS, and provides more robust information than can be gleaned from MS analysis alone[153]. This distinction can even be automated to determine in real time which CCS values represent species of interest[154]. The application of TIMS with another structural proteomic technique, fast photochemical oxidation of

proteins (FPOP), improves quality of analysis and even enables resolution of different modifications on the same amino acid residue, as well as the ability to differentiate peptides based on location of backbone oxidation[155, 156]. The structural applications of TIMS are apparent in top-down investigations as well, where it has been used to differentiate protein conformations based on differing amounts of intra-protein disulfide bonds[157].

TIMS has also been used successfully to analyze isomeric opioid metabolites in human urine, and does so with better precision and reproducibility than standard multiple reaction monitoring (MRM) techniques[158]. Analysis of isomeric compounds is an especially important task in analysis of lipids since much of the diversity in the lipidome stems from isomeric species. The addition of TIMS in lipidomic workflows allows for more robust characterization while maintaining high sensitivity and vastly increases MS/MS coverage in the resulting data[159, 160]. TIMS' ability to interface with MALDI imaging produces a system that can provide a useful dimension to deconvolute lipidomic imaging data[161].

Relevance to -omics Applications

The Bruker timsTOF lineup is tailored to -omics applications first and foremost. The convenient and rapid high-quality IMS separations pair extremely well with high-resolution imaging capabilities and rapid MS acquisition. Proteomics is currently the most mature of the -omic disciplines, and TIMS analysis aids in pushing the envelope on both targeted and discovery-focused investigations. TIMS has been shown to decrease spectral complexity in proteomics by separating peptides prior to MS analysis[162]. The reduction in co-fragmentation is an added benefit to the extra dimension of separation provided by IMS in general. TIMS has been utilized to successfully improve quantitation using isobaric tags without increasing experiment time[163], and is suitable for label-free phosphoproteomics[164]. Due to the relationship between timescale of

TIMS scan and MS scan, it is possible to measure multiple different peptides in each ion mobility scan. Using this in a targeted proteomics approach has been achieved in parallel reaction monitoring (PRM-PASEF) workflows[165, 166]. This methodology allows for absolute quantitation of endogenous peptides when isotopically labeled standards are spiked into bottom-up samples. With the high sensitivity enabled via PASEF, this technique has been successfully employed to monitor pathogenesis, progression, and biomarkers of various diseases, and in the localization of glycation sites in human serum albumin[167-169]. Software packages that leverage PASEF parameters can be employed to improve run-to-run reproducibility of PRM experiments[170].

Yet more nanoLC-coupled workflows have been optimized for peptide biomarker detection and targeting[171]. The human cardiac proteome has been analyzed and demonstrates high reproducibility and number of protein identifications[172]. The timsTOF is also well suited to the generation of spectral libraries due to its high sensitivity[173, 174]. It has also been shown that the TIMS cell is capable of interfacing with ECD and can readily differentiate the histone proteoform in this modality[175]. On the subject of peptide and protein fragmentation, TIMS parameters can be utilized to fragment these larger biomolecules in the TIMS cell itself, providing a "pseudo-MS^3" analysis for top-down or middle-down proteomics[176].

Single-cell proteomic analyses are substantially improved through utilization of timsTOF technology, as the high sensitivity and duty cycle capabilities have shown promise in dealing with the inherent extreme sample-limited conditions[177-180]. These capabilities have even been leveraged to measure peptide stereochemistry within a single cell[181]. Even so, new developments offer a glimpse of an even higher sensitivity for the instrument in the future, with work being done to produce a brighter ion beam and lower the limit of detection even further[182]. Recent work has been

done to enable sub-cellular MALDI MS analysis of single organelles thanks to the timsTOF's MALDI capabilities and aptitude for sample-limited conditions[183].

Metabolic and lipidomic workflows on the timsTOF are able to reap the rewards of TIMS' capabilities as well. Taking advantage of the system's high resolution mass spectrometry and ion mobility, human urine metabolites can be targeted and analyzed with relatively little sample preparation[184, 185]. The high resolution capabilities of TIMS analysis also allow highly effective separation and discrimination of biologically relevant lipid species based on CCS differences far below 1%[186]. As always, PASEF remains relevant in metabolomic workflows, increasing the number of features with high-quality MS/MS spectra associated[187]. These qualities have enabled targeted and untargeted metabolic profiling of such wide-ranging systems as extra virgin olive oil, mosquitoes, and breast cancer cell lines[187-189]. Additionally, the MALDI capabilities of the timsTOF Pro are effectively supplemented by the high-quality IMS separation. Traditionally, MALDI imaging of small molecules is impeded by high amounts of low-mass interferences that originate from the matrix. By performing a high-quality IMS separation prior to mass spectrometry analysis, these interferences can be cleaned up, and higher quality data is obtained. The high performance of the timsTOF MALDI imaging system has been leveraged in spatial metabolomic inquiries on human kidneys[190]. TIMS has been used to study human colorectal cancer from a metabolic and multi-omic point of view and has been used to identify genomic perturbations associated with mitochondrial dysfunction and poor disease prognosis[191, 192]. Recent work has used MALDI-TIMS imaging for structural elucidation of modified Lipid A in bacterial colonies[193].

Due to the complexity inherent in acquiring and analyzing 4-dimensional chromatography-TIMS MS/MS data, many different software packages have been published for a wide variety of TIMS-specific applications. MaxQuant and the associated MaxDIA are broadly applicable

software packages for shotgun proteomics and DIA analysis respectively[194, 195]. OpenSWATH is an open-source DIA tool that has been adapted for use with diaPASEF experiments[124, 196]. MaxLynx is another package built upon the MaxQuant environment and is designed specifically for cross-linking mass spectrometry analysis[197]. MSFragger and IonQuant are two PASEF-compatible software packages from the Nesvizhskii group that purport to improve unique peptide IDs and fast label-free quantitation when compared to MaxQuant and the popular software PEAKS [198]. Analyzing DIA data can be accomplished with and without a spectral library, and these two approaches each come with perks and drawbacks. According to Wen *et al.*, with the current state of the workflow, libraries will provide a higher number of both precursor identifications and missing values. Despite this, both library-based and library-free analyses lead to comparable conclusions[199]. Machine and deep learning strategies have been harnessed to improve peptidomic identifications, and to build computational CCS libraries with high accuracy[200, 201]. Neural networks (NN) have recently proven to be effective in analyzing DIA proteomics data, and this DIA-NN technology has been added to the Bruker PaSER 2022 software release[202]. Finally, for simple visualization and indexing of large datasets, AlphaTims and the OpenTIMS suite are designed for fast access to raw data[202-205].

Future Directions and Considerations

Although this may seem like a plethora of software, the relative novelty of the platform dictates that many needs are still unfulfilled. As a newer entrant to the field, TIMS technology lacks pragmatic guidance on method building, and there is a need to rigorously evaluate and develop methodologies that can be disseminated throughout the community. For more mature proteomic platforms, this seminal development work has already been conducted, and users are able to select published and validated methods for common analyses with high confidence.

Community efforts to develop standard proteomic, metabolomic, and lipidomic methods on the timsTOF would enable greater access to such a promising platform. There are also certain formative questions that remain unanswered about the system. For example, there is evidence that although the system is indeed suitable for native proteomics, CCS distributions can vary widely across trapping methods[206]. In a similar vein, recent work indicates that small molecule CCS values are influenced by solvent conditions and trapping parameters, and these effects should be noted by users[207]. These drawbacks are certainly unfortunate, but many of these gaps in knowledge are due in large part to the relative novelty of the platform as a whole. With time and increased utilization, it is anticipated that many of these needs will be addressed as the scientific community continues to embrace this convenient and powerful family of instruments. One future development that may prove to be impactful is tandem TIMS[208]. The yet uncommercialized technology has been utilized to probe protein structure changes as a function of the proteoform and shows the ability to maintain highly charged non-covalent protein assemblies[209, 210]. It remains to be seen whether this development is useful enough for widespread adoption, or whether drawbacks will render it a niche addition to the TIMS family. Despite these unknowns, the extensive publication record of this instrumental regime in its nascent lifetime indicates TIMS and the Bruker timsTOF have a bright future in the analytical fields and promise to push the boundaries of modern ion-mobility spectrometry.

Differential Ion Mobility & Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS)

Standing in contrast to each of the previously discussed ion mobility paradigms, differential ion mobility spectrometry (DIMS) provides a unique entry into the realm of high-end ion mobility instrumentation in the form of Field Asymmetric-Waveform Ion Mobility Spectrometry (FAIMS).

Whereas TWIMS, TIMS, and DTIMS exploit the instantaneous mobility of gas phase ions in the presence of a constant electric field, DIMS is so named for its ability to exploit the different mobilities of gas phase ions when in the presence of low and high electric fields. While the storied development and analytical success of DIMS has been detailed elsewhere[211], we will confine our discussion to that of FAIMS, as this sector of differential ion mobility has seen successful commercialization and significant implementation in structural elucidation and cutting-edge mass spectrometry based biomolecular investigations. As demonstrated across the past decades of development, the flexible, compact design of various FAIMS implementations have been specially tailored to provide higher separation capacity and analytical sensitivity while providing topical considerations in experimental design.

Background and Engineering Developments

While the conceptualization and invention of FAIMS is unclear, credit may most broadly be given to Russian scientists circa 1980, with the earliest literature appearing in the 1990s, prior to the technology reaching the United States[212]. At the time, traditional IM modalities capable of exploiting gas phase mobility of ions required sub-ambient pressure, forcing confinement of these separation strategies to settings where high vacuum and stable temperature could be achieved. In contrast, FAIMS, which operates at atmospheric pressure and near-ambient conditions, offered the first suitable mechanism for field deployment and detection of a broad range of chemical species[212]. In FAIMS, ions are introduced between two planar or cylindrical electrodes and are the propelled toward the detector by a carrier gas. An asymmetric waveform voltage – referred to as the Dispersion Voltage (DV) – is then applied to a single electrode such that the oscillating high and low electric fields yield equivalent time-voltage integral (Figure 9.5)[212, 213]. The oscillating high and low fields, each of opposite polarity, impart a force on gas phase ions orthogonal to the

carrier gas, causing ions to migrate towards the opposing detector in a fashion consistent with their response to high and low fields. The dispersion voltage alone would ultimately cause all ions to contact the electrodes and be neutralized. To account for this, a second voltage – the Compensation Voltage (CV), constant in both polarity and amplitude – may be applied to the opposite electrode, allowing for responsive ions to drift safely towards the detector. When the DV and CV are held constant, FAIMS devices operate as an ion filter, as only those ions compatible with chosen voltages will reach the detector. This strategy is fundamental to biomolecular separation, as discussed below, but would present infinitesimally low duty cycle in complex mixture analysis. As such, all but the most specialized FAIMS devices have the capacity to sweep through a range of compensation voltages or may otherwise operate at a range of voltages for the duration of the experiment. These FAIMS principles, which are comprehensively explained elsewhere[213], give rise to multiple unique FAIMS implementations, with planar and cylindrical instruments being the focus of our discussion.

Though fundamental differences in planar and cylindrical FAIMS devices may be the subject of a separate, comprehensive review, topical considerations arise from how each geometry influences the achievable electric field. Planar FAIMS, using flat plate electrodes, reminiscent of traditional DIMS, allows for homogenous electric fields across the FAIMS device, which provides narrow peak widths and therefore higher peak capacity[214, 215]. Shvartsburg and colleagues have repeatedly demonstrated the utility of the FAIMS regime for biomolecular separation and distinction of isoforms, as shown below. Early on, Shvartsburg *et al.*[216] validated that smaller planar gaps allow for higher electric fields, allowing for shorter analyses without sacrificing resolution and later demonstrated that reducing gas flow allows for even higher peak capacity[217]. Later hybrid FAIMS-IMS implementations were able to increase gas-phase resolution even

further[218]. While FAIMS' operation at atmospheric pressure offers utility in field applications[219], this can present limitations when coupling FAIMS to modern mass spectrometers with strict gas requirements and pressure limits. Baird *et al.*[220] demonstrated a means of removing buffer gas constraints when coupling planar FAIMS to orbitrap mass spectrometers, while Shvartsburg *et al.*[221] demonstrated FAIMS may be used in low pressure regimes, paving the way for future instrumental implementations and coupling to activation electronics.

Contrasting planar FAIMS, electric fields applied in cylindrical FAIMS are inherently non-homogenous. This lack of uniformity ultimately reduces the achievable peak capacity compared to planar FAIMS, lowering the achievable gas phase resolution. However, non-homogenous electric fields provide a means of ion focusing, resulting in significantly higher sensitivity, and making this FAIMS regime more suited to discovery-based -omics investigations. In addition, the geometry of cylindrical FAIMS (Figure 9.5) is more amenable to controlling temperature stability, which greatly improves ion transmission[214, 222]. Similar to planar FAIMS, recent reports have improved analytical performance by decreasing electrode gap widths, which provided a four-fold increase in peak capacity and 98% increase in identifiable proteins[223]. Noting the complementary benefits in these two popular FAIMS implementations, it becomes clear to what research interests each lends its analytical capability.

Structural Investigations

Due to the propensity for operating at higher electric fields, higher peak capacity, and demonstrated gas phase resolution, planar FAIMS has been extensively employed for the separation of diverse sets of isomeric and isobaric biomolecules. Recently employed to identify isotopic shifts on organic molecules[224], separate polyproline isomers[225] and classify lipid isomers[226, 227], this separation modality exhibits innate propensity to glean minute structural

differences from relevant biomolecules. A more active sector of research, however, is the utilization of FAIMS to separate and identify peptide/protein isoforms. Given the ubiquity of protein phosphorylation and challenges in site localization, these biomolecules made for an excellent analytical subject in early analyses[228-231]. Concurrently, Shvartsburg *et al.*[232] demonstrated separation of peptide sequence isomers employing a similar strategy in the analysis of methylated histones[233]. This latter work was further expanded by Garabedian *et al.*[234], Shliaha *et al.*[235], and Baird *et al.*[236], the latter of which detailed high mobility and mass resolution. Given its suitability for localizing and distinguishing post-translational modifications (PTMs), planar FAIMS has also been used to separate and identify isomeric glycopeptides[237]. While the structural complexity of glycosylation still outpaces the analytical power of FAIMS, Pathak *et al.*[238] successfully separated isomeric glycopeptides. Additionally, FAIMS has also been employed to analyze both small and large proteins[239]. These targeted studies demonstrate the capability of FAIMS as an ion separation and ion filtering technique and highlight the distinct analytical power it may present. Even more noteworthy is the extensive utilization of FAIMS within -omics investigations, which have aided in cementing FAIMS as capable far beyond structural investigations[240].

Relevance to -omics Applications

Though several commercial FAIMS offerings exist, the rise in access to and adoption of orbitrap mass spectrometers within -omics investigations has served to establish the Thermo Fisher FAIMS Pro as the most ubiquitous analytical platform. Given the ability of FAIMS to operate at a range of CVs – providing unique ion filtering at each – the FAIMS pro has been extensively employed in proteomics, as it provides a facile avenue towards comprehensive sample coverage. This is exemplified in the numerous reports of FAIMS within bottom-up analyses, namely Hebert

et al.[241] who provided early reports of comprehensive bottom-up analyses with FAIMS, demonstrating how ion separation can be used to augment existing analytical techniques[242-247]. Of significant interest at the time of review is the analytical sensitivity enabled through FAIMS[248, 249]. FAIMS has aided in analysis of iPSC-derived neurons[250], has shown utility in single-shot proteomics when using short LC gradients[251, 252], and has even been coupled with machine learning[252] to identify >2,800 proteins from a single nanogram of material[253]. Demonstrating extreme sensitivity – having demonstrated the ability to identify >1,000 proteins from individual cells[254], FAIMS presents itself as a meaningful modality in the nascent field of single-cell proteomics[178, 255] that can be further enhanced through instrumental application programming interfaces (APIs)[256].

Though discovery-based analyses are a permanent fixture within proteomics, FAIMS has also been extensively employed to benefit quantitative proteomics investigations. Beyond improvements in sample coverage, FAIMS – operating both as an ion focusing agent and ion filter – improves quantitative accuracy by reducing the quantity of co-isolated precursors, with numerous reports detailing the improvements achieved[257-259]. Several reports have also detailed the ability of FAIMS to provide equivalent or higher quantitative accuracy[260] without reducing identification rates[259, 261], as well as showing utility in creating custom analysis pipelines[262]. The analytical capacity of FAIMS is further expanded when targeting a specific biomolecular class. As in all ion mobility regimes, biomolecule subspecies (i.e., phosphopeptides, glycopeptides, etc.) bear similarities in mobility with one another, making them easily distinguishable from other analytes. As such, FAIMS parameters may be tuned to specifically target these species and has been employed in quantitative investigations[261, 263, 264]. PTM analyses via FAIMS extend beyond

quantitative proteomics[265] as FAIMS also provides benefits in global[247, 266-268] and targeted[269] PTM profiling from a variety of biological sources.

FAIMS has also been leveraged in more niche scenarios, including those where significant sample limitations exist. Of interest, Cooper and colleagues have utilized liquid extraction surface analysis (LESA) in combination with FAIMS to analyze both heat preserved tissue[270] and dried blood spots[271], resulting in a 50% increase in protein identifications from the latter report. Other unique applications of FAIMS include the use of real time searching (RTS) to determine temporal protein expression[272], monitoring host cell proteins produced during expression of biotherapeutics[273], as well as using FAIMS filtering to selectively analyze the cysteinome[246], SUMOylated peptides[274], crosslinked peptides[275] and identify PTM cross-talk sites[276]. However, among these unique applications, special attention may be drawn to direct-infusion shotgun proteome analysis (DI-SPA), recently demonstrated by Meyer *et al.*[277]. Given the value in quantitative proteomic measurements, innovations that provide higher throughput, lower analysis time, and higher accuracy are of paramount importance. Meyer demonstrated that utilizing FAIMS filtering in combination with data independent acquisition (DIA) results in astonishingly high throughput, having acquired >45,000 quantitative measurements from 132 samples in ~4.4 hours. DI-SPA is sure to provide a framework for future high-throughput proteomics workflows, and was even recently adapted to PRM analyses[278].

