MS-BASED CHEMICAL PROTEOMICS STUDIES OF EXTRACELLULAR GLYCOPROTEINS: IDENTIFICATION, QUANTIFICATION, AND DYNAMICS

A Dissertation
Presented to
The Academic Faculty

by

Suttipong Suttapitugsakul

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
May 2021

MS-BASED CHEMICAL PROTEOMICS STUDIES OF EXTRACELLULAR GLYCOPROTEINS: IDENTIFICATION, QUANTIFICATION, AND DYNAMICS

Approved by:

Dr. Ronghu Wu, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Matthew Torres
School of Biological Sciences
Georgia Institute of Technology

Dr. Facundo M. Fernández
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Loren D. Williams
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Pamela Pollet
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: March 16, 2021

ACKNOWLEDGEMENTS

This so-called journey to a Ph.D. has been a big mixed bag of emotions. It was long, short, fun, sad, exciting, depressing, and productive. I did not know nor expect what I would go through when I started this journey five years ago. When you start a Ph.D., people think that you are smart. Your parents think it is five or more years of classes. Your friends tell you their experience and warn you to stay away. Meanwhile, you are far away from home, family, and friends, dealing with work and life while wondering what will happen tomorrow, next week, or next month. Will this experiment work? Will I pass this 200th exam? It has been quite a journey. Looking back, I could not believe how much I have grown in the past five years not only as a scientist but also as a person, a friend, a son, and a colleague. While I would not suggest anyone (at least anyone who is sane!) to pursue a Ph.D., I would truly and honestly do it all over again. This would not have been completed without these people who supported me throughout my time here, and I would like to thank them for their invaluable assistance throughout the years.

I would first like to express my deepest gratitude and thank my advisor, Dr. Ronghu Wu, for the continuous support, expertise, valuable suggestions, and insightful feedbacks provided in the past five years. I remember seeing his talk at Pittcon when I was an undergraduate in college and thought that I would want to go into this direction in the future. Luckily enough, I decided to apply to Georgia Tech, got accepted, decided to come here, and eventually got to work in this lab. I could not have asked for a better advisor who cares about the success of his students. Ronghu helped me greatly with both my research and life problems, which, sadly, I have a lot. Deciding which graduate school to attend was not a hard choice for me. People said I was crazy to not pick an Ivy league school or a

school with the best analytical chemistry program in the US to pursue my education. Looking back, I do not regret my decision a single bit.

I would like to thank my committee members, including Dr. Facundo M. Fernández, Dr. Pamela Pollet, Dr. Matthew Torres, and Dr. Loren D. Williams, who offered useful advice regarding my research and life. I am grateful that our graduate program requires several meetings with the members, and through these meetings, I have learned more about each of you and am grateful that you do not only care about my research progress but also my life quality and my future career. Who would have thought that I would be discussing my electric guitar lesson during one of these meetings?

I would like to thank my former advisor from college, Dr. Linda B. McGown, who taught and shaped me into the analytical scientist that I am today. I was very lucky to have worked in your lab and got into the field when I was only a college sophomore. I would also like to thank my former teachers and professors from schools in both the US and Thailand who shaped me into this person today, particularly Dr. Sirirat Choosakoonkriang of Silpakorn University where I interned at back in Thailand.

I would like to thank many friends who I have met during the time here at Tech. I would first like to thank Dr. Johanna Smeekens, who was my desk mate and taught me several experimental skills that I still use these days. I would like to thank Dr. Haopeng Xiao, who gave me several insights into my experiments. While we only overlapped for a few years, I am truly grateful for the knowledge and advice, both research and life, provided by these two. I would like to thank Dr. Ming Tong and Fangxu Sun; we struggled and had fun through the several years of graduate school, and I am grateful for the useful discussion

and meals we had together. I would also like to thank other members of the Wu lab, including Chendi Jiang, Senhan Xu, Kejun Yin, and Xing Xu. I was also lucky enough to have mentored several undergraduate students, including Lindsay D. Ulmer and Claire Christman, and REU students, notably Jarett Eshima. It is a pleasure to watch you grow into a person and a scientist.

I would like to thank my Thai friends who encouraged, supported, and got me through the past five years. While we knew each other from a long time ago, I am glad that we are still connected and that I have a chance to watch you all succeed. Particularly, I would like to thank Thamonwan Diteepeng who encouraged me all the way from Zurich to step out of my comfort zone so I can grow and become a better person and taught me how to deal with obstacles in life. I am forever grateful for that.

Lastly, I would like to thank my parents, Charoon and Namkang Suttapitugsakul, who taught me the importance of education and devoted their lives to my success. They emphasized the opportunities that are available throughout life if one is highly educated. I would like to thank you for your love, support, and advice during stressful times. I came to the US when I was a high school student, and I still remember that we would chat every morning before I went to class, through tough and happy times. And I would not be here or have what I have without these two. I would also like to thank my older sister, Lalinthip Suttapitugsakul, for the support throughout the years, as well.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

2-ME	2-mercaptoethanol
4-VP	4-vinylpyridine
AA	Acrylamide
Ac_4GalNAz	N-azidoacetylgalactosamine-tetraacylated
Ac_4GlcNAz	N-azidoacetylglucosamine-tetraacylated
Ac_4ManNAz	N-azidoacetylmannosamine-tetraacylated
ACN	Acetonitrile
AGC	Automatic gain control
ANOVA	Analysis of variance
CD	Cluster of differentiation
CID	Collision induced dissociation
CSC	Cell Surface Capture
CTX	Cyclophosphamide
DAVID	Database for Annotation, Visualization and Integrated Discovery
DBCO	Dibenzocyclooctyne
DMEM	Dulbecco's modified eagle's medium
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EMT	Epithelial-to-mesenchymal transition
ER	Endoplasmic reticulum
FA	Formic acid
FBS	Fetal bovine serum
FDR	False discovery rate