While bottom-up proteomics applications occupy the largest swath of FAIMS-based -omics investigations, there is growing interest in using FAIMS for larger protein fragments and intact proteins[279], as well as metabolomics analyses[280, 281]. Though analysis of histone isoforms was repeatedly mentioned in the discussion of planar FAIMS for structural characterization, reinforcing the biological relevance and complexity of these biomolecules, FAIMS has also been

employed for discovery-based middle-down histone investigations. Utilizing multiple ion separation regimes, Garabedian *et al.*[234] were able to identify methylated, trimethylated, acetylated and phosphorylated histone variants, while Shliaha *et al.*[282] were able to confidently identify histone isoforms from mouse embryonic stem cells. Moving beyond these middle-down analyses, top-down and native mass spectrometry have also demonstrated improvements when FAIMS is incorporated. Fulcher *et al.*[283] successfully employed FAIMS for top-down proteomic investigations of Alzheimer's Disease brain tissue, while Griffiths *et al.*[284] identified proteoforms from tissue samples using LESA-FAIMS. Gerbasi *et al.*[285] reported increased proteoform identification through inclusion of FAIMS, while Fulcher *et al.*[286] further illustrates this point in their utilization of FAIMS CV stepping. Circular FAIMS has also been used for native mass spectrometry analyses[287], illustrating that FAIMS may be tailored for fragile ions where maintaining tertiary structure is paramount. Moving in the opposite direction, FAIMS is also increasingly employed for metabolomic investigations, aided largely by the capacity to sample airborne analytes and design miniaturization[288]. Traditional untargeted metabolomics studies have detailed the utility of FAIMS in separation and distinction of metabolite isomers[289], screening potential biofluid biomarkers[290], identifying airborne chemical constituents[291] and comprehending differences in fecal microbes in response to disease[292]. These studies, however, are more akin to the numerous proteomics investigations listed above, as analytes of interest are obtained from solution or tissue. More interestingly, volatile organic solvents[293] (VOCs) are easily sampled from above urine[294-298] and stool[299] for the detection of irritable bowel syndrome (IBS)[294, 299], diagnosis of diabetes[297], and identification of cancer[295, 296, 298]. These recent applications, covering a broad spectrum of biomolecular species, serve to highlight the applicability and utility of modern FAIMS implementations and commercial offerings.

Considerations and Future Directions

While the analytical advantages of incorporating FAIMS to an existing workflow have been extensively described, this technique is not without drawbacks or topical considerations. Namely, when it comes to structural investigations of biomolecules, FAIMS offers no ability to measure collisional cross section (CCS) directly[4, 42]. This is in stark contrast to other ion mobility regimes that place this ability within reach. As such, the correct assignment of ion conformations in FAIMS analyses hinges on the availability and purity of biomolecule standards, which may limit the breadth of discovery sought in high-throughput experiments. In addition, FAIMS devices display a physical limitation of only being able to operate at a single CV at a given time. As such, FAIMS is ultimately a scanning technique, allowing only compatible ions to travel safely between the electrodes and is therefore significantly lower in throughput than other IM modalities. These limitations being well-known, future FAIMS innovations are likely to center on achieving faster ion separations without sacrificing sensitivity or resolution, and incorporation of rapid electric field switching that would allow broader collection of ions to be scanned within a unit time. Further, given the power and utility of instrument APIs, one can imagine a scenario in which FAIMS' voltages are controlled in a similar fashion, providing a means of targeted ion selection and intelligent precursor isolation. Nevertheless, modern FAIMS is a powerful analytical technique, providing significant improvements in sample coverage and high sensitivity across -omics investigations. This IM paradigm, relatively nascent within the field, is sure to experience significant growth and higher utilization in coming years.

Conclusion

Modern ion mobility mass spectrometry instrumentation grants unparalleled access in metabolomics, proteomics, and structural investigations. As biological mass spectrometry continues to grow in ubiquity and access to high-end instrumentation becomes more achievable, utilization and expansion of IM-based methodologies will grow in capability and efficacy. While it is likely the next decade of instrumentation will present new, improved capabilities that far outpace current capacity, today's high end ion mobility instrumentation may be remembered as an inflection point in the history of IMS technology. Having described relevant innovations, meaningful applications, and potential limitations and drawbacks of four high-end IM instrument paradigms, this review may serve as a reference point for novice and established researchers seeking to begin or further their ion mobility-based biomolecular investigations.

Acknowledgments

Preparation of this manuscript was funded in part by the National Institutes of Health through grants R01DK071801 and RF1 AG052324, and the National Science Foundation (CHE-2108223 and IOS-2010789). L.L. acknowledges a Vilas Distinguished Achievement Professorship and Charles Melbourne Johnson Distinguished Chair Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy.

Tables

Instrument platform	Agilent 6560 IM-QTOF	Waters Select Series Cyclic IMS	Bruker TimsTof	Thermo FAIMS Pro
IMS type	DTIMS	CIMS	TIMS	FAIMS
CCS measurement	Direct; more accurate	Need calibration against known CCS	Need calibration against known CCS	None
IM Resolving power	100-200	60-80 for single pass; >750 at 100+ passes	>300	< 30; substantial increases in planar FAIMS.
Advantages	Allows for first-principles measurement of CCS	Can achieve super high IM resolving power and is capable of tandem IMS	Very high duty cycle, built-in tools augment MS^2 coverage	Easy to interface with MS platforms
Drawbacks	Low duty cycle requires multiplexing	Reduced transmission and low duty cycle with multiple passes	Analysis of the collected data can be complex	Cannot obtain CCS value
Recommended Applications	Accurate CCS measurement	Small molecule mixtures with small CCS difference; IM separation of fragmentation ions.	Discovery-based omics; complex biological mixtures.	Biological screening, discovery-based or targeted omics

Table 9.1 General information on each IMS archetype and their advantages and drawbacks.

Model	Lauch Year	IMS	Advantages	Drawbacks
Select Series Cyclic IMS	2019	Cyclic IMS technology. Resolving power: 60-80 for single pass; >750 for 100+ passes	High ion mobility resolving power; suitable for structural investigations of small molecules and protein complexes	Reduced transmission and "wrap-up" effect due to multiple passes in cIM devices hinder omics-related applications
Synapt XS	2019	T-Wave IMS technology. Resolving power ~25	Enhanced sensitivity and resolution compared to previous Synapt Model; suitable for omics-related research	Limited ion mobility resolving power
MALDI Synapt G2-Si HDMS	2013		Versatile MS platform for ESI and MALDI; suitable for mass spectrometry imaging	

Table 9.2 Comparison of three different IM-MS system from the Waters company.

Figures

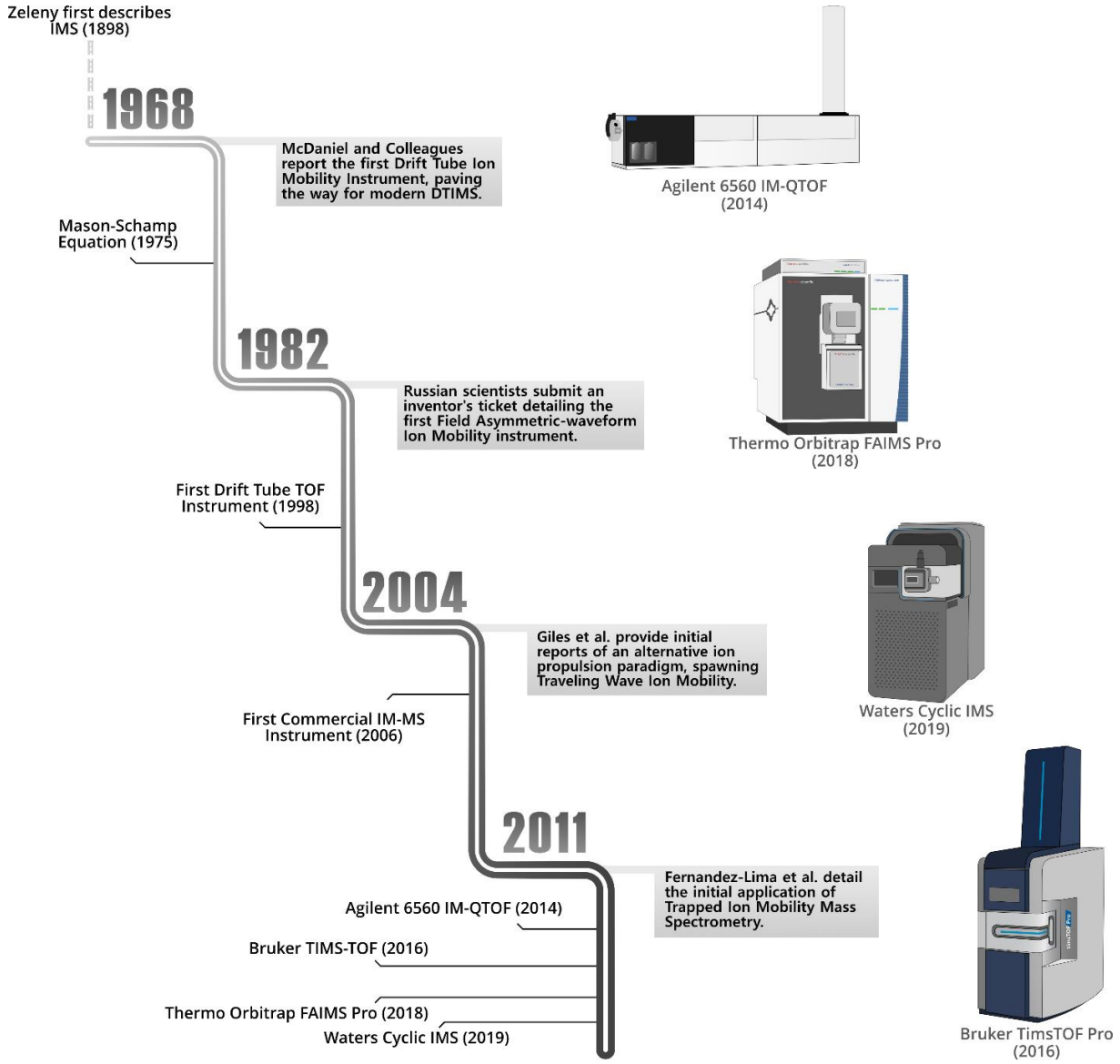


Figure 9.1 Timeline of analytical innovations and fundamental reports that gave rise to the current iterations of high-end ion mobility spectrometry instrumentation.

Though current high-end ion mobility instrumentation has seemed to arrive instantaneously and in proximity, these technical advances are the result of decades on incremental improvements. Renderings of current IM instrumentation are placed next to the milestones responsible for their eventual development.

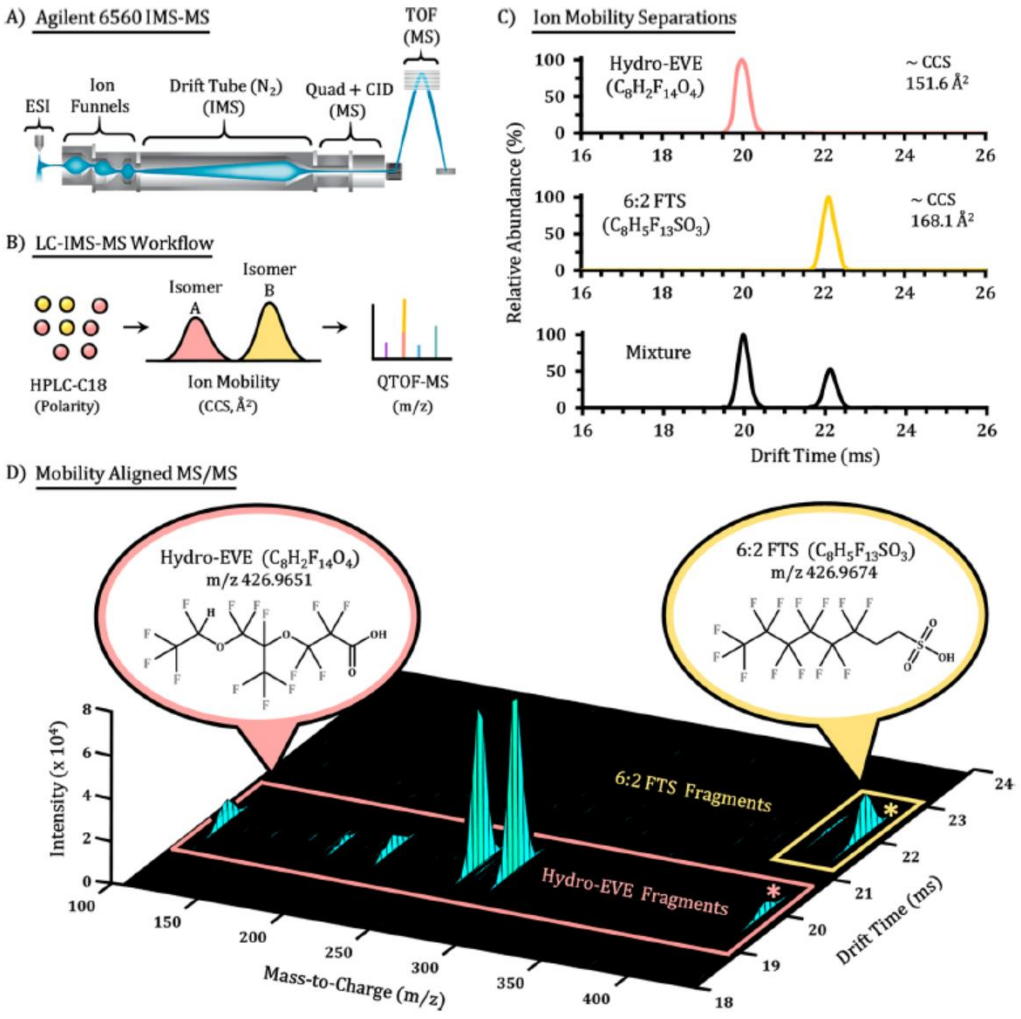


Figure 9.2 Agilent 6560 IM-QTOF.

(A) A representative schematic of the Agilent 6560 instrumentation. (B) The common RPLC-IMS-MS workflow to characterize isomers. (C) Isobaric and isomeric separation can be achieved by the IMS distribution. (D) Drift time aligned MS/MS fragments of isobars can further validate the structural differences. Reprinted from Dodds *et al.*[44] with permission.

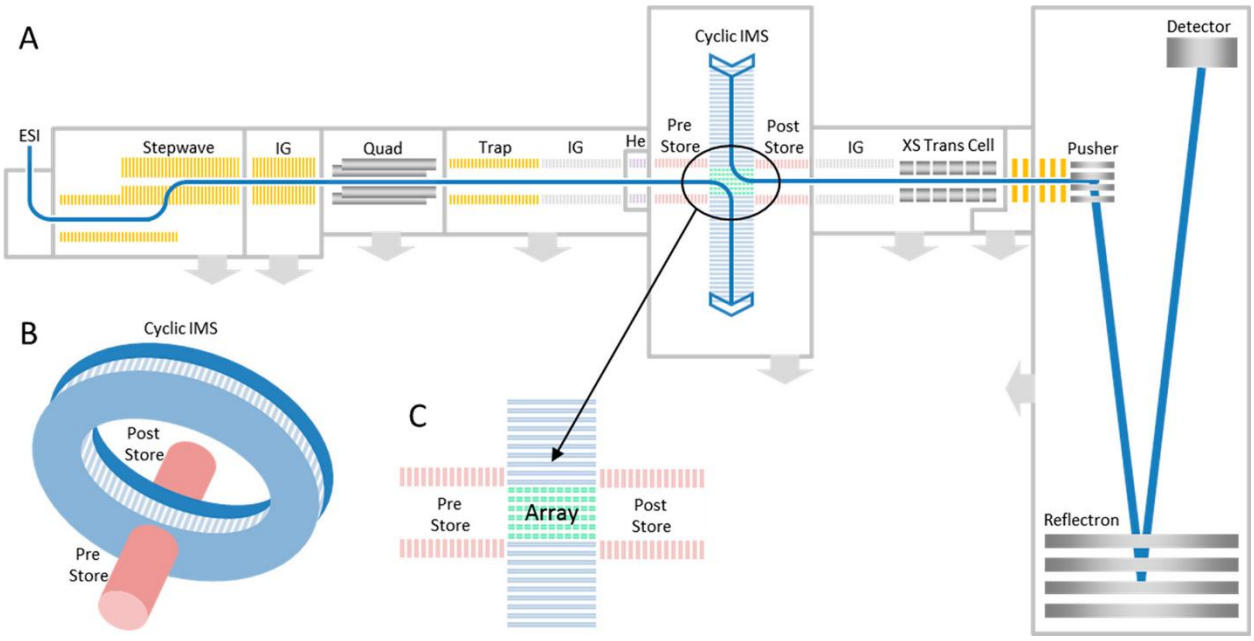


Figure 9.3 Waters Cyclic Traveling Wave Ion Mobility.

(A) A schematic of the Waters cyclic IMS instrumentation. (B) The cyclic IMS cell. (C) Pre- and post-store devices enable the multifunction of cyclic IMS. Reprinted from Eldrid, *et al.*[300] with permission. Copyright 2019 American Chemical Society.

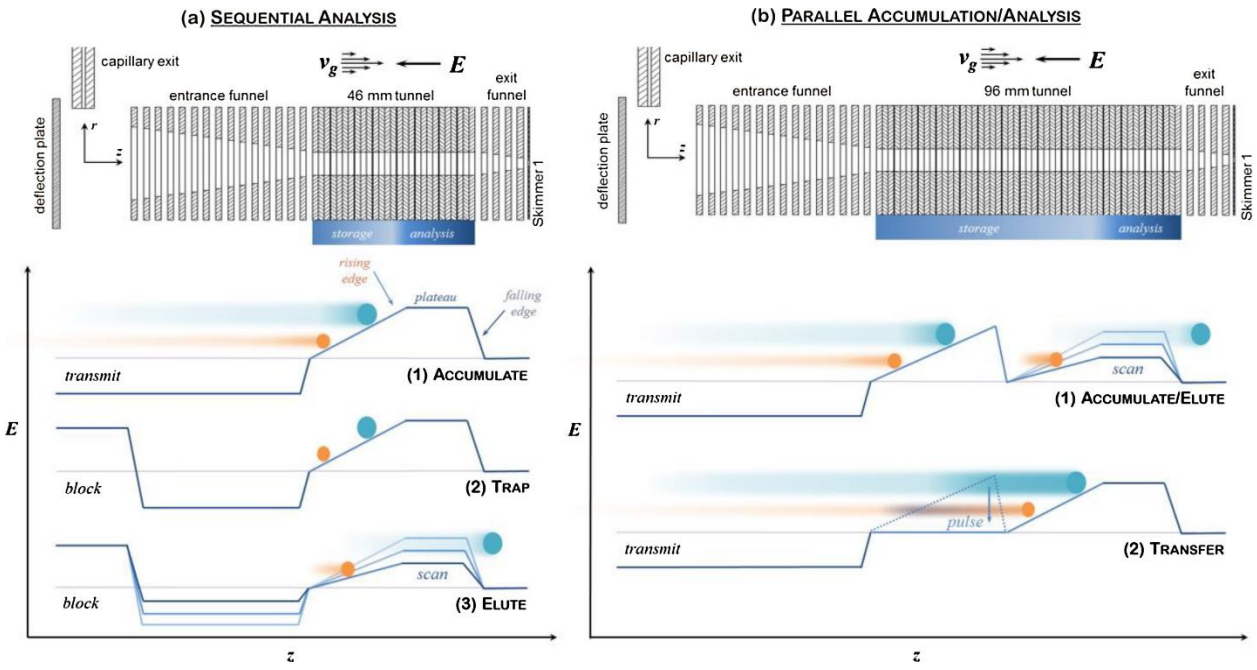


Figure 9.4 Bruker Trapped Ion Mobility.

A schematic representation of the position and time-dependent potential inside the TIMS cell. By establishing a rising edge in the accumulation and trapping steps, ions that receive more energy from the carrier gas are physically pushed further along the cell, and this separation in space during the trapping step allows for sequential elution once the potential gradient is lowered. Reprinted from Ridgeway *et al*.[142] with permission.

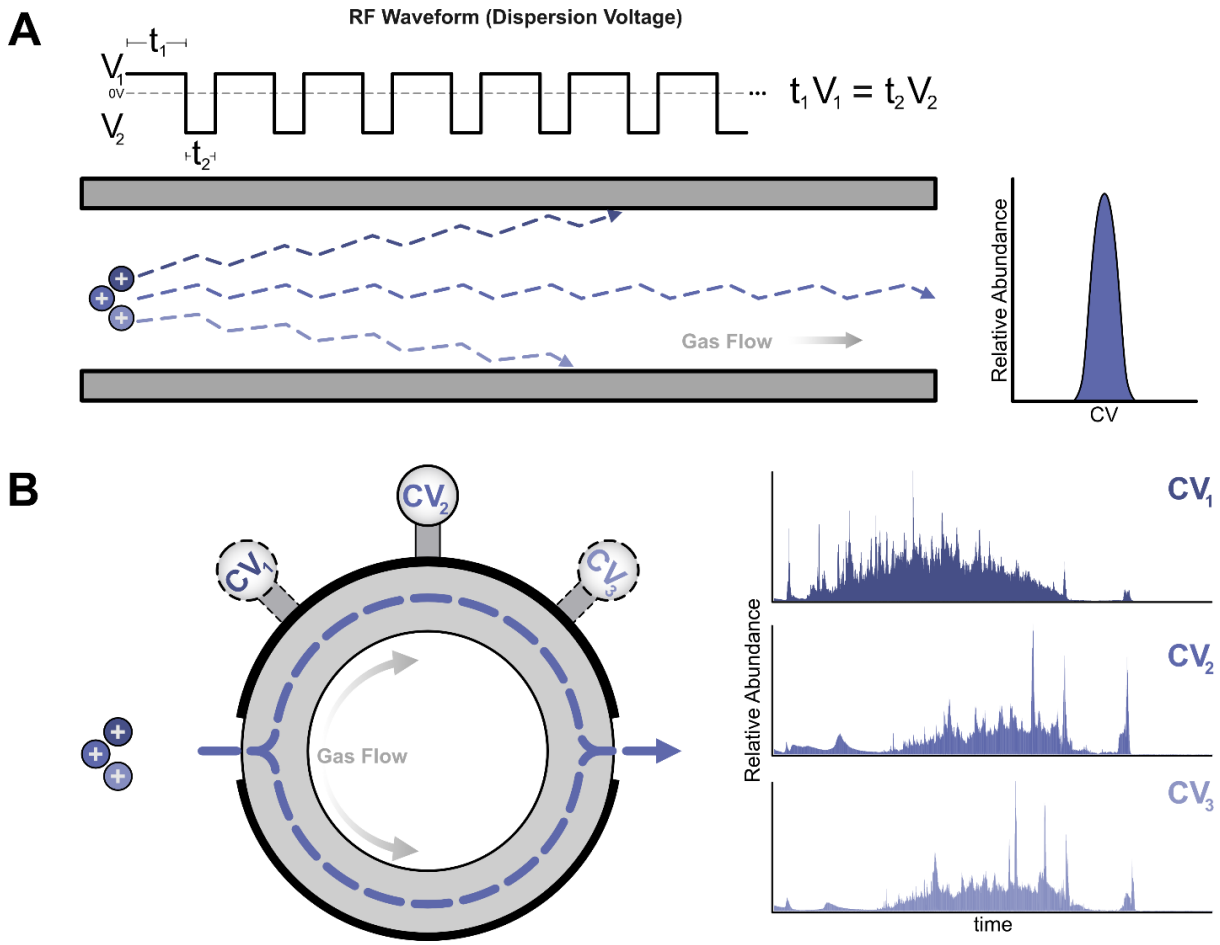


Figure 9.5 High Field Asymmetric-Waveform Ion Mobility

A) Schematic of planar FAIMS depicting typical ion movement in response to a given asymmetrical dispersion voltage (DV); a compensation voltage applied to the opposing electrode allows for detection of ions with compatible electrophoretic character. **B)** Schematic of cylindrical FAIMS, the implementation available in the Thermo FAIMS Pro. Multiple unique CVs may be applied in each run to improve sample coverage and profiling depth.

References

1. Zeleny, J., VI. On the ratio of the velocities of the two ions produced in gases by Röntgen radiation; and on some related phenomena. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1898,** *46* (278), 120-154.

2. McAfee, K. B.; Edelson, D. In *Identification and Mobility of Ions in a Townsend Discharge by Time-resolved Mass Spectrometry*, 1963.

3. Mason, E. A.; Schamp Jr, H. W., Mobility of gaseous lons in weak electric fields. *Annals of physics* **1958,** *4* (3), 233-270.

4. Dodds, J. N.; Baker, E. S., Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead. *Journal of The American Society for Mass Spectrometry* **2019,** *30* (11), 2185-2195.

5. Stow, S.; Fjeldsted, J., Fundamentals of Uniform-field Drift Tube Ion Mobility and Collision Cross Section. In *Ion Mobility-Mass Spectrometry*, 2021; pp 52-82.

6. May, J. C.; McLean, J. A., Ion mobility-mass spectrometry: time-dispersive instrumentation. *Anal Chem* **2015,** *87* (3), 1422-36.

7. Hoaglund, C. S.; Valentine, S. J.; Sporleder, C. R.; Reilly, J. P.; Clemmer, D. E., Three-dimensional ion mobility/TOFMS analysis of electrosprayed biomolecules. *Analytical chemistry* **1998,** *70* (11), 2236-2242.

8. Tang, K.; Shvartsburg, A. A.; Lee, H.-N.; Prior, D. C.; Buschbach, M. A.; Li, F.; Tolmachev, A. V.; Anderson, G. A.; Smith, R. D., High-sensitivity ion mobility spectrometry/mass spectrometry using electrodynamic ion funnel interfaces. *Analytical chemistry* **2005,** *77* (10), 3330-3339.

9. Baker, E. S.; Clowers, B. H.; Li, F.; Tang, K.; Tolmachev, A. V.; Prior, D. C.; Belov, M. E.; Smith, R. D., Ion mobility spectrometry-mass spectrometry performance using electrodynamic ion funnels and elevated drift gas pressures. *J Am Soc Mass Spectrom* **2007,** *18* (7), 1176-87.

10. May, J. C.; Goodwin, C. R.; Lareau, N. M.; Leaptrot, K. L.; Morris, C. B.; Kurulugama, R. T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E.; Overney, G.; Imatani, K.; Stafford, G. C.; Fjeldsted, J. C.; McLean, J. A., Conformational ordering of biomolecules in the gas phase: nitrogen collision cross sections measured on a prototype high resolution drift tube ion mobility-mass spectrometer. *Anal Chem* **2014,** *86* (4), 2107-16.

11. Kurulugama, R. T.; Darland, E.; Kuhlmann, F.; Stafford, G.; Fjeldsted, J., Evaluation of drift gas selection in complex sample analyses using a high performance drift tube ion mobility-QTOF mass spectrometer. *Analyst* **2015,** *140* (20), 6834-44.

12. Groessl, M.; Graf, S.; Knochenmuss, R., High resolution ion mobility-mass spectrometry for separation and identification of isomeric lipids. *Analyst* **2015,** *140* (20), 6904-11.

13. Causon, T. J.; Si-Hung, L.; Newton, K.; Kurulugama, R. T.; Fjeldsted, J.; Hann, S., Fundamental study of ion trapping and multiplexing using drift tube-ion mobility time-of-flight mass spectrometry for non-targeted metabolomics. *Analytical and bioanalytical chemistry* **2019,** *411* (24), 6265-6274.

14. Belov, M. E.; Buschbach, M. A.; Prior, D. C.; Tang, K.; Smith, R. D., Multiplexed ion mobility spectrometry-orthogonal time-of-flight mass spectrometry. *Analytical chemistry* **2007,** *79* (6), 2451-2462.

15. Morrison, K. A.; Siems, W. F.; Clowers, B. H., Augmenting Ion Trap Mass Spectrometers Using a Frequency Modulated Drift Tube Ion Mobility Spectrometer. *Anal Chem* **2016,** *88* (6), 3121-9.

16. Causon, T. J.; Si-Hung, L.; Newton, K.; Kurulugama, R. T.; Fjeldsted, J.; Hann, S., Fundamental study of ion trapping and multiplexing using drift tube-ion mobility time-of-flight mass spectrometry for non-targeted metabolomics. *Anal Bioanal Chem* **2019,** *411* (24), 6265-6274.

17. May, J. C.; Knochenmuss, R.; Fjeldsted, J. C.; McLean, J. A., Resolution of isomeric mixtures in ion mobility using a combined demultiplexing and peak deconvolution technique. *Analytical Chemistry* **2020,** *92* (14), 9482-9492.

18. Clowers, B. H.; Siems, W. F.; Hill, H. H.; Massick, S. M., Hadamard transform ion mobility spectrometry. *Analytical chemistry* **2006,** *78* (1), 44-51.

19. Prost, S. A.; Crowell, K. L.; Baker, E. S.; Ibrahim, Y. M.; Clowers, B. H.; Monroe, M. E.; Anderson, G. A.; Smith, R. D.; Payne, S. H., Detecting and removing data artifacts in Hadamard transform ion mobility-mass spectrometry measurements. *J Am Soc Mass Spectrom* **2014,** *25* (12), 2020-2027.

20. May, J. C.; Knochenmuss, R.; Fjeldsted, J. C.; McLean, J. A., Resolution of Isomeric Mixtures in Ion Mobility Using a Combined Demultiplexing and Peak Deconvolution Technique. *Anal Chem* **2020,** *92* (14), 9482-9492.

21. Harris, G. A.; Kwasnik, M.; Fernandez, F. M., Direct analysis in real time coupled to multiplexed drift tube ion mobility spectrometry for detecting toxic chemicals. *Anal Chem* **2011,** *83* (6), 1908-15.

22. Kwasnik, M.; Caramore, J.; Fernández, F. M., Digitally-multiplexed nanoelectrospray ionization atmospheric pressure drift tube ion mobility spectrometry. *Analytical chemistry* **2009,** *81* (4), 1587-1594.

23. Dodds, J. N.; Baker, E. S., Improving the Speed and Selectivity of Newborn Screening Using Ion Mobility Spectrometry-Mass Spectrometry. *Anal Chem* **2021,** *93* (51), 17094-17102.

24. Campuzano, I.; Bush, M. F.; Robinson, C. V.; Beaumont, C.; Richardson, K.; Kim, H.; Kim, H. I., Structural characterization of drug-like compounds by ion mobility mass spectrometry: comparison of theoretical and experimentally derived nitrogen collision cross sections. *Anal Chem* **2012,** *84* (2), 1026-33.

25. Salbo, R.; Bush, M. F.; Naver, H.; Campuzano, I.; Robinson, C. V.; Pettersson, I.; Jørgensen, T. J.; Haselmann, K. F., Traveling-wave ion mobility mass spectrometry of protein complexes: accurate calibrated collision cross-sections of human insulin oligomers. *Rapid Communications in Mass Spectrometry* **2012,** *26* (10), 1181-1193.

26. Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T., Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Analytical chemistry* **2010,** *82* (22), 9557-9565.

27. Allen, S. J.; Schwartz, A. M.; Bush, M. F., Effects of polarity on the structures and charge states of native-like proteins and protein complexes in the gas phase. *Analytical chemistry* **2013,** *85* (24), 12055-12061.

28. Allen, S. J.; Giles, K.; Gilbert, T.; Bush, M. F., Ion mobility mass spectrometry of peptide, protein, and protein complex ions using a radio-frequency confining drift cell. *Analyst* **2016,** *141* (3), 884-891.

29. Forsythe, J. G.; Petrov, A. S.; Walker, C. A.; Allen, S. J.; Pellissier, J. S.; Bush, M. F.; Hud, N. V.; Fernández, F. M., Collision cross section calibrants for negative ion mode traveling wave ion mobility-mass spectrometry. *Analyst* **2015,** *140* (20), 6853-6861.

30. Hines, K. M.; Ross, D. H.; Davidson, K. L.; Bush, M. F.; Xu, L., Large-scale structural characterization of drug and drug-like compounds by high-throughput ion mobility-mass spectrometry. *Analytical chemistry* **2017,** *89* (17), 9023-9030.

31. Bush, M. F.; Campuzano, I. D.; Robinson, C. V., Ion mobility mass spectrometry of peptide ions: effects of drift gas and calibration strategies. *Anal Chem* **2012,** *84* (16), 7124-30.

32. Stow, S. M.; Causon, T. J.; Zheng, X.; Kurulugama, R. T.; Mairinger, T.; May, J. C.; Rennie, E. E.; Baker, E. S.; Smith, R. D.; McLean, J. A.; Hann, S.; Fjeldsted, J. C., An Interlaboratory Evaluation of Drift Tube Ion Mobility-Mass Spectrometry Collision Cross Section Measurements. *Anal Chem* **2017,** *89* (17), 9048-9055.

33. Zhou, Z.; Luo, M.; Chen, X.; Yin, Y.; Xiong, X.; Wang, R.; Zhu, Z.-J., Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. *Nature communications* **2020,** *11* (1), 1-13.

34. Zhou, Z.; Shen, X.; Tu, J.; Zhu, Z.-J., Large-scale prediction of collision cross-section values for metabolites in ion mobility-mass spectrometry. *Analytical chemistry* **2016,** *88* (22), 11084-11091.

35. Zhou, Z.; Xiong, X.; Zhu, Z.-J., MetCCS predictor: a web server for predicting collision cross-section values of metabolites in ion mobility-mass spectrometry based metabolomics. *Bioinformatics* **2017,** *33* (14), 2235-2237.

36. Zhou, Z.; Tu, J.; Xiong, X.; Shen, X.; Zhu, Z.-J., LipidCCS: prediction of collision cross-section values for lipids with high precision to support ion mobility–mass spectrometry-based lipidomics. *Analytical chemistry* **2017,** *89* (17), 9559-9566.

37. Plante, P.-L.; Francovic-Fontaine, É.; May, J. C.; McLean, J. A.; Baker, E. S.; Laviolette, F.; Marchand, M.; Corbeil, J., Predicting ion mobility collision cross-sections using a deep neural network: DeepCCS. *Analytical chemistry* **2019,** *91* (8), 5191-5199.

38. Ross, D. H.; Xu, L., Determination of drugs and drug metabolites by ion mobility-mass spectrometry: A review. *Analytica Chimica Acta* **2021,** *1154*, 338270.

39. Tu, J.; Zhou, Z.; Li, T.; Zhu, Z.-J., The emerging role of ion mobility-mass spectrometry in lipidomics to facilitate lipid separation and identification. *TrAC Trends in Analytical Chemistry* **2019,** *116*, 332-339.

40. Unsihuay, D.; Yin, R.; Sanchez, D. M.; Yang, M.; Li, Y.; Sun, X.; Dey, S. K.; Laskin, J., High-resolution imaging and identification of biomolecules using Nano-DESI coupled to ion mobility spectrometry. *Analytica Chimica Acta* **2021,** *1186*, 339085.

41. Mu, Y.; Schulz, B. L.; Ferro, V., Applications of ion mobility-mass spectrometry in carbohydrate chemistry and glycobiology. *Molecules* **2018,** *23* (10), 2557.

42. Li, G.; Delafield, D. G.; Li, L., Improved structural elucidation of peptide isomers and their receptors using advanced ion mobility-mass spectrometry. *TrAC Trends in Analytical Chemistry* **2020,** *124*.