GO	Gene ontology
HCD	Higher-energy collisional dissociation
HEPES	*N*-(2-hydroxyethyl)piperazine-*N*'-2-ethanesulfonic acid
HILIC	Hydrophilic interaction liquid chromatography
HPLC	High-performance liquid chromatography
IAA	Iodoacetamide
iBAQ	Intensity-based absolute quantification
KEGG	Kyoto encyclopedia of genes and genomes
LC	Liquid chromatography
LDA	Linear discriminant analysis
LFQ	Label-free quantification
LPS	Lipopolysaccharides
LTQ	Linear trap quadrupole
MS	Mass spectrometry
NaCl	Sodium chloride
N-EM	*N*-ethylmaleimide
NHS	*N*-hydroxysuccinimide
PBS	Phosphate buffered saline
PMA	Phorbol 12-myristate 13-acetate
PNGase F	Peptide N-glycosidase F
PTM	Post-translational modification
RPMI	Roswell park memorial institute medium
SDC	Sodium deoxycholate
SDS	Sodium dodecyl sulfate
SFM	Serum-free medium

SILAC	Stable isotope labeling by amino acid in cell culture
STRING	Search tool for the retrieval of interacting genes/proteins
TCEP	tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
TGF-β	Transforming growth factor-beta
THPP	tris(3-hydroxypropyl)phosphine
TMT	Tandem mass tag
TOPO	Topotecan

SUMMARY

Extracellular proteins, including cell-surface and secreted proteins, play crucial roles in cell-cell communication, cell-matrix interactions, and the regulation of most extracellular biological processes. The majority of these proteins, if not all, are glycosylated, which affects their functions, properties, stability, and dynamics. Cell-surface glycoproteins function as receptors, transporters, and adhesion molecules, while secreted glycoproteins, including cytokines, chemokines, and growth factors, serve as signaling molecules that mediate the communication between cells. The abundance of extracellular glycoproteins reflects the disease or developmental states of the cells and, hence, can be used as biomarkers for disease diagnosis. Despite their importance, extracellular glycoproteins cannot be easily studied on a large-scale due to their naturally low abundance compared with cytosolic proteins. Several enrichment methods are not specific for glycoproteins or do not target all glycan types. Others lack extracellular specificity or require the harsh, oxidative conditions that may lead to cell death. Traditional antibody-based methods such as flow cytometry or ELISA do not allow for large-scale analysis. Furthermore, the glycosylation information, especially the glycosylation sites, cannot be obtained through these traditional methods. As a result, the knowledge of extracellular glycoproteins, including their presence, distribution, abundance, and dynamics in different cell types and biological settings, is not widely available.

This thesis work is focused on the study of extracellular glycoproteins, including surface and secreted glycoproteins, through mass spectrometry (MS)-based proteomics. First, an enrichment method based on metabolic labeling with a sugar analog, copper-free click chemistry, and MS-based proteomics is employed to target cell-surface glycoproteins

from multiple cell lines. The method is universal for all glycan types, targets surface glycoproteins only on the extracellular side, works under physiological conditions, and reveals site-specific glycosylation information. The occurrence and distribution of cell-surface glycoproteins across eight cell types are investigated. Both the relative abundance across the cells and the absolute abundance within the cells are quantified to generate a useful resource for surface glycoprotein distribution. Additionally, the glycosylation sites, non-canonical glycosylation motifs, protein secondary structure, and solvent accessibility of these sites are determined through several bioinformatics tools. The approach to study surface glycoproteins is then combined with multiplexed proteomics quantification using the tandem mass tag (TMT) reagents to determine the dynamics of cell-surface glycoproteins in the innate immune response to bacterial infection. Apart from surface glycoproteins that increase or decrease their abundance during the immune process, transient changes of several glycoproteins that may not be easily observed are captured. The dynamics of surface glycoproteins on monocytes and macrophages are determined and compared, and the monocyte-to-macrophage differentiation is demonstrated as a source of the different responses in the two cell types. The dynamics of surface glycoproteins in cells during the epithelial-to-mesenchymal transition (EMT) is then investigated, and proteins that participate in the partial EMT state are uncovered. For secreted glycoproteins, especially in the cell culture model, the analysis is often not comprehensive due to the presence of serum proteins in the media required for cell growth and proliferation. Several studies employ serum-free media, but it has been shown that this starvation alters the secretion of several proteins even after a short period of time. Therefore, a method to study secreted glycoproteins in the cell culture model without serum depletion is developed by

combining the enrichment method for surface glycoproteins, benefitting from the synthesis of extracellular glycoproteins through the classical secretory pathway, with a signal boosting technique through multiplexed proteomics quantification. The potential biological sources of secreted glycoproteins to use as a boosting sample are first determined, and the importance of the boosting-to-sample ratio is also demonstrated. This approach quantifies hundreds of secreted glycoproteins, including cytokines, from monocytes and macrophages in the innate immune response to bacterial infection model and from Hep G2 cells during the EMT process. Additionally, the reduction and alkylation steps of the bottom-up proteomics workflow are optimized to maximize the number of proteins and peptides that can be identified by MS. An enrichment step with thiopropyl Sepharose 6B beads for cysteine-containing peptides is further investigated. Other post-translational modifications are also explored, including phosphorylation in neuroblastoma tumors to understand their response to chemoimmunotherapeutic treatment. Not only the methods developed from this thesis will be useful for the study of extracellular glycoproteins, but the information regarding cell-surface and secreted glycoproteins is also unprecedented and will lead to a better understanding of their functions, benefiting the biological and biomedical fields.

CHAPTER 1. INTRODUCTION

Glycosylation is one of the most common protein modifications and is essential for cells. This modification is exceptionally complex because glycans are highly diverse and can be covalently attached to several amino acid residues in proteins through various configurations. There are two major types of protein glycosylation, i.e., N-linked glycosylation and O-linked glycosylation, where glycans are attached to asparagine and serine or threonine, respectively.[1-2] Glycosylation plays vital roles in cells, including determination of protein folding, trafficking, and stability, signal transduction, and regulation of nearly every extracellular activity such as cell-cell communication and cell-matrix interactions.[3-4] Aberrant protein glycosylation is directly related to multiple diseases, including cancer, neurodegenerative disorders, pulmonary diseases, blood disorders, and genetic diseases.[5-6] Due to the importance and complexity of protein glycosylation in biological systems, there is a long-standing interest to develop innovative methods to study glycoproteins and apply them for biomedical research. Investigation of protein glycosylation has become more popular with the development of modern instrumentation and computational methods. According to a PubMed search using the keyword “glycosylation”, sixteen publications were listed during 1960-1970 while over twenty thousand studies were reported in the past ten years. With the growing interests in protein glycosylation, this trend is expected to continue in the next decades.