43. Lee, H. H. L.; Chae, S. Y.; Son, M. K.; Kim, H. I., Ion Mobility Signatures of Glutamine-Containing Tryptic Peptides in the Gas Phase. *Mass Spectrometry Letters* **2021,** *12* (4), 137-145.

44. Dodds, J. N.; Hopkins, Z. R.; Knappe, D. R. U.; Baker, E. S., Rapid Characterization of Per- and Polyfluoroalkyl Substances (PFAS) by Ion Mobility Spectrometry-Mass Spectrometry (IMS-MS). *Anal Chem* **2020,** *92* (6), 4427-4435.

45. Kirkwood, K. I.; Fleming, J.; Nguyen, H.; Reif, D. M.; Baker, E. S.; Belcher, S. M., Utilizing Pine Needles to Temporally and Spatially Profile Per-and Polyfluoroalkyl Substances (PFAS). *Environmental Science & Technology* **2022,** *56* (6), 3441-3451.

46. Davis, D. E., Jr.; Leaptrot, K. L.; Koomen, D. C.; May, J. C.; Cavalcanti, G. A.; Padilha, M. C.; Pereira, H. M. G.; McLean, J. A., Multidimensional Separations of Intact Phase II Steroid Metabolites Utilizing LC-Ion Mobility-HRMS. *Anal Chem* **2021,** *93* (31), 10990-10998.

47. Chouinard, C. D.; Beekman, C. R.; Kemperman, R. H.; King, H. M.; Yost, R. A., Ion mobility-mass spectrometry separation of steroid structural isomers and epimers. *International Journal for Ion Mobility Spectrometry* **2017,** *20* (1), 31-39.

48. Maddox, S. W.; Olsen, S. S.; Velosa, D. C.; Burkus-Matesevac, A.; Peverati, R.; Chouinard, C. D., Improved identification of isomeric steroids using the Paterno-Buchi reaction with ion mobility-mass spectrometry. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (10), 2086-2092.

49. Velosa, D. C.; Rivera, M. E.; Neal, S. P.; Olsen, S. S.; Burkus-Matesevac, A.; Chouinard, C. D., Toward Routine Analysis of Anabolic Androgenic Steroids in Urine Using Ion Mobility-Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2021,** *33* (1), 54-61.

50. Zheng, X.; Smith, F. B.; Aly, N. A.; Cai, J.; Smith, R. D.; Patterson, A. D.; Baker, E. S., Evaluating the structural complexity of isomeric bile acids with ion mobility spectrometry. *Anal Bioanal Chem* **2019,** *411* (19), 4673-4682.

51. Butler, K. E.; Dodds, J. N.; Flick, T.; Campuzano, I. D.; Baker, E. S., High-Resolution Demultiplexing (HRdm) Ion Mobility Spectrometry–Mass Spectrometry for Aspartic and Isoaspartic Acid Determination and Screening. *Analytical Chemistry* **2022,** *94* (16), 6191-6199.

52. Poad, B. L. J.; Zheng, X.; Mitchell, T. W.; Smith, R. D.; Baker, E. S.; Blanksby, S. J., Online Ozonolysis Combined with Ion Mobility-Mass Spectrometry Provides a New Platform for Lipid Isomer Analyses. *Anal Chem* **2018,** *90* (2), 1292-1300.

53. Harris, R. A.; May, J. C.; Stinson, C. A.; Xia, Y.; McLean, J. A., Determining Double Bond Position in Lipids Using Online Ozonolysis Coupled to Liquid Chromatography and Ion Mobility-Mass Spectrometry. *Anal Chem* **2018,** *90* (3), 1915-1924.

54. Kuo, S.-T.; Tang, S.; Russell, D. H.; Yan, X., Characterization of lipid carbon–carbon double-bond isomerism via ion mobility-mass spectrometry (IMS-MS) combined with cuprous ion-induced fragmentation. *International Journal of Mass Spectrometry* **2022**, 116889.

55. Xie, C.; Wu, Q.; Zhang, S.; Wang, C.; Gao, W.; Yu, J.; Tang, K., Improving glycan isomeric separation via metal ion incorporation for drift tube ion mobility-mass spectrometry. *Talanta* **2020,** *211*, 120719.

56. Zheng, X.; Zhang, X.; Schocker, N. S.; Renslow, R. S.; Orton, D. J.; Khamsi, J.; Ashmus, R. A.; Almeida, I. C.; Tang, K.; Costello, C. E., Enhancing glycan isomer separations with metal ions and positive and negative polarity ion mobility spectrometry-mass spectrometry analyses. *Analytical and bioanalytical chemistry* **2017,** *409* (2), 467-476.

57. Butler, K. E.; Kalmar, J. G.; Muddiman, D. C.; Baker, E. S., Utilizing liquid chromatography, ion mobility spectrometry, and mass spectrometry to assess INLIGHT™ derivatized N-linked glycans in biological samples. *Analytical and bioanalytical chemistry* **2022,** *414* (1), 623-637.

58. Mesa Sanchez, D.; Creger, S.; Singla, V.; Kurulugama, R. T.; Fjeldsted, J.; Laskin, J., Ion Mobility-Mass Spectrometry Imaging Workflow. *J Am Soc Mass Spectrom* **2020,** *31* (12), 2437-2442.

59. Unsihuay, D.; Qiu, J.; Swaroop, S.; Nagornov, K. O.; Kozhinov, A. N.; Tsybin, Y. O.; Kuang, S.; Laskin, J., Imaging of triglycerides in tissues using nanospray desorption electrospray ionization (Nano-DESI) mass spectrometry. *International journal of mass spectrometry* **2020,** *448*, 116269.

60. Ekelöf, M. n.; Dodds, J.; Khodjaniyazova, S.; Garrard, K. P.; Baker, E. S.; Muddiman, D. C., Coupling IR-MALDESI with drift tube ion mobility-mass spectrometry for high-throughput screening and imaging applications. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (3), 642-650.

61. Ben-Nissan, G.; Sharon, M., The application of ion-mobility mass spectrometry for structure/function investigation of protein complexes. *Current Opinion in Chemical Biology* **2018,** *42*, 25-33.

62. Heo, C. E.; Kim, M.; Son, M. K.; Hyun, D. G.; Heo, S. W.; Kim, H. I., Ion Mobility Mass Spectrometry Analysis of Oxygen Affinity-Associated Structural Changes in Hemoglobin. *Journal of the American Society for Mass Spectrometry* **2021,** *32* (10), 2528-2535.

63. Gabelica, V.; Livet, S.; Rosu, F., Optimizing native ion mobility Q-TOF in helium and nitrogen for very fragile noncovalent structures. *Journal of The American Society for Mass Spectrometry* **2018,** *29* (11), 2189-2198.

64. Dixit, S. M.; Polasky, D. A.; Ruotolo, B. T., Collision induced unfolding of isolated proteins in the gas phase: past, present, and future. *Current opinion in chemical biology* **2018,** *42*, 93-100.

65. Vallejo, D. D.; Polasky, D. A.; Kurulugama, R. T.; Eschweiler, J. D.; Fjeldsted, J. C.; Ruotolo, B. T., A Modified Drift Tube Ion Mobility-Mass Spectrometer for Charge-Multiplexed Collision-Induced Unfolding. *Anal Chem* **2019,** *91* (13), 8137-8146.

66. Vallejo, D. D.; Kang, J.; Coghlan, J.; Ramírez, C. R.; Polasky, D. A.; Kurulugama, R. T.; Fjeldsted, J. C.; Schwendeman, A. A.; Ruotolo, B. T., Collision-Induced Unfolding Reveals Stability Differences in Infliximab Therapeutics under Native and Heat Stress Conditions. *Analytical Chemistry* **2021,** *93* (48), 16166-16174.

67. Gadkari, V. V.; Ramírez, C. R.; Vallejo, D. D.; Kurulugama, R. T.; Fjeldsted, J. C.; Ruotolo, B. T., Enhanced collision induced unfolding and electron capture dissociation of native-like protein ions. *Analytical chemistry* **2020,** *92* (23), 15489-15496.

68. Vallejo, D. D.; Jeon, C. K.; Parson, K. F.; Herderschee, H. R.; Eschweiler, J. D.; Filoti, D. I.; Ruotolo, B. T., Ion Mobility–Mass Spectrometry Reveals the Structures and Stabilities of Biotherapeutic Antibody Aggregates. *Analytical Chemistry* **2022**.

69. Zheng, X.; Aly, N. A.; Zhou, Y.; Dupuis, K. T.; Bilbao, A.; Paurus, V. L.; Orton, D. J.; Wilson, R.; Payne, S. H.; Smith, R. D., A structural examination and collision cross section database for over 500 metabolites and xenobiotics using drift tube ion mobility spectrometry. *Chemical Science* **2017,** *8* (11), 7724-7736.

70. Feuerstein, M. L.; Kurulugama, R. T.; Hann, S.; Causon, T., Novel acquisition strategies for metabolomics using drift tube ion mobility-quadrupole resolved all ions time-of-flight mass spectrometry (IM-QRAI-TOFMS). *Analytica Chimica Acta* **2021,** *1163*, 338508.

71. Nichols, C. M.; Dodds, J. N.; Rose, B. S.; Picache, J. A.; Morris, C. B.; Codreanu, S. G.; May, J. C.; Sherrod, S. D.; McLean, J. A., Untargeted molecular discovery in primary metabolism: collision cross section as a molecular descriptor in ion mobility-mass spectrometry. *Analytical chemistry* **2018,** *90* (24), 14484-14492.

72. Zhang, X.; Kew, K.; Reisdorph, R.; Sartain, M.; Powell, R.; Armstrong, M.; Quinn, K.; Cruickshank-Quinn, C.; Walmsley, S.; Bokatzian, S.; Darland, E.; Rain, M.; Imatani, K.; Reisdorph, N., Performance of a High-Pressure Liquid Chromatography-Ion Mobility-Mass Spectrometry System for Metabolic Profiling. *Anal Chem* **2017,** *89* (12), 6384-6391.

73. Picache, J. A.; Rose, B. S.; Balinski, A.; Leaptrot, K. L.; Sherrod, S. D.; May, J. C.; McLean, J. A., Collision cross section compendium to annotate and predict multi-omic compound identities. *Chemical science* **2019,** *10* (4), 983-993.

74. Odenkirk, M. T.; Baker, E. S., Utilizing drift tube ion mobility spectrometry for the evaluation of metabolites and xenobiotics. In *Ion Mobility-Mass Spectrometry*, Springer: 2020; pp 35-54.

75. Zhou, Z.; Luo, M.; Chen, X.; Yin, Y.; Xiong, X.; Wang, R.; Zhu, Z. J., Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. *Nat Commun* **2020,** *11* (1), 4334.

76. Drouin, N.; Mielcarek, A.; Wenz, C.; Rudaz, S., Evaluation of ion mobility in capillary electrophoresis coupled to mass spectrometry for the identification in metabolomics. *Electrophoresis* **2021,** *42* (4), 342-349.

77. Blaženović, I.; Shen, T.; Mehta, S. S.; Kind, T.; Ji, J.; Piparo, M.; Cacciola, F.; Mondello, L.; Fiehn, O., Increasing compound identification rates in untargeted lipidomics research with liquid chromatography drift time–ion mobility mass spectrometry. *Analytical chemistry* **2018,** *90* (18), 10758-10764.

78. da Silva, K. M.; Iturrospe, E.; Heyrman, J.; Koelmel, J. P.; Cuykx, M.; Vanhaecke, T.; Covaci, A.; van Nuijs, A. L., Optimization of a liquid chromatography-ion mobility-high resolution mass spectrometry platform for untargeted lipidomics and application to HepaRG cell extracts. *Talanta* **2021,** *235*, 122808.

79. Chen, X.; Yin, Y.; Zhou, Z.; Li, T.; Zhu, Z.-J., Development of a combined strategy for accurate lipid structural identification and quantification in ion-mobility mass spectrometry based untargeted lipidomics. *Analytica Chimica Acta* **2020,** *1136*, 115-124.

80. Rose, B. S.; Leaptrot, K. L.; Harris, R. A.; Sherrod, S. D.; May, J. C.; McLean, J. A., High Confidence Shotgun Lipidomics Using Structurally Selective Ion Mobility-Mass Spectrometry. In *Mass Spectrometry-Based Lipidomics*, Springer: 2021; pp 11-37.

81. Silva, A. C. R.; Garrett, R.; Rezende, C. M.; Meckelmann, S. W., Lipid characterization of arabica and robusta coffee beans by liquid chromatography-ion mobility-mass spectrometry. *Journal of Food Composition and Analysis* **2022,** *111*, 104587.

82. Kirkwood, K. I.; Christopher, M. W.; Burgess, J. L.; Littau, S. R.; Foster, K.; Richey, K.; Pratt, B. S.; Shulman, N.; Tamura, K.; MacCoss, M. J., Development and Application of Multidimensional Lipid Libraries to Investigate Lipidomic Dysregulation Related to Smoke Inhalation Injury Severity. *Journal of proteome research* **2021,** *21* (1), 232-242.

83. Odenkirk, M. T.; Horman, B. M.; Dodds, J. N.; Patisaul, H. B.; Baker, E. S., Combining Micropunch Histology and Multidimensional Lipidomic Measurements for In-Depth Tissue Mapping. *ACS Measurement Science Au* **2021,** *2* (1), 67-75.

84. Baker, E. S.; Burnum-Johnson, K. E.; Jacobs, J. M.; Diamond, D. L.; Brown, R. N.; Ibrahim, Y. M.; Orton, D. J.; Piehowski, P. D.; Purdy, D. E.; Moore, R. J.; Danielson, W. F., 3rd; Monroe, M. E.; Crowell, K. L.; Slysz, G. W.; Gritsenko, M. A.; Sandoval, J. D.; Lamarche, B. L.; Matzke, M. M.; Webb-Robertson, B. J.; Simons, B. C.; McMahon, B. J.; Bhattacharya, R.; Perkins, J. D.; Carithers, R. L., Jr.; Strom, S.; Self, S. G.; Katze, M. G.; Anderson, G. A.; Smith, R. D., Advancing the high throughput identification of liver fibrosis protein signatures using multiplexed ion mobility spectrometry. *Mol Cell Proteomics* **2014,** *13* (4), 1119-27.

85. Kedia, K.; Wendler, J. P.; Baker, E. S.; Burnum-Johnson, K. E.; Jarsberg, L. G.; Stratton, K. G.; Wright, A. T.; Piehowski, P. D.; Gritsenko, M. A.; Lewinsohn, D. M.; Sigal, G.

B.; Weiner, M. H.; Smith, R. D.; Jacobs, J. M.; Nahid, P., Application of multiplexed ion mobility spectrometry towards the identification of host protein signatures of treatment effect in pulmonary tuberculosis. *Tuberculosis (Edinb)* **2018,** *112*, 52-61.

86. Belova, L.; Caballero-Casero, N.; van Nuijs, A. L. N.; Covaci, A., Ion Mobility-High-Resolution Mass Spectrometry (IM-HRMS) for the Analysis of Contaminants of Emerging Concern (CECs): Database Compilation and Application to Urine Samples. *Anal Chem* **2021,** *93* (16), 6428-6436.

87. Bilbao, A.; Gibbons, B. C.; Stow, S. M.; Kyle, J. E.; Bloodsworth, K. J.; Payne, S. H.; Smith, R. D.; Ibrahim, Y. M.; Baker, E. S.; Fjeldsted, J. C., A Preprocessing Tool for Enhanced Ion Mobility-Mass Spectrometry-Based Omics Workflows. *J Proteome Res* **2022,** *21* (3), 798-807.

88. Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H., Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Communications in Mass Spectrometry* **2004,** *18* (20), 2401-2414.

89. Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H., An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry* **2007,** *261* (1), 1-12.

90. Dodds, J. N.; May, J. C.; McLean, J. A., Correlating Resolving Power, Resolution and Collision Cross Section: Unifying Cross Platform Assessment of Separation Efficiency in Ion Mobility Spectrometry. *Analytical chemistry* **2017,** *89* (22), 12176-12184.

91. Shvartsburg, A. A.; Smith, R. D., Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Analytical chemistry* **2008,** *80* (24), 9689-9699.

92. Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S.; Richardson, K.; Langridge, D.; Green, M., A Cyclic Ion Mobility-Mass Spectrometry System. *Analytical Chemistry* **2019,** *91* (13), 8564-8573.

93. Merenbloom, S. I.; Glaskin, R. S.; Henson, Z. B.; Clemmer, D. E., High-Resolution Ion Cyclotron Mobility Spectrometry. *Analytical Chemistry* **2009,** *81* (4), 1482-1487.

94. Deng, L.; Webb, I. K.; Garimella, S. V. B.; Hamid, A. M.; Zheng, X.; Norheim, R. V.; Prost, S. A.; Anderson, G. A.; Sandoval, J. A.; Baker, E. S.; Ibrahim, Y. M.; Smith, R. D., Serpentine Ultralong Path with Extended Routing (SUPER) High Resolution Traveling Wave Ion Mobility-MS using Structures for Lossless Ion Manipulations. *Analytical Chemistry* **2017,** *89* (8), 4628-4634.

95. Hollerbach, A. L.; Li, A.; Prabhakaran, A.; Nagy, G.; Harrilal, C. P.; Conant, C. R.; Norheim, R. V.; Schimelfenig, C. E.; Anderson, G. A.; Garimella, S. V. B.; Smith, R. D.; Ibrahim, Y. M., Ultra-High-Resolution Ion Mobility Separations Over Extended Path

Lengths and Mobility Ranges Achieved using a Multilevel Structures for Lossless Ion Manipulations Module. *Analytical Chemistry* **2020,** *92* (11), 7972-7979.