Mass spectrometry (MS)-based proteomics provides an excellent opportunity to globally analyze proteins and their modifications.[7-16] Nonetheless, it is still extremely challenging to comprehensively analyze protein glycosylation.[17] Unlike many other modifications with a fixed structure for the modified group, such as phosphorylation, the diversity of glycans prevents us from employing the commonly used database searching methods such as SEQUEST and Mascot to identify glycopeptides in bottom-up proteomics. Several glycoproteins are expressed in the extracellular regions, and these glycoproteins often serve as biomarkers for disease detection or drug targets. The high-abundance protein background hinders the detection of these low-abundance glycoproteins in complex biological samples. Furthermore, glycans can interfere with the fragmentation of the peptide backbone.[17] Innovative and effective methods are critical to overcome these hurdles and to allow for comprehensive analysis of glycoproteins using MS. In this chapter, after briefly introducing protein glycosylation, several enrichment methods to analyze protein glycosylation in the extracellular region are described, including those on the cell surface, the secretome, and the extracellular vesicles.

1.1 Overview of protein glycosylation

Glycosylation is a common and essential modification where glycans are covalently attached to proteins. Among several types of protein glycosylation, N- and O-linked glycosylation are the two major ones. For N-linked glycosylation, the glycan precursor $GlcNAc_2Man_9Glc_3$ is transferred *en bloc* to nascent peptides while their corresponding mRNAs are being translated by the ribosome. Thus, glycosylation is also called a co-translational modification. N-glycans normally contain a common $GlcNAc_2Man_3$ core that are further modified with various monosaccharides by enzymes in the endoplasmic

reticulum (ER) and Golgi apparatus, including the terminal sialic acid and core fucose residues. N-glycosylation sites typically have a canonical motif, i.e., N-X-S/T where X can be any amino acid residues except proline. However, the presence of this motif does not guarantee that a particular copy of the same protein would be glycosylated, which may be referred to as the macroheterogeneity of protein glycosylation. Glycans may contain several types of monosaccharides linked together through different modes of connections even at the same site on different copies of the same protein, which is referred to as the microheterogeneity of protein glycosylation, further increasing their complexity. For O-linked glycosylation, the glycans are attached to the side chains of serine and threonine. Mucin-type O-glycosylation is the most common protein O-glycosylation where, typically, monosaccharides are sequentially added to proteins by various glycosyltransferases, instead of the *en bloc* glycan transfer in N-glycosylation. O-GlcNAcylation, where the GlcNAc group is dynamically added or removed from serine and threonine similar to phosphorylation, belongs to this category as well. Unlike N-linked glycosylation, there is no canonical motif for O-linked glycosylation.

In a typical glycoproteomic study, proteins are first extracted from cells or tissues before glycoproteins or glycopeptides are enriched or separated from the complex samples. MS analysis can then be performed at different levels with different glyco-species, including glycoproteins, intact or derivatized glycopeptides, deglycosylated peptides, as well as the released glycans, which provide different information from glycosylation sites to glycan structures. As described in the introduction, glycosylation is essential for cells, and glycoproteins contain much valuable information regarding the statuses of cellular

development and disease. Therefore, it is critically important to identify glycosylation sites on glycoproteins and elucidate glycan structures.

1.2 Cell-surface glycoprotein analysis

The surface of eukaryotic cells is typically covered with sugars that are attached to various embedded proteins and lipids. This thick layer of glycoproteins and glycolipids, including N-linked and mucin-type O-glycoproteins, glycosylphosphatidylinositol (GPI)-anchored proteins, proteoglycans, and glycosphingolipids, surrounding the cell is called the glycocalyx.[18] Surface glycoproteins are generally synthesized through the classical secretory pathway and modified into the mature glycoforms by many enzymes in the ER and Golgi apparatus that add or remove sugars in the glycan moeity.[19-20] These cell-surface glycoproteins participate in many intra- and extracellular activities, including cell-matrix adhesion, cell-cell interaction, cell migration and motility, signal reception and activation of intracellular signaling pathways, vesicle-mediated transport, and molecule transportation across the plasma membrane.[21-22] The presence of specific surface glycoproteins can be used as a marker for the classification of cell types.[23] Aberrant protein glycosylation or abundance changes of surface glycoproteins can reflect the cellular statuses.[24-25]

Surface glycoproteins are conventionally studied by employing antibodies to specifically target proteins of interest. This is normally coupled with methods such as fluorescence microscopy or flow cytometry in immunophenotyping.[26-27] Besides the low throughput of these methods, the availability and specificity of antibodies could be an issue. MS can be exploited to alleviate these problems because it can detect thousands of proteins with high confidence in one experiment. Furthermore, antibodies are not required, and thus

no prior knowledge is needed for surface protein detection. However, many surface glycoproteins have low abundance, and it is challenging to distinguish surface glycoproteins from those inside cells. Therefore, selective separation and enrichment of surface glycoproteins are critical prior to MS analysis.

1.2.1 Hydrazide chemistry

One powerful method to enrich surface glycoproteins is through hydrazide chemistry where hydroxyl groups on glycans are converted to aldehyde groups and subsequently conjugated with a hydrazide group. In 2003, Zhang et al. developed a highly innovative MS method for surface glycoproteomic analysis based on this approach.[8] They identified 104 unique peptides from 64 surface glycoproteins in LNCaP cells. Zarif et al. isolated CD14+ monocytes from human blood and differentiated them *in vitro* into pro-inflammatory and anti-inflammatory macrophages. Using solid-phase extraction of N-linked glycopeptides (SPEG) by glycan oxidation and glycopeptide enrichment with hydrazide beads, they specifically identified glycoproteins on the surface of anti-inflammatory macrophages that may be involved in prostate cancer proliferation and metastasis, as opposed to the pro-inflammatory ones.[28]

Later on, the Cell Surface Capture (CSC) method was reported in 2009, where glycans were first oxidized with sodium periodate and a biocytin-hydrazide tag was employed to generate a chemical handle for the following enrichment with streptavidin beads.[29] About 100 surface glycoproteins were identified and quantified in each experiment. The CSC method was also applied to study surface glycoproteins in 41 human and 31 mouse cell types and, on average, 284 surface glycoproteins were identified from each cell type.[30] The data were compiled for the Cell Surface Protein Atlas (CSPA)

database. Using CSC, surface glycoproteins were identified from four human lymphocyte cell lines and human induced pluripotent stem cells, respectively, and the results may be used in cell type classification and drug discovery.[31-32] CSC has been modified into an automated system with smaller number of cells required, termed autoCSC.[33]

1.2.2 Enzymatic approach

Chemical oxidization of *cis*-diol groups using the oxidants such as sodium periodate in hydrazide chemistry-based methods may affect the cell viability. Minimizing cell death is critical for surface glycoprotein analysis because many highly abundant intracellular proteins could leak out and interfere with the tagging and enrichment of surface glycoproteins. Recently, a milder approach using galactose oxidase was reported. Ramya et al. employed this method, together with aminooxy-biotin tagging and streptavidin enrichment, to identify 68 glycoproteins on the cell surface.[34] We optimized this method and coupled the enzymatic reaction with hydrazide chemistry-based enrichment (Figure 1.1). The approach was further improved by the pretreatment of cells with neuraminidase to remove the terminal sialic acid residues. Moreover, the addition of horseradish peroxidase during the oxidation reaction, which consumes hydrogen peroxide (one of the oxidation products), pushes the reaction to completion.[35] Using this approach, we identified, on average, 953 N-glycosylation sites from 393 surface glycoproteins per experiment in MCF7 cells. Combining with SILAC, the approach enabled us to quantify 909 unique N-glycopeptides from 334 surface glycoproteins with 65% being down-regulated by over two-fold in cells treated with brefeldin A, which inhibits protein secretion through the classical secretory pathway.

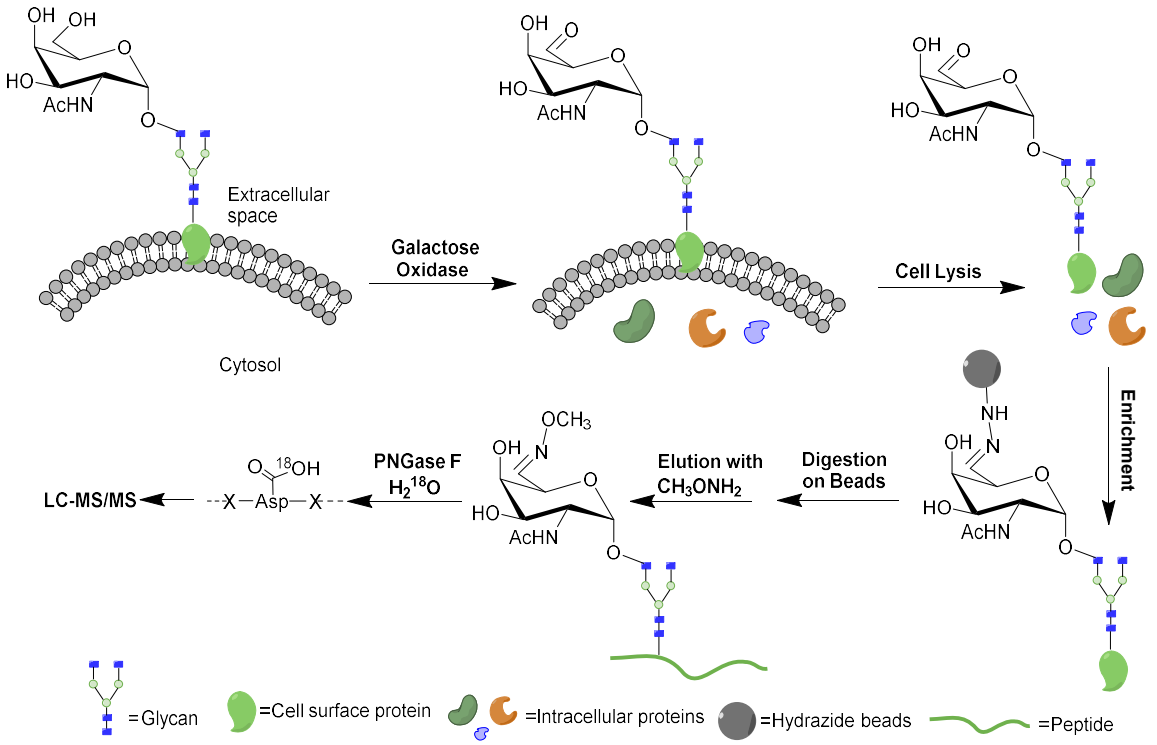

Figure 1.1. Enzymatic approach to identify cell-surface glycoproteins. Glycans on the cell surface were first oxidized with galactose oxidase (GAO) and immobilized on hydrazide beads after cell lysis. Surface glycopeptides were eluted from the beads with methoxylamine before PNGase F treatment and LC-MS/MS analysis. Reprinted with permission from Sun, F.; Suttapitugsakul, S.; Wu, R., Enzymatic Tagging of Glycoproteins on the Cell Surface for Their Global and Site-Specific Analysis with Mass Spectrometry. *Anal. Chem.* 2019, 91 (6), 4195-4203. Copyright 2019 American Chemical Society.