96. Gabelica, V.; Shvartsburg, A. A.; Afonso, C.; Barran, P.; Benesch, J. L. P.; Bleiholder, C.; Bowers, M. T.; Bilbao, A.; Bush, M. F.; Campbell, J. L.; Campuzano, I. D. G.; Causon, T.; Clowers, B. H.; Creaser, C. S.; De Pauw, E.; Far, J.; Fernandez-Lima, F.; Fjeldsted, J. C.; Giles, K.; Groessl, M.; Hogan Jr, C. J.; Hann, S.; Kim, H. I.; Kurulugama, R. T.; May, J. C.; McLean, J. A.; Pagel, K.; Richardson, K.; Ridgeway, M. E.; Rosu, F.; Sobott, F.; Thalassinos, K.; Valentine, S. J.; Wyttenbach, T., Recommendations for reporting ion mobility Mass Spectrometry measurements. *Mass Spectrometry Reviews* **2019,** *38* (3), 291-320.

97. Richardson, K.; Langridge, D.; Dixit, S. M.; Ruotolo, B. T., An Improved Calibration Approach for Traveling Wave Ion Mobility Spectrometry: Robust, High-Precision Collision Cross Sections. *Analytical Chemistry* **2021,** *93* (7), 3542-3550.

98. Ruger, C. P.; Le Maitre, J.; Maillard, J.; Riches, E.; Palmer, M.; Afonso, C.; Giusti, P., Exploring Complex Mixtures by Cyclic Ion Mobility High-Resolution Mass Spectrometry: Application Toward Petroleum. *Anal Chem* **2021,** *93* (14), 5872-5881.

99. Lillja, J.; Lanekoff, I., Quantitative determination of sn-positional phospholipid isomers in MSn using silver cationization. *Analytical and Bioanalytical Chemistry* **2022**, 1-10.

100. Gao, T.; Lott, A. A.; Huang, F.; Rohokale, R.; Li, Q.; Olivos, H. J.; Chen, S.; Guo, Z., Structural Characterization and Analysis of Different Epimers of Neutral Glycosphingolipid LcGg4 by Ion Mobility Spectrometry-Mass Spectrometry. *Analyst* **2022**.

101. Cho, E.; Riches, E.; Palmer, M.; Giles, K.; Ujma, J.; Kim, S., Isolation of Crude Oil Peaks Differing by m/z approximately 0.1 via Tandem Mass Spectrometry Using a Cyclic Ion Mobility-Mass Spectrometer. *Anal Chem* **2019,** *91* (22), 14268-14274.

102. Cho, E.; Cho, Y.; Rakhmat, S.; Kim, Y. H.; Kim, S., Molecular-Level Structural Analysis of Hydrotreated and Untreated Atmospheric Residue Oils via Atmospheric Pressure Photoionization Cyclic Ion Mobility Mass Spectrometry and Ultrahigh-Resolution Mass Spectrometry. *Energy & Fuels* **2021,** *35* (22), 18163-18169.

103. Ruger, C. P.; Le Maitre, J.; Riches, E.; Palmer, M.; Orasche, J.; Sippula, O.; Jokiniemi, J.; Afonso, C.; Giusti, P.; Zimmermann, R., Cyclic Ion Mobility Spectrometry Coupled to High-Resolution Time-of-Flight Mass Spectrometry Equipped with Atmospheric Solid Analysis Probe for the Molecular Characterization of Combustion Particulate Matter. *J Am Soc Mass Spectrom* **2021,** *32* (1), 206-217.

104. Riches, E.; Palmer, M. E., Application of a novel cyclic ion mobility-mass spectrometer to the analysis of synthetic polymers: A preliminary evaluation. *Rapid Commun Mass Spectrom* **2020,** *34 Suppl 2*, e8710.

105. McCullagh, M.; Goscinny, S.; Palmer, M.; Ujma, J., Investigations into pesticide charge site isomers using conventional IM and cIM systems. *Talanta* **2021,** *234*, 122604.

106. Khabazbashi, S.; Engelhardt, J.; Möckel, C.; Weiss, J.; Kruve, A., Estimation of the concentrations of hydroxylated polychlorinated biphenyls in human serum using ionization efficiency prediction for electrospray. *Analytical and Bioanalytical Chemistry* **2022**, 1-10.

107. Colson, E.; Decroo, C.; Cooper-Shepherd, D.; Caulier, G.; Henoumont, C.; Laurent, S.; De Winter, J.; Flammang, P.; Palmer, M.; Claereboudt, J.; Gerbaux, P., Discrimination of Regioisomeric and Stereoisomeric Saponins from Aesculus hippocastanum Seeds by Ion Mobility Mass Spectrometry. *J Am Soc Mass Spectrom* **2019,** *30* (11), 2228-2237.

108. Cooper-Shepherd, D. A.; Olivos, H. J.; Wu, Z.; Palmer, M. E., Exploiting Self-Association to Evaluate Enantiomeric Composition by Cyclic Ion Mobility–Mass Spectrometry. *Analytical Chemistry* **2022**.

109. McCullagh, M.; Giles, K.; Richardson, K.; Stead, S.; Palmer, M., Investigations into the performance of travelling wave enabled conventional and cyclic ion mobility systems to characterise protomers of fluoroquinolone antibiotic residues. *Rapid Communications in Mass Spectrometry* **2019,** *33*, 11-21.

110. Higton, D.; Palmer, M. E.; Vissers, J. P. C.; Mullin, L. G.; Plumb, R. S.; Wilson, I. D., Use of Cyclic Ion Mobility Spectrometry (cIM)-Mass Spectrometry to Study the Intramolecular Transacylation of Diclofenac Acyl Glucuronide. *Anal Chem* **2021,** *93* (20), 7413-7421.

111. Williamson, D. L.; Bergman, A. E.; Heider, E. C.; Nagy, G., Experimental Measurements of Relative Mobility Shifts Resulting from Isotopic Substitutions with High-Resolution Cyclic Ion Mobility Separations. *Anal Chem* **2022,** *94* (6), 2988-2995.

112. Williamson, D. L.; Nagy, G., Evaluating the Utility of Temporal Compression in High-Resolution Traveling Wave-Based Cyclic Ion Mobility Separations. *ACS Measurement Science Au* **2022**.

113. Ujma, J.; Ropartz, D.; Giles, K.; Richardson, K.; Langridge, D.; Wildgoose, J.; Green, M.; Pringle, S., Cyclic Ion Mobility Mass Spectrometry Distinguishes Anomers and Open-Ring Forms of Pentasaccharides. *J Am Soc Mass Spectrom* **2019,** *30* (6), 1028-1037.

114. Ropartz, D.; Fanuel, M.; Ujma, J.; Palmer, M.; Giles, K.; Rogniaux, H., Structure Determination of Large Isomeric Oligosaccharides of Natural Origin through Multipass and Multistage Cyclic Traveling-Wave Ion Mobility Mass Spectrometry. *Anal Chem* **2019,** *91* (18), 12030-12037.

115. Ropartz, D.; Fanuel, M.; Ollivier, S.; Lissarrague, A.; Benkoulouche, M.; Mulard, L. A.; Andre, I.; Guieysse, D.; Rogniaux, H., Combination of High-Resolution Multistage

Ion Mobility and Tandem MS with High Energy of Activation to Resolve the Structure of Complex Chemoenzymatically Synthesized Glycans. *Anal Chem* **2022,** *94* (4), 2279-2287.

116. Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S.; Richardson, K.; Langridge, D.; Green, M., A Cyclic Ion Mobility-Mass Spectrometry System. *Anal Chem* **2019,** *91* (13), 8564-8573.

117. Ollivier, S.; Tarquis, L.; Fanuel, M.; Li, A.; Durand, J.; Laville, E.; Potocki-Veronese, G.; Ropartz, D.; Rogniaux, H., Anomeric Retention of Carbohydrates in Multistage Cyclic Ion Mobility (IMS(n)): De Novo Structural Elucidation of Enzymatically Produced Mannosides. *Anal Chem* **2021,** *93* (15), 6254-6261.

118. Peterson, T. L.; Nagy, G., Toward Sequencing the Human Milk Glycome: High-Resolution Cyclic Ion Mobility Separations of Core Human Milk Oligosaccharide Building Blocks. *Anal Chem* **2021,** *93* (27), 9397-9407.

119. Peterson, T. L.; Nagy, G., Rapid cyclic ion mobility separations of monosaccharide building blocks as a first step toward a high-throughput reaction screening platform for carbohydrate syntheses. *RSC Advances* **2021,** *11* (63), 39742-39747.

120. Sastre Torano, J.; Gagarinov, I. A.; Vos, G. M.; Broszeit, F.; Srivastava, A. D.; Palmer, M.; Langridge, J. I.; Aizpurua-Olaizola, O.; Somovilla, V. J.; Boons, G. J., Ion-Mobility Spectrometry Can Assign Exact Fucosyl Positions in Glycans and Prevent Misinterpretation of Mass-Spectrometry Data After Gas-Phase Rearrangement. *Angew Chem Int Ed Engl* **2019,** *58* (49), 17616-17620.

121. Cavallero, G. J.; Zaia, J., Resolving Heparan Sulfate Oligosaccharide Positional Isomers Using Hydrophilic Interaction Liquid Chromatography-Cyclic Ion Mobility Mass Spectrometry. *Anal Chem* **2022,** *94* (5), 2366-2374.

122. Sanda, M.; Morrison, L.; Goldman, R., N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Anal Chem* **2021,** *93* (4), 2003-2009.

123. Jia, C.; Lietz, C. B.; Yu, Q.; Li, L., Site-specific characterization of (D)-amino acid containing peptide epimers by ion mobility spectrometry. *Anal Chem* **2014,** *86* (6), 2972-81.

124. Meier, F.; Brunner, A.-D.; Frank, M.; Ha, A.; Bludau, I.; Voytik, E.; Kaspar-Schoenefeld, S.; Lubeck, M.; Raether, O.; Bache, N.; Aebersold, R.; Collins, B. C.; Röst, H. L.; Mann, M., diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nature Methods* **2020,** *17* (12), 1229-1236.

125. Gibson, K.; Cooper-Shepherd, D. A.; Pallister, E.; Inman, S. E.; Jackson, S. E.; Lindo, V., Toward Rapid Aspartic Acid Isomer Localization in Therapeutic Peptides Using Cyclic Ion Mobility Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2022**.

126. Tomczyk, N.; Giles, K.; Richardson, K.; Ujma, J.; Palmer, M.; Nielsen, P. K.; Haselmann, K. F., Mapping Isomeric Peptides Derived from Biopharmaceuticals Using High-Resolution Ion Mobility Mass Spectrometry. *Anal Chem* **2021,** *93* (49), 16379-16384.

127. Liu, Y.; Liu, Y.; Nytka, M.; Huang, S. R.; Lemr, K.; Turecek, F., Probing d- and l-Adrenaline Binding to beta2-Adrenoreceptor Peptide Motifs by Gas-Phase Photodissociation Cross-Linking and Ion Mobility Mass Spectrometry. *J Am Soc Mass Spectrom* **2021,** *32* (4), 1041-1052.

128. Desligniere, E.; Botzanowski, T.; Diemer, H.; Cooper-Shepherd, D. A.; Wagner-Rousset, E.; Colas, O.; Bechade, G.; Giles, K.; Hernandez-Alba, O.; Beck, A.; Cianferani, S., High-Resolution IMS-MS to Assign Additional Disulfide Bridge Pairing in Complementarity-Determining Regions of an IgG4 Monoclonal Antibody. *J Am Soc Mass Spectrom* **2021,** *32* (10), 2505-2512.

129. Eldrid, C.; Ujma, J.; Kalfas, S.; Tomczyk, N.; Giles, K.; Morris, M.; Thalassinos, K., Gas Phase Stability of Protein Ions in a Cyclic Ion Mobility Spectrometry Traveling Wave Device. *Anal Chem* **2019,** *91* (12), 7554-7561.

130. Eldrid, C.; Ben-Younis, A.; Ujma, J.; Britt, H.; Cragnolini, T.; Kalfas, S.; Cooper-Shepherd, D.; Tomczyk, N.; Giles, K.; Morris, M.; Akter, R.; Raleigh, D.; Thalassinos, K., Cyclic Ion Mobility-Collision Activation Experiments Elucidate Protein Behavior in the Gas Phase. *J Am Soc Mass Spectrom* **2021,** *32* (6), 1545-1552.

131. Deslignière, E. n.; Ollivier, S.; Ehkirch, A.; Martelet, A.; Ropartz, D.; Lechat, N.; Hernandez-Alba, O.; Menet, J.-M.; Clavier, S.; Rogniaux, H., Combination of IM-Based Approaches to Unravel the Coexistence of Two Conformers on a Therapeutic Multispecific mAb. *Analytical Chemistry* **2022**.

132. Shaw, J. B.; Cooper-Shepherd, D. A.; Hewitt, D.; Wildgoose, J. L.; Beckman, J. S.; Langridge, J. I.; Voinov, V. G., Enhanced Top-Down Protein Characterization with Electron Capture Dissociation and Cyclic Ion Mobility Spectrometry. *Anal Chem* **2022,** *94* (9), 3888-3896.

133. Snyder, D. T.; Jones, B. J.; Lin, Y. F.; Cooper-Shepherd, D. A.; Hewitt, D.; Wildgoose, J.; Brown, J. M.; Langridge, J. I.; Wysocki, V. H., Surface-induced dissociation of protein complexes on a cyclic ion mobility spectrometer. *Analyst* **2021,** *146* (22), 6861-6873.

134. Snyder, D. T.; Panczyk, E. M.; Somogyi, A.; Kaplan, D. A.; Wysocki, V., Simple and Minimally Invasive SID Devices for Native Mass Spectrometry. *Anal Chem* **2020,** *92* (16), 11195-11203.

135. Park, M. A. Apparatus and method for parallel flow ion mobility spectrometry combined with mass spectrometry. US7838826B1, 2010/11/23, 2010.

136. Fernandez-Lima, F.; Kaplan, D. A.; Suetering, J.; Park, M. A., Gas-phase separation using a trapped ion mobility spectrometer. *International Journal for Ion Mobility Spectrometry* **2011,** *14* (2), 93-98.

137. Meier, F.; Brunner, A.-D.; Koch, S.; Koch, H.; Lubeck, M.; Krause, M.; Goedecke, N.; Decker, J.; Kosinski, T.; Park, M. A.; Bache, N.; Hoerning, O.; Cox, J.; Räther, O.; Mann, M., Online parallel accumulation – serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Molecular & Cellular Proteomics* **2018**, mcp.TIR118.000900.

138. Fernandez-Lima, F. A.; Kaplan, D. A.; Park, M. A., Note: Integration of trapped ion mobility spectrometry with mass spectrometry. *Review of Scientific Instruments* **2011,** *82* (12), 126106.

139. Spraggins, J. M.; Djambazova, K. V.; Rivera, E. S.; Migas, L. G.; Neumann, E. K.; Fuetterer, A.; Suetering, J.; Goedecke, N.; Ly, A.; Van de Plas, R.; Caprioli, R. M., High-Performance Molecular Imaging with MALDI Trapped Ion-Mobility Time-of-Flight (timsTOF) Mass Spectrometry. *Analytical Chemistry* **2019,** *91* (22), 14552-14560.

140. Soltwisch, J.; Heijs, B.; Koch, A.; Vens-Cappell, S.; Höhndorf, J.; Dreisewerd, K., MALDI-2 on a Trapped Ion Mobility Quadrupole Time-of-Flight Instrument for Rapid Mass Spectrometry Imaging and Ion Mobility Separation of Complex Lipid Profiles. *Analytical Chemistry* **2020,** *92* (13), 8697-8703.

141. Meier, F.; Beck, S.; Grassl, N.; Lubeck, M.; Park, M. A.; Raether, O.; Mann, M., Parallel Accumulation–Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device. *Journal of Proteome Research* **2015,** *14* (12), 5378-5387.

142. Ridgeway, M. E.; Lubeck, M.; Jordens, J.; Mann, M.; Park, M. A., Trapped ion mobility spectrometry: A short review. *International Journal of Mass Spectrometry* **2018,** *425*, 22-35.

143. Michelmann, K.; Silveira, J. A.; Ridgeway, M. E.; Park, M. A., Fundamentals of Trapped Ion Mobility Spectrometry. *Journal of The American Society for Mass Spectrometry* **2015,** *26* (1), 14-24.

144. Benigni, P.; Porter, J.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F., Increasing Analytical Separation and Duty Cycle with Nonlinear Analytical Mobility Scan Functions in TIMS-FT-ICR MS. *Analytical Chemistry* **2018,** *90* (4), 2446-2450.

145. Silveira, J. A.; Ridgeway, M. E.; Laukien, F. H.; Mann, M.; Park, M. A., Parallel accumulation for 100% duty cycle trapped ion mobility-mass spectrometry. *International Journal of Mass Spectrometry* **2017,** *413*, 168-175.

146. Chai, M.; Young, M. N.; Liu, F. C.; Bleiholder, C., A Transferable, Sample-Independent Calibration Procedure for Trapped Ion Mobility Spectrometry (TIMS). *Analytical Chemistry* **2018,** *90* (15), 9040-9047.

147. Schroeder, M.; Meyer, S. W.; Heyman, H. M.; Barsch, A.; Sumner, L. W., Generation of a Collision Cross Section Library for Multi-Dimensional Plant Metabolomics Using UHPLC-Trapped Ion Mobility-MS/MS. *Metabolites* **2020,** *10* (1), 13.

148. Naylor, C. N.; Reinecke, T.; Ridgeway, M. E.; Park, M. A.; Clowers, B. H., Validation of Calibration Parameters for Trapped Ion Mobility Spectrometry. *Journal of the American Society for Mass Spectrometry* **2019,** *30* (10), 2152-2162.

149. Will, J. M.; Behrens, A.; Macke, M.; Quarles, C. D.; Karst, U., Automated Chiral Analysis of Amino Acids Based on Chiral Derivatization and Trapped Ion Mobility–Mass Spectrometry. *Analytical Chemistry* **2021,** *93* (2), 878-885.