1.2.3 Metabolic labeling

Another approach to analyze cell-surface glycoproteins is through metabolic labeling with sugar analogs.[36-37] This method resulted from the development of bioorthogonal chemistry, where a specific chemical reaction in biological systems is achieved through functional groups that do not naturally exist or interfere with normal biological activities.[38-39] It allows for selective studies of biomolecules including glycoproteins on the cell surface.[40] The sugar analog is usually peracetylated to increase

its passive diffusion rate into cells, such as N-azidoacetylmannosamine-tetraacylated (Ac_4ManNAz). The acetyl groups are deacetylated by intracellular esterases into azidosialic acid, which is incorporated into glycoproteins including the surface ones, through the Roseman-Warren biosynthetic pathway.[41] ManNAz was successfully applied to identify sialylated surface glycoproteins in both cultured cells and model animals such as mouse.[40, 42] Notably, metabolic labeling with sugar analogs is very compatible with cultured cell because cells can be grown in a medium containing these sugar analogs and incorporate them into their surface glycoproteins. Furthermore, Spiciarich et al. applied this method to study human prostate cancer tissue *ex vivo* through tissue slice culture, in which the tissues stayed metabolically active for days.[43] The tissues were cultured in the presence of Ac_4ManNAz, and surface glycoproteins were tagged with the biotin-alkyne reagent before enrichment with avidin resins. Over 900 proteins were detected from the normal and cancerous prostate tissues. Among those, 68% were membrane or secreted proteins and 45% were known glycoproteins.

Metabolic labeling with Ac_4ManNAz may be applied to determine the glycan-protein interactions. Li et al. developed a beautiful method called protein oxidation of sialic acid environments (POSE) to study surface protein interactions (Figure 1.2).[44] Cells were first metabolically labeled and tagged with dibenzocyclooctyne (DBCO)-(S)-1-(p-bromoacetamidobenzyl)ethylenediaminetetraacetate (FeBABE), followed by hydrogen peroxide treatment and quenched with methionine amide hydrochloride. The treatment with hydrogen peroxide generated radical species that can oxidize proteins nearby. The study identified 150-200 proteins that were oxidized from each cell line. Interestingly, they also evaluated the incorporation efficiency of azidosialic acid into glycans. While the

incorporation rate in PNT2 cells was as high as 87%, the efficiency in Caco-2 cells was only 18%. The conjugation efficiency between DBCO and azide was estimated to be over 86%. The variation of the incorporation efficiency needs to be improved in metabolic labeling with the sugar analog among different cell types in order to increase the quantification accuracy.

Another concern with Ac_4ManNAz labeling is its effects on cell viability and proliferation. In two recent studies from Han et al., cells were treated with different concentrations of Ac_4ManNAz, Ac_4GlcNAz, and Ac_4GalNAz from 10-50 μM for 72 hours.[45-46] Higher concentrations of Ac_4ManNAz were found to affect biological processes such as the MAPK activity, apoptotic process, and immune and inflammatory response according to transcriptomic analysis. The authors suggested that labeling with 10 μM Ac_4ManNAz should be sufficient. However, the labeling time and the cell type need to be further considered. Our lab previously compared the three sugar analogs in a surface glycoprotein identification experiment and found that labeling with 100 μM Ac_4GalNAz resulted in the highest coverage of surface glycoproteins from HepG2 cells.[47] We also compared the labeling with 10-250 μM Ac_4GalNAz, 100 μM Ac_4GalNAc, and a vehicle control group in A549 cells. Within the incubation time of 24 hours, nearly all quantified proteins in these tested conditions were not affected by Ac_4GalNAz treatment (unpublished data).

Ac_4GalNAz has been used in several applications for surface glycoprotein analysis.[37] For example, Xiao et al. combined the method with pulse-chase labeling to study the dynamics of cell-surface glycoproteins and measure their half-lives.[48] Cells were first labeled with Ac_4GalNAz and then tagged with DBCO-biotin before being switched to

the medium without the sugar analog. The cells were collected at different time points, and the tagged surface glycoproteins were separated. After digestion and enrichment, the glycopeptides were labeled with the TMT reagents. It was found that the half-lives of surface glycoproteins were generally longer than those of newly synthesized proteins, which may be due to the presence of glycans that can protect proteins from being degraded.

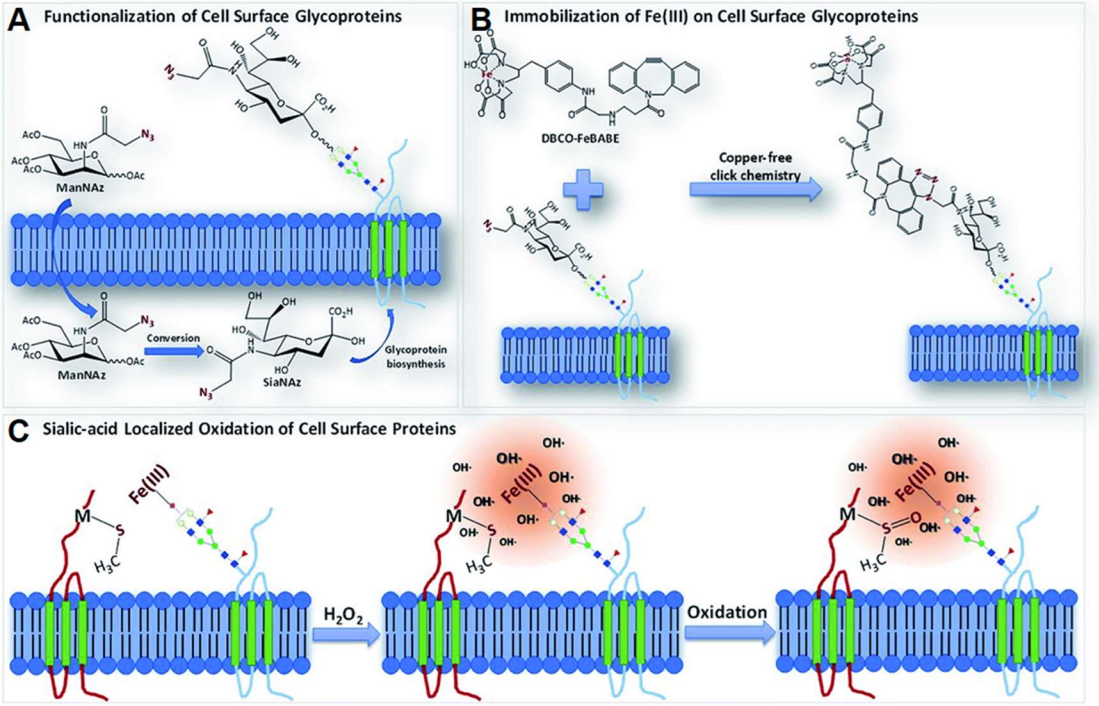

Figure 1.2. Protein oxidation of sialic acid environments (POSE) approach to map protein environment of sialic acid. Cells were first labeled with ManNAz (A) and tagged with DBCO-FeBABE (B) before hydrogen peroxide treatment (C). Amino acids in the vicinity were oxidized by radical species and proteins were subsequently analyzed with LC-MS/MS. Reproduced from Li, Q.; Xie, Y.; Xu, G.; Lebrilla, C. B., Identification of potential sialic acid binding proteins on cell membranes by proximity chemical labeling. *Chem. Sci.* 2019, 10 (24), 6199-6209. under the Creative Commons Attribution-Non-Commercial 3.0 Unported License – Published by The Royal Society of Chemistry.