150. Gomez, J. D.; Ridgeway, M. E.; Park, M. A.; Fritz, K. S., Utilizing ion mobility to identify isobaric post-translational modifications: resolving acrolein and propionyl lysine adducts by TIMS mass spectrometry. *International Journal for Ion Mobility Spectrometry* **2018,** *21* (3), 65-69.

151. Kruse, T.; Benz, C.; Garvanska, D. H.; Lindqvist, R.; Mihalic, F.; Coscia, F.; Inturi, R.; Sayadi, A.; Simonetti, L.; Nilsson, E.; Ali, M.; Kliche, J.; Moliner Morro, A.; Mund, A.; Andersson, E.; McInerney, G.; Mann, M.; Jemth, P.; Davey, N. E.; Överby, A. K.; Nilsson, J.; Ivarsson, Y., Large scale discovery of coronavirus-host factor protein interaction motifs reveals SARS-CoV-2 specific mechanisms and vulnerabilities. *Nature Communications* **2021,** *12* (1), 6761.

152. Larson, E. J.; Roberts, D. S.; Melby, J. A.; Buck, K. M.; Zhu, Y.; Zhou, S.; Han, L.; Zhang, Q.; Ge, Y., High-Throughput Multi-attribute Analysis of Antibody-Drug Conjugates Enabled by Trapped Ion Mobility Spectrometry and Top-Down Mass Spectrometry. *Analytical Chemistry* **2021,** *93* (29), 10013-10021.

153. Ihling, C. H.; Piersimoni, L.; Kipping, M.; Sinz, A., Cross-Linking/Mass Spectrometry Combined with Ion Mobility on a timsTOF Pro Instrument for Structural Proteomics. *Analytical Chemistry* **2021,** *93* (33), 11442-11450.

154. Steigenberger, B.; Toorn, v. d. H. W. P.; Bijl, E.; Greisch, J.-F.; Räther, O.; Lubeck, M.; Pieters, R. J.; Heck, A. J. R.; Scheltema, R. A., Benefits of Collisional Cross Section Assisted Precursor Selection (caps-PASEF) for Cross-linking Mass Spectrometry. *Molecular & Cellular Proteomics* **2020,** *19* (10), 1677-1687.

155. Loginov, D. S.; Fiala, J.; Brechlin, P.; Kruppa, G.; Novak, P., Hydroxyl radical footprinting analysis of a human haptoglobin-hemoglobin complex. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2022,** *1870* (2), 140735.

156. Loginov, D. S.; Fiala, J.; Chmelik, J.; Brechlin, P.; Kruppa, G.; Novak, P., Benefits of Ion Mobility Separation and Parallel Accumulation–Serial Fragmentation Technology on timsTOF Pro for the Needs of Fast Photochemical Oxidation of Protein Analysis. *ACS Omega* **2021,** *6* (15), 10352-10361.

157. Schmitz, T.; Pengelley, S.; Belau, E.; Suckau, D.; Imhof, D., LC-Trapped Ion Mobility Spectrometry-TOF MS Differentiation of 2- and 3-Disulfide-Bonded Isomers of the μ-Conotoxin PIIIA. *Analytical Chemistry* **2020,** *92* (16), 10920-10924.

158. Adams, K. J.; Ramirez, C. E.; Smith, N. F.; Muñoz-Muñoz, A. C.; Andrade, L.; Fernandez-Lima, F., Analysis of isomeric opioids in urine using LC-TIMS-TOF MS. *Talanta* **2018,** *183*, 177-183.

159. Vasilopoulou, C. G.; Sulek, K.; Brunner, A.-D.; Meitei, N. S.; Schweiger-Hufnagel, U.; Meyer, S. W.; Barsch, A.; Mann, M.; Meier, F., Trapped ion mobility spectrometry and PASEF enable in-depth lipidomics from minimal sample amounts. *Nature Communications* **2020,** *11* (1), 331.

160. Helmer, P. O.; Nordhorn, I. D.; Korf, A.; Behrens, A.; Buchholz, R.; Zubeil, F.; Karst, U.; Hayen, H., Complementing Matrix-Assisted Laser Desorption Ionization-Mass Spectrometry Imaging with Chromatography Data for Improved Assignment of Isobaric and Isomeric Phospholipids Utilizing Trapped Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2021,** *93* (4), 2135-2143.

161. Djambazova, K. V.; Klein, D. R.; Migas, L. G.; Neumann, E. K.; Rivera, E. S.; Van de Plas, R.; Caprioli, R. M.; Spraggins, J. M., Resolving the Complexity of Spatial Lipidomics Using MALDI TIMS Imaging Mass Spectrometry. *Analytical Chemistry* **2020,** *92* (19), 13290-13297.

162. Charkow, J.; Röst, H. L., Trapped Ion Mobility Spectrometry Reduces Spectral Complexity in Mass Spectrometry-Based Proteomics. *Analytical Chemistry* **2021,** *93* (50), 16751-16758.

163. Ogata, K.; Ishihama, Y., Extending the Separation Space with Trapped Ion Mobility Spectrometry Improves the Accuracy of Isobaric Tag-Based Quantitation in Proteomic LC/MS/MS. *Analytical Chemistry* **2020,** *92* (12), 8037-8040.

164. Almeida, F. A.; Passamani, L. Z.; Santa-Catarina, C.; Mooney, B. P.; Thelen, J. J.; Silveira, V., Label-Free Quantitative Phosphoproteomics Reveals Signaling Dynamics Involved in Embryogenic Competence Acquisition in Sugarcane. *Journal of Proteome Research* **2020,** *19* (10), 4145-4157.

165. Lesur, A.; Schmit, P.-O.; Bernardin, F.; Letellier, E.; Brehmer, S.; Decker, J.; Dittmar, G., Highly Multiplexed Targeted Proteomics Acquisition on a TIMS-QTOF. *Analytical Chemistry* **2021,** *93* (3), 1383-1392.

166. Brzhozovskiy, A.; Kononikhin, A.; Bugrova, A. E.; Kovalev, G. I.; Schmit, P.-O.; Kruppa, G.; Nikolaev, E. N.; Borchers, C. H., The Parallel Reaction Monitoring-Parallel Accumulation–Serial Fragmentation (prm-PASEF) Approach for Multiplexed Absolute Quantitation of Proteins in Human Plasma. *Analytical Chemistry* **2022,** *94* (4), 2016-2022.

167. Li, W.; Zhang, Q.; Wang, X.; Wang, H.; Zuo, W.; Xie, H.; Tang, J.; Wang, M.; Zeng, Z.; Cai, W.; Tang, D.; Dai, Y., Comparative Proteomic Analysis to Investigate the Pathogenesis of Oral Adenoid Cystic Carcinoma. *ACS Omega* **2021,** *6* (29), 18623-18634.

168. Lesur, A.; Dittmar, G., The clinical potential of prm-PASEF mass spectrometry. *Expert Review of Proteomics* **2021,** *18* (2), 75-82.

169. Shin, A.; Vazmitsel, Y.; Connolly, S.; Kabytaev, K., Comprehensive profiling and kinetic studies of glycated lysine residues in human serum albumin. *Analytical and Bioanalytical Chemistry* **2022,** *414* (17), 4861-4875.

170. Zhu, H.; Ficarro, S. B.; Alexander, W. M.; Fleming, L. E.; Adelmant, G.; Zhang, T.; Willetts, M.; Decker, J.; Brehmer, S.; Krause, M.; East, M. P.; Gray, N. S.; Johnson, G. L.; Kruppa, G.; Marto, J. A., PRM-LIVE with Trapped Ion Mobility Spectrometry and Its Application in Selectivity Profiling of Kinase Inhibitors. *Analytical Chemistry* **2021,** *93* (41), 13791-13799.

171. Garabedian, A.; Benigni, P.; Ramirez, C. E.; Baker, E. S.; Liu, T.; Smith, R. D.; Fernandez-Lima, F., Towards Discovery and Targeted Peptide Biomarker Detection Using nanoESI-TIMS-TOF MS. *Journal of the American Society for Mass Spectrometry* **2018,** *29* (5), 817-826.

172. Aballo, T. J.; Roberts, D. S.; Melby, J. A.; Buck, K. M.; Brown, K. A.; Ge, Y., Ultrafast and Reproducible Proteomics from Small Amounts of Heart Tissue Enabled by Azo and timsTOF Pro. *Journal of Proteome Research* **2021,** *20* (8), 4203-4211.

173. Fossati, A.; Richards, A. L.; Chen, K.-H.; Jaganath, D.; Cattamanchi, A.; Ernst, J. D.; Swaney, D. L., Toward Comprehensive Plasma Proteomics by Orthogonal Protease Digestion. *Journal of Proteome Research* **2021,** *20* (8), 4031-4040.

174. Qin, G.; Zhang, P.; Sun, M.; Fu, W.; Cai, C., Comprehensive spectral libraries for various rabbit eye tissue proteomes. *Sci Data* **2022,** *9* (1), 111.

175. Jeanne Dit Fouque, K.; Kaplan, D.; Voinov, V. G.; Holck, F. H. V.; Jensen, O. N.; Fernandez-Lima, F., Proteoform Differentiation using Tandem Trapped Ion Mobility, Electron Capture Dissociation, and ToF Mass Spectrometry. *Analytical Chemistry* **2021,** *93* (27), 9575-9582.

176. Borotto, N. B.; Graham, K. A., Fragmentation and Mobility Separation of Peptide and Protein Ions in a Trapped-Ion Mobility Device. *Analytical Chemistry* **2021,** *93* (29), 9959-9964.

177. Petelski, A. A.; Emmott, E.; Leduc, A.; Huffman, R. G.; Specht, H.; Perlman, D. H.; Slavov, N., Multiplexed single-cell proteomics using SCoPE2. *Nature Protocols* **2021,** *16* (12), 5398-5425.

178. Orsburn, B. C., Evaluation of the Sensitivity of Proteomics Methods Using the Absolute Copy Number of Proteins in a Single Cell as a Metric. *Proteomes* **2021,** *9* (3), 34.

179. Brunner, A.-D.; Thielert, M.; Vasilopoulou, C.; Ammar, C.; Coscia, F.; Mund, A.; Hoerning, O. B.; Bache, N.; Apalategui, A.; Lubeck, M.; Richter, S.; Fischer, D. S.; Raether, O.; Park, M. A.; Meier, F.; Theis, F. J.; Mann, M., Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Molecular Systems Biology* **2022,** *18* (3), e10798.

180. Yuan, Y.; Orsburn, B.; Bumpus, N., Single-Cell Proteomic Profiling of Hepatocytes Reveals Heterogeneous Responses to Hepatotoxic Drugs Nevirapine, Efavirenz, and Acetaminophen. *The FASEB Journal* **2022,** *36* (S1).

181. Mast, D. H.; Liao, H.-W.; Romanova, E. V.; Sweedler, J. V., Analysis of Peptide Stereochemistry in Single Cells by Capillary Electrophoresis–Trapped Ion Mobility Spectrometry Mass Spectrometry. *Analytical Chemistry* **2021,** *93* (15), 6205-6213.

182. Nagaraj, N.; Kosinski, T.; Apalategui, A.; Schmit, P.-O.; Lubeck, M.; Raether, O.; Kruppa, G., Ultrahigh Sensitivity Proteomics on the timsTOF SCP. *LC-GC North America* **2021,** *39* (9), 450-452.

183. Castro, D. C.; Xie, Y. R.; Rubakhin, S. S.; Romanova, E. V.; Sweedler, J. V., Image-guided MALDI mass spectrometry for high-throughput single-organelle characterization. *Nature Methods* **2021,** *18* (10), 1233-1238.

184. Adams, K. J.; Smith, N. F.; Ramirez, C. E.; Fernandez-Lima, F., Discovery and targeted monitoring of polychlorinated biphenyl metabolites in blood plasma using LC-TIMS-TOF MS. *International Journal of Mass Spectrometry* **2018,** *427*, 133-140.

185. Di Poto, C.; Tian, X.; Peng, X.; Heyman, H. M.; Szesny, M.; Hess, S.; Cazares, L. H., Metabolomic Profiling of Human Urine Samples Using LC-TIMS-QTOF Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2021,** *32* (8), 2072-2080.

186. Jeanne Dit Fouque, K.; Ramirez, C. E.; Lewis, R. L.; Koelmel, J. P.; Garrett, T. J.; Yost, R. A.; Fernandez-Lima, F., Effective Liquid Chromatography–Trapped Ion Mobility Spectrometry–Mass Spectrometry Separation of Isomeric Lipid Species. *Analytical Chemistry* **2019,** *91* (8), 5021-5027.

187. Drakopoulou, S. K.; Damalas, D. E.; Baessmann, C.; Thomaidis, N. S., Trapped Ion Mobility Incorporated in LC–HRMS Workflows as an Integral Analytical Platform of High

Sensitivity: Targeted and Untargeted 4D-Metabolomics in Extra Virgin Olive Oil. *Journal of Agricultural and Food Chemistry* **2021,** *69* (51), 15728-15737.

188. Horvath, T. D.; Dagan, S.; Scaraffia, P. Y., Unraveling mosquito metabolism with mass spectrometry-based metabolomics. *Trends in Parasitology* **2021,** *37* (8), 747-761.

189. SHARAF, B. M.; GIDDEY, A. D.; ALNISS, H.; AL-HROUB, H. M.; EL-AWADY, R.; MOUSA, M.; ALMEHDI, A.; SOARES, N. C.; SEMREEN, M. H., Untargeted Metabolomics of Breast Cancer Cells MCF-7 and SkBr3 Treated With Tamoxifen/Trastuzumab. *Cancer Genomics - Proteomics* **2022,** *19* (1), 79-93.

190. Neumann, E. K.; Migas, L. G.; Allen, J. L.; Caprioli, R. M.; Van de Plas, R.; Spraggins, J. M., Spatial Metabolomics of the Human Kidney using MALDI Trapped Ion Mobility Imaging Mass Spectrometry. *Analytical Chemistry* **2020,** *92* (19), 13084-13091.

191. Zhang, W.; Tang, D.; Lin, L.; Fan, T.; Xia, L.; Cai, W.; Dai, W.; Zou, C.; Yin, L.; Xu, Y.; Dai, Y., Integrative multiplatform-based molecular profiling of human colorectal cancer reveals proteogenomic alterations underlying mitochondrial inactivation. *Am J Cancer Res* **2021,** *11* (6), 2893-2910.

192. Zhang, W.; Lin, L.; Xia, L.; Cai, W.; Dai, W.; Zou, C.; Yin, L.; Tang, D.; Xu, Y.; Dai, Y., Multi-omics analyses of human colorectal cancer revealed three mitochondrial genes potentially associated with poor outcomes of patients. *Journal of Translational Medicine* **2021,** *19* (1), 273.

193. Yang, H.; Smith, R. D.; Chandler, C. E.; Johnson, J. K.; Jackson, S. N.; Woods, A. S.; Scott, A. J.; Goodlett, D. R.; Ernst, R. K., Lipid A Structural Determination from a Single Colony. *Analytical Chemistry* **2022,** *94* (21), 7460-7465.

194. Prianichnikov, N.; Koch, H.; Koch, S.; Lubeck, M.; Heilig, R.; Brehmer, S.; Fischer, R.; Cox, J., MaxQuant software for ion mobility enhanced shotgun proteomics. *bioRxiv* **2019**, 651760.

195. Sinitcyn, P.; Hamzeiy, H.; Salinas Soto, F.; Itzhak, D.; McCarthy, F.; Wichmann, C.; Steger, M.; Ohmayer, U.; Distler, U.; Kaspar-Schoenefeld, S.; Prianichnikov, N.; Yılmaz, Ş.; Rudolph, J. D.; Tenzer, S.; Perez-Riverol, Y.; Nagaraj, N.; Humphrey, S. J.; Cox, J., MaxDIA enables library-based and library-free data-independent acquisition proteomics. *Nature Biotechnology* **2021,** *39* (12), 1563-1573.

196. Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; Aebersold, R., OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nature Biotechnology* **2014,** *32* (3), 219-223.

197. Yılmaz, Ş.; Busch, F.; Nagaraj, N.; Cox, J., Accurate and Automated High-Coverage Identification of Chemically Cross-Linked Peptides with MaxLynx. *Analytical Chemistry* **2022,** *94* (3), 1608-1617.

198. Yu, F.; Haynes, S. E.; Teo, G. C.; Avtonomov, D. M.; Polasky, D. A.; Nesvizhskii, A. I., Fast Quantitative Analysis of timsTOF PASEF Data with MSFragger and IonQuant. *Molecular & Cellular Proteomics* **2020,** *19* (9), 1575-1585.

199. Wen, C.; Gan, G.; Xu, X.; Lin, G.; Chen, X.; Wu, Y.; Xu, Z.; Wang, J.; Xie, C.; Wang, H.-R.; Zhong, C.-Q., Investigation of Effects of the Spectral Library on Analysis of diaPASEF Data. *Journal of Proteome Research* **2022,** *21* (2), 507-518.

200. Peeters, M. K. R.; Baggerman, G.; Gabriels, R.; Pepermans, E.; Menschaert, G.; Boonen, K., Ion Mobility Coupled to a Time-of-Flight Mass Analyzer Combined With Fragment Intensity Predictions Improves Identification of Classical Bioactive Peptides and Small Open Reading Frame-Encoded Peptides. *Frontiers in Cell and Developmental Biology* **2021,** *9*.

201. Meier, F.; Köhler, N. D.; Brunner, A.-D.; Wanka, J.-M. H.; Voytik, E.; Strauss, M. T.; Theis, F. J.; Mann, M., Deep learning the collisional cross sections of the peptide universe from a million experimental values. *Nature Communications* **2021,** *12* (1), 1185.

202. Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M., DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods* **2020,** *17* (1), 41-44.

203. Willems, S.; Voytik, E.; Skowronek, P.; Strauss, M. T.; Mann, M., AlphaTims: Indexing Trapped Ion Mobility Spectrometry–TOF Data for Fast and Easy Accession and Visualization. *Molecular & Cellular Proteomics* **2021,** *20*.