1.2.4 Direct tagging of surface proteins

Other methods target all proteins on the cell surface even though the majority, if not all, of surface proteins are glycosylated. For membrane separation by

ultracentrifugation, contamination of membrane proteins from other cellular compartments could be an issue.[49-50] Alternatively, surface proteins may be directly biotinylated for further separation. In a study from Hormann et al., three surface protein purification methods were compared, including sulfo-NHS-SS-biotinylation, aminooxy-biotinylation, and surface coating with silica beads. Sulfo-NHS-SS-biotinylation outperformed the other two in identifying surface proteins, with the localization of over 90% of the proteins on the plasma membrane.[51] In the same study, instead of a typical avidin denaturation elution by SDS, elution through biotin competition further increased the coverage.

Matta et al. employed a method similar to CSC, but the aminooxy-biotin tag was used, to compare the surfaceomes of chondrogenic progenitor cells and bone marrow mesenchymal stem cells. Even though the two types of cells are very similar, distinct populations of surface glycoproteins were observed.[52] Kalxdorf et al. successfully combined chemical oxidation of glycans with alkoxylamine-PEG_4-biotin tagging and multiplexed proteomics to analyze changes of the surfaceome during macrophage differentiation over 72 hours, as well as the effects of a kinase inhibitor on the cell differentiation.[53]

1.2.5 In silico prediction of surface glycoproteins

Although we can currently detect a few hundred surface glycoproteins in a single experiment, the total number of glycoproteins present on the cell surface still remains elusive. Bausch-Fluck and Goldman et al. employed machine learning to predict if a glycoprotein can be localized on the surface based on the experimentally derived surface glycoproteins from the CSPA database and those from inside of cells in other subcellular compartments. The predictor, called SURFY, scores proteins based on factors determined

from a random forest classifier, including the frequencies of the N-X-S/T motif and cysteine residues, the presence of C-glycosylation site, and the length and number of transmembrane domains. This approach predicted a total of 2,886 proteins that could be present on the surface of human cells with the accuracy of 93.5%, corresponding to 14.3% proteins in the human proteome that may be present on the cell surface.[54] When the results were matched with transcriptomic data from an RNA-seq experiment, only a subset of these proteins was identified in a specific cell type, such as 507 proteins from HeLa cells, corresponding well with the number of surface glycoproteins typically identified from experiments.

1.3 Secreted glycoprotein characterization

Cells secrete proteins into the extracellular environment, and most of them are glycosylated through the classical secretory pathway.[55-57] These glycoproteins participate in many extracellular events including cellular signaling and cell-cell communication. Combining the results from antibody detection with those from MS analysis, Uhlen et al. determined the data set of human secreted proteins based on protein characteristics such as the presence of signal peptides or the transmembrane region, and showed that the number of potential genes encoding secreted proteins was 2,641, which is about 13% of all human protein-coding genes.[58] Secreted glycoproteins in bodily fluids can serve as excellent non-invasive sources of biomarkers for disease detection.[59-60] Particularly, glycosylated proteins also have enhanced stability, increasing the possibility of their detection.[61] Several molecular biology approaches, especially antibody-based methods, have been employed to study secreted glycoproteins.[62-63] Nevertheless, these methods do not allow for large-scale analysis of secreted glycoproteins nor reveal the information of the glycosylation site or

the glycans. MS-based proteomics is powerful for secreted glycoprotein analysis. However, there are some challenges that must be overcome to effectively identify and quantify secreted glycoproteins on a large scale. In this section, several approaches to study secreted glycoproteins from different biological samples using MS and their applications in the biological and biomedical fields are reviewed.

1.3.1 Extensive and multidimensional fractionation

One major obstacle in global analysis of secreted glycoprotein is the extremely high dynamic range of proteins present in the culture medium, blood, or plasma.[64-65] In serum, the concentration of albumin could be in the tens of grams per liter,[66] while others may be present in the micrograms per liter or less.[67] Because MS is biased for abundant proteins/peptides, especially with the popular data-dependent acquisition approach, those highly abundant proteins such as albumin are typically dominant in secreted protein analysis. However, many important proteins including those released from diseased cells normally present at a very low abundance are extraordinarily challenging to be detected using MS-based proteomics.[68] The simplest method to decrease the complexity of proteins/peptides is to perform extensive, multidimensional fractionation. After protein digestion, peptides are fractionated through orthogonal separation methods prior to MS analysis. The fractionation decreases the complexity and improves the chance for the detection of low-abundance peptides by MS. While extensive fractionation greatly improves the coverage of secreted proteins, the analysis is time-consuming, as seen by the long analysis time from the study by Dey (>500 hours) to cover ~5,000 proteins.[69]

1.3.2 Depletion of highly abundant proteins

Since high-abundance proteins can affect the detection during MS analysis, another approach to reduce the complexity is through depletion columns. Several types of these columns are commercially available. These columns are based on antibodies that can capture the proteins with the high abundance in samples. In one example, CaptureSelect HumanPlasma 14 affinity resin was used to remove proteins including albumin, immunoglobulins, transferrin, and others from plasma samples before multi-lectin affinity chromatography for glycoprotein enrichment.[70] The authors identified glycoproteins that are differently secreted from prostate cancer and benign prostatic hyperplasia patients. Similarly, Berva et al. combined protein depletion with solid phase extraction of N-linked glycoproteins (SPEG), which is based on a reaction between oxidized glycans and hydrazine resins, to enrich glycoproteins from human plasma.[71] Compared with the experiment without the depletion, the number of identified glycoproteins increased by 24%.

While depletion columns can effectively remove high-abundance proteins from serum samples, studies have shown that these columns may also result in the sample loss, which affects the quantification of proteins and the reproducibility.[72-73] One of the proteins with the highest abundance is albumin, which can bind to several secreted proteins.[74] Therefore, the removal of albumin caused the loss of other secreted proteins.[75] The capacity of these columns is normally low, and thus the approach is not very cost-effective. Additionally, non-specific binding of proteins to the material used in these columns may occur.[76] De Jesus et al. compared three depletion methods to remove the most abundant proteins from cells, including magnetic nanoparticles, sequential application of dithiothreitol and acetonitrile, and the commercial ProteoMiner apparatus based on

immunoaffinity.[77] The results revealed that the magnetic nanoparticle depletion was more versatile, reproducible, and removed proteins with high molecular weights (>80 kDa), resulting in the highest number of identified proteins. For the chemical method, while the depletion of proteins was also observed, the number of identified proteins was lower. The ProteoMiner apparatus was not cost-effective and showed similar results to the chemical method.

1.3.3 Serum starvation

In cell culture models, serum such as Fetal Bovine Serum (FBS) is often used in the growth media since it contains proteins and other small molecules required for *in vitro* cell growth.[78] The concentration of proteins, especially albumin, is very high in the final culture media, which affects the detection of low-abundance, secreted glycoproteins. One solution to this problem is to grow cells in a serum-free medium for a short period of time.[79-80] Before switching to the serum-free medium, the cells must also be washed thoroughly to remove residual serum proteins. Roger et al. showed that three washes are optimal to remove serum proteins, while more washes can decrease cellular viability.[81] Additionally, washing cells with serum-free medium leads to the contamination of proteins from lysed cells.

One major problem with this approach is the effects of cellular starvation on cell growth and proliferation, and thus the alteration of protein secretion from cells. Cell death may also occur, resulting in the leakage of intracellular proteins into the extracellular space. This complicates and affects secreted glycoproteomic analysis. One study focused on enriching newly synthesized, secreted proteins from cells by combining azidohomoalanine

(AHA) labeling with pulsed SILAC.[82] AHA is a methionine analog that can be conjugated by aminoacyl tRNA synthetase during protein synthesis instead of methionine. Proteins labeled with AHA containing the azide group can be selectively enriched through click chemistry. Over 1,000 secreted proteins were identified and quantified including some extracellular matrix proteins and cytokines. The authors further applied this method for a time-resolved analysis of secreted proteins from mouse macrophages (RAW264.7) treated with lipopolysaccharides after 0, 6, and 17 hours. Several inflammatory cytokines were identified and quantified within a two-hour labeling period at each time point. The authors also showed that the secretion of proteins could be affected by incubating cells under serum-free conditions only after three hours, raising the concern that the results from secreted protein analysis under such conditions should be further examined and validated by other methods.

1.3.4 Lectin-based enrichment

Lectins are proteins that bind specifically to glycans and have been extensively used for glycoprotein enrichment.[83-86] They were employed to enrich secreted glycoproteins in the extracellular compartment.[87] In a study by Boersema et al., the authors profiled secreted glycoproteins from 11 breast cancer cell lines in different stages (Figure 1.3).[61] Two lectins, concanavalin A (Con A) and WGA, were combined with filter aided sample preparation (FASP) to enrich secreted N-glycoproteins under the serum-free conditions. They identified and quantified peptides with 1,398 N-glycosylation sites from 701 proteins. The results also demonstrated differential protein secretions from cancer cells in the different stages. While the authors showed that the serum-free condition did not affect the viability of the cells, the biological interpretation of the quantitation results from cells under such

condition might need to be carefully considered. Due to their glycan specificity, multiple lectins are often required to capture glycoproteins with diverse glycans in the secretome. Li and colleagues developed a method termed lectin affinity capture followed by solid-phase extraction of glycosite-containing peptides (LecSPEG).[88] Here, *Aleuria aurantia* lectin (AAL)-, *Sambucus nigra* lectin (SNA)-, and WGA-conjugated beads were used to enrich glycoproteins from serum samples. The eluted proteins were digested, and the resulting glycopeptides were further oxidized with sodium periodate for their enrichment with hydrazide beads. With this approach, the number of unique glycosites increased from 2 with lectin enrichment alone to 46 with the LecSPEG method, and the specificity increased from 2% to 82%.