204. Łącki, M. K.; Startek, M. P.; Brehmer, S.; Distler, U.; Tenzer, S., OpenTIMS, TimsPy, and TimsR: Open and Easy Access to timsTOF Raw Data. *Journal of Proteome Research* **2021,** *20* (4), 2122-2129.

205. 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. *Nature Methods* **2017,** *14* (5), 513-520.

206. Morsa, D.; Hanozin, E.; Eppe, G.; Quinton, L.; Gabelica, V.; Pauw, E. D., Effective Temperature and Structural Rearrangement in Trapped Ion Mobility Spectrometry. *Analytical Chemistry* **2020,** *92* (6), 4573-4582.

207. Zhang, L.; Wang, Y.; Zheng, F.; Zhu, D.; Liang, Y.; Shi, Q., Influence Exerted by the Solvent Effect on the Mobility Peak of 1,8-Naphthalic Anhydride in Ion Mobility Spectrometry. *Journal of the American Society for Mass Spectrometry* **2022,** *33* (3), 457-462.

208. C. Liu, F.; E. Ridgeway, M.; A. Park, M.; Bleiholder, C., Tandem trapped ion mobility spectrometry. *Analyst* **2018,** *143* (10), 2249-2258.

209. Liu, F. C.; Cropley, T. C.; Ridgeway, M. E.; Park, M. A.; Bleiholder, C., Structural Analysis of the Glycoprotein Complex Avidin by Tandem-Trapped Ion Mobility Spectrometry–Mass Spectrometry (Tandem-TIMS/MS). *Analytical Chemistry* **2020,** *92* (6), 4459-4467.

210. Kirk, S. R.; Liu, F. C.; Cropley, T. C.; Carlock, H. R.; Bleiholder, C., On the Preservation of Non-covalent Peptide Assemblies in a Tandem-Trapped Ion Mobility Spectrometer-Mass Spectrometer (TIMS-TIMS-MS). *Journal of the American Society for Mass Spectrometry* **2019,** *30* (7), 1204-1212.

211. Winter, D. L.; Wilkins, M. R.; Donald, W. A., Differential Ion Mobility–Mass Spectrometry for Detailed Analysis of the Proteome. *Trends in Biotechnology* **2019,** *37* (2), 198-213.

212. Guevremont, R., High-field asymmetric waveform ion mobility spectrometry: A new tool for mass spectrometry. *Journal of Chromatography A* **2004,** *1058* (1-2), 3-19.

213. Shvartsburg, A. A., *Differential ion mobility spectrometry : nonlinear ion transport and fundamentals of FAIMS*. CRC Press: Boca Raton, 2009; p xxix, 299 p.

214. Purves, R. W.; Prasad, S.; Belford, M.; Vandenberg, A.; Dunyach, J.-J., Optimization of a New Aerodynamic Cylindrical FAIMS Device for Small Molecule Analysis. *Journal of the American Society for Mass Spectrometry* **2017,** *28* (3), 525-538.

215. Li, J.; Gao, W.; Wu, H.; Shi, S.; Yu, J.; Tang, K., On the resolution, sensitivity and ion transmission efficiency of a planar FAIMS. *International Journal of Mass Spectrometry* **2022,** *471*, 116727.

216. Shvartsburg, A. A.; Smith, R. D.; Wilks, A.; Koehl, A.; Ruiz-Alonso, D.; Boyle, B., Ultrafast Differential Ion Mobility Spectrometry at Extreme Electric Fields in Multichannel Microchips. *Analytical Chemistry* **2009,** *81* (15), 6489-6495.

217. Shvartsburg, A. A.; Smith, R. D., Ultrahigh-Resolution Differential Ion Mobility Spectrometry Using Extended Separation Times. *Analytical Chemistry* **2011,** *83* (1), 23-29.

218. Li, J.; Li, L.; Gao, W.; Shi, S.; Yu, J.; Tang, K., Two-Dimensional FAIMS-IMS Characterization of Peptide Conformers with Resolution Exceeding 1000. *Analytical Chemistry* **2022,** *94* (16), 6363-6370.

219. Kabir, K. M. M.; Donald, W. A., Microscale differential ion mobility spectrometry for field deployable chemical analysis. *TrAC Trends in Analytical Chemistry* **2017,** *97*, 399-427.

220. Baird, M. A.; Shliaha, P. V.; Anderson, G. A.; Moskovets, E.; Laiko, V.; Makarov, A. A.; Jensen, O. N.; Shvartsburg, A. A., High-Resolution Differential Ion Mobility Separations/Orbitrap Mass Spectrometry without Buffer Gas Limitations. *Analytical Chemistry* **2019,** *91* (10), 6918-6925.

221. Shvartsburg, A. A.; Haris, A.; Andrzejewski, R.; Entwistle, A.; Giles, R., Differential Ion Mobility Separations in the Low-Pressure Regime. *Analytical Chemistry* **2018,** *90* (1), 936-943.

222. Barnett, D. A.; Belford, M.; Dunyach, J.-J.; Purves, R. W., Characterization of a temperature-Controlled FAIMS system. *Journal of the American Society for Mass Spectrometry* **2007,** *18* (9), 1653-1663.

223. Swearingen, K. E.; Winget, J. M.; Hoopmann, M. R.; Kusebauch, U.; Moritz, R. L., Decreased Gap Width in a Cylindrical High-Field Asymmetric Waveform Ion Mobility Spectrometry Device Improves Protein Discovery. *Analytical Chemistry* **2015,** *87* (24), 12230-12237.

224. Pathak, P.; Baird, M. A.; Shvartsburg, A. A., Structurally Informative Isotopic Shifts in Ion Mobility Spectra for Heavier Species. *Journal of the American Society for Mass Spectrometry* **2019**.

225. Creese, A. J.; Cooper, H. J., Separation of cis and trans Isomers of Polyproline by FAIMS Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2016,** *27* (12), 2071-2074.

226. Shvartsburg, A. A.; Isaac, G.; Leveque, N.; Smith, R. D.; Metz, T. O., Separation and Classification of Lipids Using Differential Ion Mobility Spectrometry. *Journal of the American Society for Mass Spectrometry* **2011,** *22* (7), 1146-1155.

227. Berthias, F.; Poad, B. L. J.; Thurman, H. A.; Bowman, A. P.; Blanksby, S. J.; Shvartsburg, A. A., Disentangling Lipid Isomers by High-Resolution Differential Ion Mobility Spectrometry/Ozone-Induced Dissociation of Metalated Species. *Journal of the American Society for Mass Spectrometry* **2021,** *32* (12), 2827-2836.

228. Xuan, Y.; Creese, A. J.; Horner, J. A.; Cooper, H. J., High-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled with high-resolution electron transfer dissociation mass spectrometry for the analysis of isobaric phosphopeptides. *Rapid Communications in Mass Spectrometry* **2009,** *23* (13), 1963-1969.

229. Shvartsburg, A. A.; Creese, A. J.; Smith, R. D.; Cooper, H. J., Separation of Peptide Isomers with Variant Modified Sites by High-Resolution Differential Ion Mobility Spectrometry. *Analytical Chemistry* **2010,** *82* (19), 8327-8334.

230. Shvartsburg, A. A.; Singer, D.; Smith, R. D.; Hoffmann, R., Ion Mobility Separation of Isomeric Phosphopeptides from a Protein with Variant Modification of Adjacent Residues. *Analytical Chemistry* **2011,** *83* (13), 5078-5085.

231. Ibrahim, Y. M.; Shvartsburg, A. A.; Smith, R. D.; Belov, M. E., Ultrasensitive Identification of Localization Variants of Modified Peptides Using Ion Mobility Spectrometry. *Analytical Chemistry* **2011,** *83* (14), 5617-5623.

232. Shvartsburg, A. A.; Creese, A. J.; Smith, R. D.; Cooper, H. J., Separation of a Set of Peptide Sequence Isomers Using Differential Ion Mobility Spectrometry. *Analytical chemistry* **2011,** *83* (18), 6918-6923.

233. Shvartsburg, A. A.; Zheng, Y.; Smith, R. D.; Kelleher, N. L., Separation of variant methylated histone tails by differential ion mobility. *Anal. Chem. (Washington, DC, U. S.)* **2012,** *84* (15), 6317-6320.

234. Garabedian, A.; Baird, M. A.; Porter, J.; Jeanne Dit Fouque, K.; Shliaha, P. V.; Jensen, O. N.; Williams, T. D.; Fernandez-Lima, F.; Shvartsburg, A. A., Linear and Differential Ion Mobility Separations of Middle-Down Proteoforms. *Anal Chem* **2018,** *90* (4), 2918-2925.

235. Shliaha, P. V.; Baird, M. A.; Nielsen, M. M.; Gorshkov, V.; Bowman, A. P.; Kaszycki, J. L.; Jensen, O. N.; Shvartsburg, A. A., Characterization of Complete Histone Tail Proteoforms Using Differential Ion Mobility Spectrometry. *Anal. Chem. (Washington, DC, U. S.)* **2017**, Ahead of Print.

236. Baird, M. A.; Anderson, G. A.; Shliaha, P. V.; Jensen, O. N.; Shvartsburg, A. A., Differential Ion Mobility Separations/Mass Spectrometry with High Resolution in Both Dimensions. *Analytical Chemistry* **2018**.

237. Creese, A. J.; Cooper, H. J., Separation and Identification of Isomeric Glycopeptides by High Field Asymmetric Waveform Ion Mobility Spectrometry. *Anal. Chem. (Washington, DC, U. S.)* **2012,** *84* (5), 2597-2601.

238. Pathak, P.; Baird, M. A.; Shvartsburg, A. A., High-Resolution Ion Mobility Separations of Isomeric Glycoforms with Variations on the Peptide and Glycan Levels. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (7), 1603-1609.

239. Shvartsburg, A. A.; Smith, R. D., Protein Analyses Using FAIMS Microchips with Mass Spectrometry. *Analytical chemistry* **2012,** *84* (17), 7297-7300.

240. Kolakowski, B. M.; Mester, Z., Review of applications of high-field asymmetric waveform ion mobility spectrometry (FAIMS) and differential mobility spectrometry (DMS). *Analyst* **2007,** *132* (9), 842-864.

241. Hebert, A. S.; Prasad, S.; Belford, M. W.; Bailey, D. J.; McAlister, G. C.; Abbatiello, S. E.; Huguet, R.; Wouters, E. R.; Dunyach, J.-J.; Brademan, D. R.; Westphall, M. S.; Coon, J. J., Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Analytical Chemistry* **2018,** *90* (15), 9529-9537.

242. Eckert, S.; Chang, Y.-C.; Bayer, F. P.; The, M.; Kuhn, P.-H.; Weichert, W.; Kuster, B., Evaluation of Disposable Trap Column nanoLC–FAIMS–MS/MS for the Proteomic Analysis of FFPE Tissue. *Journal of Proteome Research* **2021,** *20* (12), 5402-5411.

243. Ang, C.-S.; Sacharz, J.; Leeming, M. G.; Nie, S.; Varshney, S.; Scott, N. E.; Williamson, N. A., Getting more out of FLAG-Tag co-immunoprecipitation mass spectrometry experiments using FAIMS. *Journal of Proteomics* **2022,** *254*, 104473.

244. Greguš, M.; Kostas, J. C.; Ray, S.; Abbatiello, S. E.; Ivanov, A. R., Improved Sensitivity of Ultralow Flow LC–MS-Based Proteomic Profiling of Limited Samples Using Monolithic Capillary Columns and FAIMS Technology. *Analytical Chemistry* **2020,** *92* (21), 14702-14712.

245. Pfammatter, S.; Wu, Z.; Bonneil, E.; Bailey, D. J.; Prasad, S.; Belford, M.; Rochon, J.; Picard, P.; Lacoursière, J.; Dunyach, J.-J.; Thibault, P., Integration of Segmented Ion Fractionation and Differential Ion Mobility on a Q-Exactive Hybrid Quadrupole Orbitrap Mass Spectrometer. *Analytical Chemistry* **2021,** *93* (28), 9817-9825.

246. Yan, T.; Desai, H. S.; Boatner, L. M.; Yen, S. L.; Cao, J.; Palafox, M. F.; Jami-Alahmadi, Y.; Backus, K. M., SP3-FAIMS Chemoproteomics for High-Coverage Profiling of the Human Cysteinome**. *ChemBioChem* **2021,** *22* (10), 1841-1851.

247. Besada, V.; Ramos, Y.; Espinosa, L. A.; Fu, W.; Perera, Y.; González, L. J., FAIMS-MS might contribute to phosphopeptides identification in plasma. *Journal of Proteomics* **2021,** *234*, 104102.

248. Yang, Z.; Sun, L., Recent technical progress in sample preparation and liquid-phase separation-mass spectrometry for proteomic analysis of mass-limited samples. *Analytical Methods* **2021,** *13* (10), 1214-1225.

249. Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill Jr., H. H., Ion mobility–mass spectrometry. *Journal of Mass Spectrometry* **2008,** *43* (1), 1-22.

250. Reilly, L.; Peng, L.; Lara, E.; Ramos, D.; Fernandopulle, M.; Pantazis, C. B.; Stadler, J.; Santiana, M.; Dadu, A.; Iben, J.; Faghri, F.; Nalls, M. A.; Coon, S. L.; Narayan, P.; Singleton, A. B.; Cookson, M. R.; Ward, M. E.; Qi, Y. A., A fully automated FAIMS-DIA proteomic pipeline for high-throughput characterization of iPSC-derived neurons. *bioRxiv* **2021**, 2021.11.24.469921.

251. Bekker-Jensen, D. B.; Martínez-Val, A.; Steigerwald, S.; Rüther, P.; Fort, K. L.; Arrey, T. N.; Harder, A.; Makarov, A.; Olsen, J. V., A Compact Quadrupole-Orbitrap Mass

Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients *. *Molecular & Cellular Proteomics* **2020,** *19* (4), 716-729.

252. Ivanov, M. V.; Bubis, J. A.; Gorshkov, V.; Abdrakhimov, D. A.; Kjeldsen, F.; Gorshkov, M. V., Boosting MS1-only Proteomics with Machine Learning Allows 2000 Protein Identifications in Single-Shot Human Proteome Analysis Using 5 min HPLC Gradient. *Journal of Proteome Research* **2021,** *20* (4), 1864-1873.

253. Stejskal, K.; Op de Beeck, J.; Dürnberger, G.; Jacobs, P.; Mechtler, K., Ultrasensitive NanoLC-MS of Subnanogram Protein Samples Using Second Generation Micropillar Array LC Technology with Orbitrap Exploris 480 and FAIMS PRO. *Analytical Chemistry* **2021,** *93* (25), 8704-8710.

254. Cong, Y.; Motamedchaboki, K.; Misal, S. A.; Liang, Y.; Guise, A. J.; Truong, T.; Huguet, R.; Plowey, E. D.; Zhu, Y.; Lopez-Ferrer, D.; Kelly, R. T., Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell. *Chemical Science* **2021,** *12* (3), 1001-1006.

255. Kelly, R. T., Single-cell Proteomics: Progress and Prospects. *Molecular & Cellular Proteomics* **2020,** *19* (11), 1739-1748.

256. Furtwängler, B.; Üresin, N.; Motamedchaboki, K.; Huguet, R.; Lopez-Ferrer, D.; Zabrouskov, V.; Porse, B. T.; Schoof, E. M., Real-Time Search Assisted Acquisition on a Tribrid Mass Spectrometer Improves Coverage in Multiplexed Single-Cell Proteomics. *Molecular & Cellular Proteomics* **2022**, 100219.

257. Pfammatter, S.; Bonneil, E.; McManus, F. P.; Prasad, S.; Bailey, D. J.; Belford, M.; Dunyach, J.-J.; Thibault, P., A Novel Differential Ion Mobility Device Expands the Depth of Proteome Coverage and the Sensitivity of Multiplex Proteomic Measurements *. *Molecular & Cellular Proteomics* **2018,** *17* (10), 2051-2067.

258. Pfammatter, S.; Bonneil, E.; Thibault, P., Improvement of Quantitative Measurements in Multiplex Proteomics Using High-Field Asymmetric Waveform Spectrometry. *Journal of Proteome Research* **2016,** *15* (12), 4653-4665.

259. Schweppe, D. K.; Prasad, S.; Belford, M. W.; Navarrete-Perea, J.; Bailey, D. J.; Huguet, R.; Jedrychowski, M. P.; Rad, R.; McAlister, G.; Abbatiello, S. E.; Woulters, E. R.; Zabrouskov, V.; Dunyach, J.-J.; Paulo, J. A.; Gygi, S. P., Characterization and Optimization of Multiplexed Quantitative Analyses Using High-Field Asymmetric-Waveform Ion Mobility Mass Spectrometry. *Analytical Chemistry* **2019,** *91* (6), 4010-4016.

260. Bonneil, E.; Pfammatter, S.; Thibault, P., Enhancement of mass spectrometry performance for proteomic analyses using high-field asymmetric waveform ion mobility spectrometry (FAIMS). *Journal of Mass Spectrometry* **2015,** *50* (11), 1181-1195.

261. Fang, P.; Ji, Y.; Silbern, I.; Viner, R.; Oellerich, T.; Pan, K.-T.; Urlaub, H., Evaluation and Optimization of High-Field Asymmetric Waveform Ion-Mobility Spectrometry for Multiplexed Quantitative Site-Specific N-Glycoproteomics. *Analytical Chemistry* **2021,** *93* (25), 8846-8855.

262. Gaun, A.; Lewis Hardell, K. N.; Olsson, N.; O'Brien, J. J.; Gollapudi, S.; Smith, M.; McAlister, G.; Huguet, R.; Keyser, R.; Buffenstein, R.; McAllister, F. E., Automated 16-Plex Plasma Proteomics with Real-Time Search and Ion Mobility Mass Spectrometry Enables Large-Scale Profiling in Naked Mole-Rats and Mice. *Journal of Proteome Research* **2021,** *20* (2), 1280-1295.

263. Pfammatter, S.; Bonneil, E.; McManus, F. P.; Thibault, P., Accurate Quantitative Proteomic Analyses Using Metabolic Labeling and High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS). *Journal of Proteome Research* **2019,** *18* (5), 2129-2138.

264. Schweppe, D. K.; Rusin, S. F.; Gygi, S. P.; Paulo, J. A., Optimized Workflow for Multiplexed Phosphorylation Analysis of TMT-Labeled Peptides Using High-Field Asymmetric Waveform Ion Mobility Spectrometry. *Journal of Proteome Research* **2020,** *19* (1), 554-560.

265. Zhao, H.; Creese, A. J.; Cooper, H. J., Online LC-FAIMS-MS/MS for the Analysis of Phosphorylation in Proteins. *Methods Mol Biol* **2016,** *1355*, 241-50.

266. Muehlbauer, L. K.; Hebert, A. S.; Westphall, M. S.; Shishkova, E.; Coon, J. J., Global Phosphoproteome Analysis Using High-Field Asymmetric Waveform Ion Mobility Spectrometry on a Hybrid Orbitrap Mass Spectrometer. *Analytical Chemistry* **2020,** *92* (24), 15959-15967.

267. Ahmad Izaham, A. R.; Ang, C.-S.; Nie, S.; Bird, L. E.; Williamson, N. A.; Scott, N. E., What Are We Missing by Using Hydrophilic Enrichment? Improving Bacterial Glycoproteome Coverage Using Total Proteome and FAIMS Analyses. *Journal of Proteome Research* **2021,** *20* (1), 599-612.

268. Ulasi, G. N.; Creese, A. J.; Hui, S. X.; Penn, C. W.; Cooper, H. J., Comprehensive mapping of O-glycosylation in flagellin from Campylobacter jejuni 11168: A multienzyme differential ion mobility mass spectrometry approach. *Proteomics* **2015,** *15* (16), 2733-45.

269. Zhao, H.; Cunningham, D. L.; Creese, A. J.; Heath, J. K.; Cooper, H. J., FAIMS and Phosphoproteomics of Fibroblast Growth Factor Signaling: Enhanced Identification of Multiply Phosphorylated Peptides. *J. Proteome Res.* **2015,** *14* (12), 5077-5087.

270. Griffiths, R. L.; Simmonds, A. L.; Swales, J. G.; Goodwin, R. J. A.; Cooper, H. J., LESA MS Imaging of Heat-Preserved and Frozen Tissue: Benefits of Multistep Static FAIMS. *Analytical Chemistry* **2018,** *90* (22), 13306-13314.

271. Rosting, C.; Yu, J.; Cooper, H. J., High Field Asymmetric Waveform Ion Mobility Spectrometry in Nontargeted Bottom-up Proteomics of Dried Blood Spots. *Journal of Proteome Research* **2018,** *17* (6), 1997-2004.

272. Zhang, T.; Gygi, S. P.; Paulo, J. A., Temporal Proteomic Profiling of SH-SY5Y Differentiation with Retinoic Acid Using FAIMS and Real-Time Searching. *Journal of Proteome Research* **2021,** *20* (1), 704-714.

273. Johnson, R. O. B.; Greer, T.; Cejkov, M.; Zheng, X.; Li, N., Combination of FAIMS, Protein A Depletion, and Native Digest Conditions Enables Deep Proteomic Profiling of Host Cell Proteins in Monoclonal Antibodies. *Analytical Chemistry* **2020,** *92* (15), 10478-10484.

274. Pfammatter, S.; Bonneil, E.; McManus, F. P.; Thibault, P., Gas-Phase Enrichment of Multiply Charged Peptide Ions by Differential Ion Mobility Extend the Comprehensiveness of SUMO Proteome Analyses. *Journal of the American Society for Mass Spectrometry* **2018,** *29* (6), 1111-1124.

275. Schnirch, L.; Nadler-Holly, M.; Siao, S.-W.; Frese, C. K.; Viner, R.; Liu, F., Expanding the Depth and Sensitivity of Cross-Link Identification by Differential Ion Mobility Using High-Field Asymmetric Waveform Ion Mobility Spectrometry. *Analytical Chemistry* **2020,** *92* (15), 10495-10503.

276. Adoni, K. R.; Cunningham, D. L.; Heath, J. K.; Leney, A. C., FAIMS Enhances the Detection of PTM Crosstalk Sites. *Journal of Proteome Research* **2022,** *21* (4), 930-939.

277. Meyer, J. G.; Niemi, N. M.; Pagliarini, D. J.; Coon, J. J., Quantitative shotgun proteome analysis by direct infusion. *Nature Methods* **2020,** *17* (12), 1222-1228.

278. Trujillo, E. A.; Hebert, A. S.; Rivera Vazquez, J. C.; Brademan, D. R.; Tatli, M.; Amador-Noguez, D.; Meyer, J. G.; Coon, J. J., Rapid Targeted Quantitation of Protein Overexpression with Direct Infusion Shotgun Proteome Analysis (DISPA-PRM). *Analytical Chemistry* **2022,** *94* (4), 1965-1973.

279. Cooper, H. J., To What Extent is FAIMS Beneficial in the Analysis of Proteins? *Journal of the American Society for Mass Spectrometry* **2016,** *27*, 566-577.

280. Delvaux, A.; Rathahao-Paris, E.; Alves, S., Different ion mobility-mass spectrometry coupling techniques to promote metabolomics. *Mass Spectrometry Reviews n/a* (n/a).

281. Bennet, S. M.; Keshteli, A. H.; Bercik, P.; Madsen, K. L.; Reed, D.; Vanner, S. J., Application of metabolomics to the study of irritable bowel syndrome. *Neurogastroenterology & Motility* **2020,** *32* (6), e13884.

282. Shliaha, P. V.; Gorshkov, V.; Kovalchuk, S. I.; Schwämmle, V.; Baird, M. A.; Shvartsburg, A. A.; Jensen, O. N., Middle-Down Proteomic Analyses with Ion Mobility Separations of Endogenous Isomeric Proteoforms. *Analytical Chemistry* **2020**.

283. Fulcher, J. M.; Makaju, A.; Moore, R. J.; Zhou, M.; Bennett, D. A.; De Jager, P. L.; Qian, W.-J.; Paša-Tolić, L.; Petyuk, V. A., Enhancing Top-Down Proteomics of Brain Tissue with FAIMS. *Journal of Proteome Research* **2021,** *20* (5), 2780-2795.

284. Griffiths, R. L.; Hughes, J. W.; Abbatiello, S. E.; Belford, M. W.; Styles, I. B.; Cooper, H. J., Comprehensive LESA Mass Spectrometry Imaging of Intact Proteins by Integration of Cylindrical FAIMS. *Analytical Chemistry* **2020,** *92* (4), 2885-2890.

285. Gerbasi, V. R.; Melani, R. D.; Abbatiello, S. E.; Belford, M. W.; Huguet, R.; McGee, J. P.; Dayhoff, D.; Thomas, P. M.; Kelleher, N. L., Deeper Protein Identification Using Field Asymmetric Ion Mobility Spectrometry in Top-Down Proteomics. *Analytical Chemistry* **2021,** *93* (16), 6323-6328.

286. Kaulich, P. T.; Cassidy, L.; Winkels, K.; Tholey, A., Improved Identification of Proteoforms in Top-Down Proteomics Using FAIMS with Internal CV Stepping. *Analytical Chemistry* **2022,** *94* (8), 3600-3607.

287. Hale, O. J.; Illes-Toth, E.; Mize, T. H.; Cooper, H. J., High-Field Asymmetric Waveform Ion Mobility Spectrometry and Native Mass Spectrometry: Analysis of Intact Protein Assemblies and Protein Complexes. *Analytical Chemistry* **2020,** *92* (10), 6811-6816.

288. Du, X.; Mou, J.; Zeng, H.; Zeng, R.; Jiang, Y.; Li, H., Printed Circuit Board (PCB) Brazing and Ion Source Integration of a High-Field Asymmetric Ion Mobility Spectrometry (FAIMS) Chip. *Analytical Letters* **2021,** *54* (8), 1377-1388.

289. Szykuła, K. M.; Meurs, J.; Turner, M. A.; Creaser, C. S.; Reynolds, J. C., Combined hydrophilic interaction liquid chromatography-scanning field asymmetric waveform ion mobility spectrometry-time-of-flight mass spectrometry for untargeted metabolomics. *Analytical and Bioanalytical Chemistry* **2019,** *411* (24), 6309-6317.

290. Arthur, K. L.; Turner, M. A.; Reynolds, J. C.; Creaser, C. S., Increasing Peak Capacity in Nontargeted Omics Applications by Combining Full Scan Field Asymmetric Waveform Ion Mobility Spectrometry with Liquid Chromatography–Mass Spectrometry. *Analytical Chemistry* **2017,** *89* (6), 3452-3459.

291. Li, H.; Pan, J.; Zeng, H.; Chen, Z.; Du, X.; Xiao, W., Identification of Specific Substances in the FAIMS Spectra of Complex Mixtures Using Deep Learning. *Sensors* **2021,** *21* (18), 6160.

292. Jeffery, I. B.; Das, A.; O'Herlihy, E.; Coughlan, S.; Cisek, K.; Moore, M.; Bradley, F.; Carty, T.; Pradhan, M.; Dwibedi, C.; Shanahan, F.; O'Toole, P. W., Differences in Fecal

Microbiomes and Metabolomes of People With vs Without Irritable Bowel Syndrome and Bile Acid Malabsorption. *Gastroenterology* **2020,** *158* (4), 1016-1028.e8.

293. McFarlanE, M.; MozdiaK, E.; Daulton, E.; Arasaradnam, R.; Covington, J.; Nwokolo, C., Pre-analytical and analytical variables that influence urinary volatile organic compound measurements. *PLOS ONE* **2020,** *15* (7), e0236591.

294. Arasaradnam, R. P.; Westenbrink, E.; McFarlane, M. J.; Harbord, R.; Chambers, S.; O'Connell, N.; Bailey, C.; Nwokolo, C. U.; Bardhan, K. D.; Savage, R.; Covington, J. A., Differentiating Coeliac Disease from Irritable Bowel Syndrome by Urinary Volatile Organic Compound Analysis – A Pilot Study. *PLOS ONE* **2014,** *9* (10), e107312.

295. Niemi, R. J.; Roine, A. N.; Eräviita, E.; Kumpulainen, P. S.; Mäenpää, J. U.; Oksala, N., FAIMS analysis of urine gaseous headspace is capable of differentiating ovarian cancer. *Gynecologic Oncology* **2018,** *151* (3), 519-524.

296. NISSINEN, S. I.; ROINE, A.; HOKKINEN, L.; KARJALAINEN, M.; VENÄLÄINEN, M.; HELMINEN, H.; NIEMI, R.; LEHTIMÄKI, T.; RANTANEN, T.; OKSALA, N., Detection of Pancreatic Cancer by Urine Volatile Organic Compound Analysis. *Anticancer Research* **2019,** *39* (1), 73-79.

297. Esfahani, S.; Wicaksono, A.; Mozdiak, E.; Arasaradnam, R. P.; Covington, J. A., Non-Invasive Diagnosis of Diabetes by Volatile Organic Compounds in Urine Using FAIMS and Fox4000 Electronic Nose. *Biosensors* **2018,** *8* (4), 121.

298. Mozdiak, E.; Wicaksono, A. N.; Covington, J. A.; Arasaradnam, R. P., Colorectal cancer and adenoma screening using urinary volatile organic compound (VOC) detection: early results from a single-centre bowel screening population (UK BCSP). *Techniques in Coloproctology* **2019,** *23* (4), 343-351.

299. Bosch, S.; van Gaal, N.; Zuurbier, R. P.; Covington, J. A.; Wicaksono, A. N.; Biezeveld, M. H.; Benninga, M. A.; Mulder, C. J.; de Boer, N. K. H.; de Meij, T. G. J., Differentiation Between Pediatric Irritable Bowel Syndrome and Inflammatory Bowel Disease Based on Fecal Scent: Proof of Principle Study. *Inflammatory Bowel Diseases* **2018,** *24* (11), 2468-2475.

300. Eldrid, C.; Ujma, J.; Kalfas, S.; Tomczyk, N.; Giles, K.; Morris, M.; Thalassinos, K., Gas Phase Stability of Protein Ions in a Cyclic Ion Mobility Spectrometry Traveling Wave Device. *Analytical Chemistry* **2019,** *91* (12), 7554-7561.

Chapter 10: Conclusions and Future Directions

Mass spectrometry (MS)-based proteomic analyses are likely to be a critical component of biomolecular analyses for decades to come. The ample development and technological advanced validated in this field of study have provided diversity in approach, allowing users to tailor analytical workflows to a given application or underlying pursuit. Nevertheless, there are some components of analytical workflows that remain nearly ubiquitous across proteomic analyses, meaning that the observable proteome and the information derived within is beholden to and biased towards the efficiency of these singular components.

Principally, the use of reversed-phase liquid chromatography (RPLC) is the predominant mode of biomolecular separation. While the retention capacity, achievable resolution, and various technological improvements are noted and the performance is not in question, we do present evidence that reveals RPLC separations only sample a portion of the proteome. As shown, RPLC separations are dependent on the relative hydrophobicity of peptide analytes, biasing towards retention of longer, less polar species that ionize well and are easily identified via MS. However, the work in chapters 2 and 5 demonstrate there is a significant amount of biomolecular information left untouched or undiscovered when utilizing RPLC. Without any additional treatment or preparation of samples, the addition of PGC separations increased peptide and protein identifications by as much as 43% and 24%, respectively. Not only this, we demonstrate that the additional proteins identified are not any more or less abundant than those identified in RPLC, indicating this biomolecular information is easily within reach, but is often missed. The value of more comprehensive protein profiles goes without stating, but we also demonstrate how the increased peptide identifications obtained through PGC analyses can substantially impact protein-level abundance estimations. These realities known, we also illuminate how modern proteomics

data processing pipelines are less efficient at identifying peptides from PGC separations, due largely to the inability to correctly predict peptide retention times.

Considering this, we present two logical next steps for this work. First, perform an investigation that validates the inability of processing software to identify peptides from PGC separations and present a potential remedy. As we evaluate the shortcomings of some software in Chapter 5, and we present a strategy for retention time calibration in Chapter 8, we propose combining these two sets of knowledge may be a facile strategy to improve PGC-based analyses. Software such as DIA-NN and Prosit allow for the creation of *in silico* spectral libraries that contain theoretical fragment ion intensities and retention times. Using the capsule network machine learning approach in Chapter 8, we may collect a set of empirical peptide identifications, train a model that correctly predicts PGC-specific retention time, and use this trained model to recalibrate the *in silico* libraries used for identification. We anticipate this will result in higher peptide identifications for all analyses, more accurate protein quantitation, and the ability to perform biological discovery to a modest scale. Secondly, we propose the development of informatic approaches to correctly account for PGC retention times and eliminate the shortcomings associated with poor data processing. Doing so will entail collecting numerous PGC-based datasets, analyzing the output spectra, and developing methods to correctly predict spectral information and retention time. Not only could this approach be used alone, it could be incorporated to one of the many academic-driven data processing pipelines (e.g., FragPipe) that will distribute the information to a broader audience.

In addition to these proteomic revelations, we also detail the use of PGC separations for glycoproteomic analyses. We detail in Chapter 2 the utility of PGC in untargeted investigations, highlighting the analytical advantages but suggesting some potential drawbacks of elevated

column temperatures. We investigated these drawbacks in Chapter 3, detailing the differential impact of high column temperatures on various glycopeptide classes. Taking guidance from these chapters, we propose future work may seek to leverage the isomeric resolution of glycopeptides to quantify the abundance of various glycopeptide isomers in model organisms. As we have shown the ability to distinguish isomers and have discussed quantitative approaches in Chapter 4, we believe mass defect labeling may be a suitable approach to increase the throughput of glycoproteomic analyses, as well as provide the capacity to ascertain quantitative differences in isomeric composition. Our investigations have relied on the availability of complex prostate cancer (PCa) cell lines but have not yet sought to quantify glycoproteomic differences across phenotypes. Doing so would not only further validate the use of PGC-based glycoproteomics, but would also provide the first reports of how glycan composition changes in accordance with PCa progression.

Finally, Chapters 7 and 8 have explored the utility of data-independent acquisition (DIA)-MS for biological discovery. These two studies, quantifying the proteomic alterations seen in PCa and Alzheimer's Disease, respectively, were successful in provide broad proteomic coverage and detailing numerous pertinent dysregulation events specific to a given disease. The individual discoveries may be read above, but the overarching conclusion of these works is that DIA analyses present unparalleled access to the proteome, substantially outpacing traditional DDA analyses. Based on the promise and room for growth in PCa investigation, we propose future work should seek to expand the analyses presented above (see Chapter 7) by expanding the breadth of phenotypes analyzed and through utilization of nascent DIA technology such as that offered in trapped ion mobility instrumentation (see Chapter 8). We are confident that utilization of this advanced technology, in tandem with the progress PCa cell model presented above, will present seminal reports into the biomolecular alterations specific to PCa progression.

Taken together, the alternative approaches to chromatographic separation and mass spectrometry acquisition presented here represent critical components in advancing proteomic analyses. Continual development in preparative, analytical, and informatic workflows are critical to propelling biomolecular investigations towards providing meaningful, actionable biological insight. These developments, regardless their reported success, will always be tempered against the known efficacy of the analytical techniques used for validation. The work presented here highlights the importance of including complementary and orthogonal separations to improve protein and peptide recognition in a variety of analyses. As well, we highlight the value of utilizing new, advanced data acquisition strategies that increase the breadth proteomic information achieved in a given experiment. Our work has demonstrated that consideration and utilization of these alternatives substantially enhance biomolecular investigations, alleviating in some part the obstacles hindering comprehensive proteomic profiling.