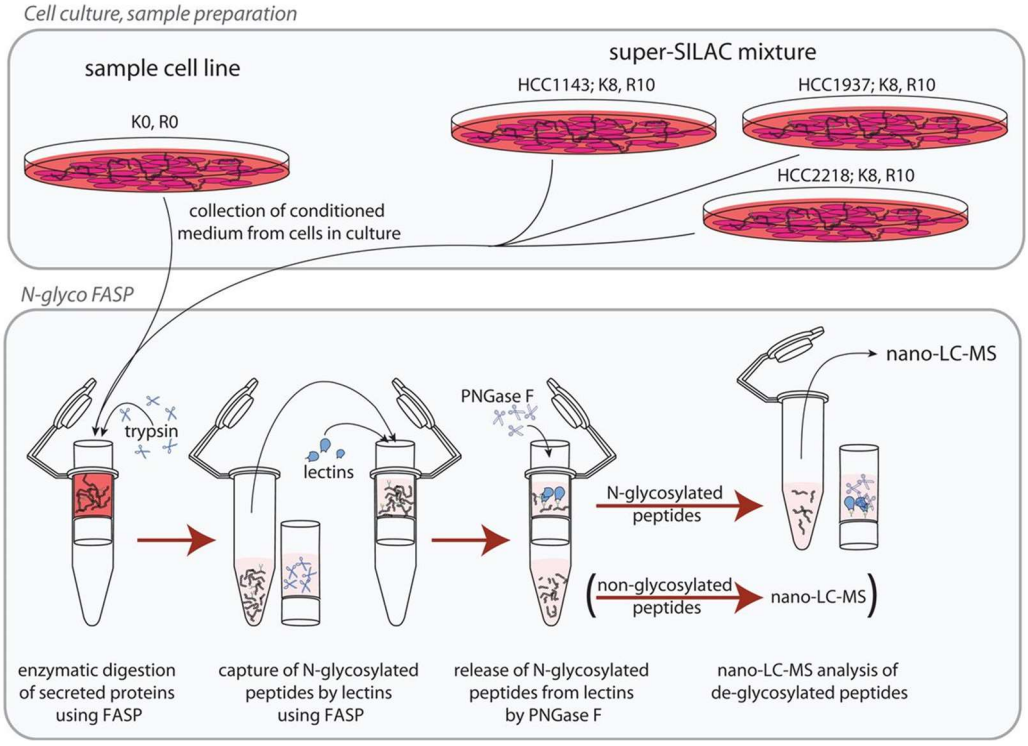

Figure 1.3. Workflow combining FASP with lectins to enrich secreted N-glycoproteins from cell culture media. Reproduced from Boersema, P. J.; Geiger, T.; Wisniewski, J. R.; Mann, M., Quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples. *Mol Cell Proteomics* 2013, 12 (1), 158-71. Copyright retained by the American Society for Biochemistry and Molecular Biology under the CC-BY license.

1.3.5 HILIC

In HILIC, hydrophilic peptides, including those with glycans, are retained in the column and eluted by an increasing amount of water in the mobile phase.[89] HILIC has been reported for the study of secreted glycoproteins in various samples.[90] ZIC-HILIC, where the stationary phase contains both positive and negative charges, was also employed for secreted glycoprotein analysis.[91-92] Other work has combined different separation modes in order to increase the specificity for glycopeptide enrichment.[93] In the work from Chen and colleagues, phosphopeptides and glycopeptides were simultaneously enriched through a method called mode switchable solid phase extraction (MS-SPE).[94] A solid-phase extraction column packed with PolyWAX was first used to separate both phosphopeptides and glycopeptides in the electrostatic repulsion hydrophilic interaction chromatography (ERLIC) mode. The column was then switched to the HILIC mode by using high concentration of organic solvent where phosphopeptides were eluted and glycopeptides were retained. Eventually, glycopeptides were eluted with water. With this approach, 136 phosphorylation sites and 283 N-glycosylation sites were identified from the secretome of A549 cells under serum-free conditions. Nevertheless, non-glycosylated peptides may also be hydrophilic and retained in these columns, which affects the enrichment specificity.

1.3.6 Boronic acid-based enrichment

Boronic acid forms reversible covalent bonds with the diol groups on glycans, allowing for the enrichment of glycoproteins in biological samples.[95-102] Boronic acid was compared with lectin-based enrichment (ConA, *Lens culinaris* agglutinin, or WGA) for glycoprotein analysis in human serum samples, and the highest number of identified

glycoproteins using boronic acid was possibly due to the wider array of glycans enriched.[103] For *Saccharomyces cerevisiae*, Smeekens and colleagues employed boronic acid-conjugated magnetic beads to enrich secreted glycopeptides from the growth medium.[104] The analysis requires the careful medium collection step to avoid cytosolic protein leakage, which can interfere with the detection of secreted glycoproteins. In this study, the cells were also treated with tunicamycin to inhibit protein N-glycosylation. Many secreted N-glycoproteins were down-regulated compared with the control group, showing the effects of N-glycosylation on protein secretion.

1.3.7 Hydrazide chemistry-based enrichment

Hydrazide chemistry-based methods have also been employed to enrich secreted glycoproteins.[105-106] In one example, combining MS analysis with hydrazide chemistry and ZIC-HILIC, Li and colleagues identified 1,213 unique N-glycosylation sites from 611 glycoproteins from two cell lines with low or high metastatic potential.[88] Differential regulation of several secreted glycoproteins was discovered based on the metastasis capability and could be used as biomarkers. Note that this was done under serum-free conditions.

While the data-dependent acquisition (DDA) method is biased for abundant peptides, data-independent acquisition (DIA) records all the peptides by spacing the measurement length into smaller sections. This minimizes the bias of MS in detecting low-abundance secreted glycoproteins. In one study, Sajic et al. employed Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH-MS) to detect secreted glycoproteins in the blood from several carcinomas.[107] Glycoproteins were enriched from the blood of

patients with localized lung, pancreas, ovary, prostate, and colorectal cancer, or their control matches using hydrazine resins, and subsequently analyzed by MS. Over 1,000 glycopeptides from >200 glycoproteins were quantified. The author identified proteins that are generally altered in patients with cancer and other glycoproteins that are unique and specific for each cancer type.

1.3.8 Metabolic labeling-based enrichment

Metabolic labeling with a sugar analog is powerful to analyze glycoproteins on the cell surface, as discussed above. This has also been used in secreted glycoprotein analysis by MS. For instance, Roper et al. employed ManNAz to label secreted glycoproteins from two stromal cell lines.[108] The azide-labeled glycoproteins are then enriched with alkyne beads through the CuAAC reaction. From the cells being serum-starved for 48 hours, 75 secreted glycoproteins were detected, while in 1% serum-containing media, 100 secreted glycoproteins were detected. Kuhn and colleagues developed the secretome protein enrichment with click sugars (SPECS) method for shed and secreted protein analysis of primary cells in the presence of serum proteins.[109] In this work, glycoproteins were first labeled using ManNAz. The media were then collected, and the glycoproteins were tagged with DBCO-PEG12-biotin before streptavidin enrichment and MS analysis. Later, this method was modified in the high-performance secretome-protein-enrichment-with-click-sugars method (hiSPEC), which decreased the number of cells required for secretome studies (Figure 1.4).[110] The authors also labeled glycoproteins with ManNAz and enriched secreted glycoproteins using Concanavalin A and DBCO beads. Secreted glycoproteins from different systems were identified, including those in brain slices upon LPS-induced neuroinflammation and primary astrocytes, microglia, neurons, and oligodendrocytes in a

cell-specific manner. The results need to be carefully considered since non-specific binding may occur. Additionally, the incorporation efficiency of these sugar analogs in different cell types must be considered.

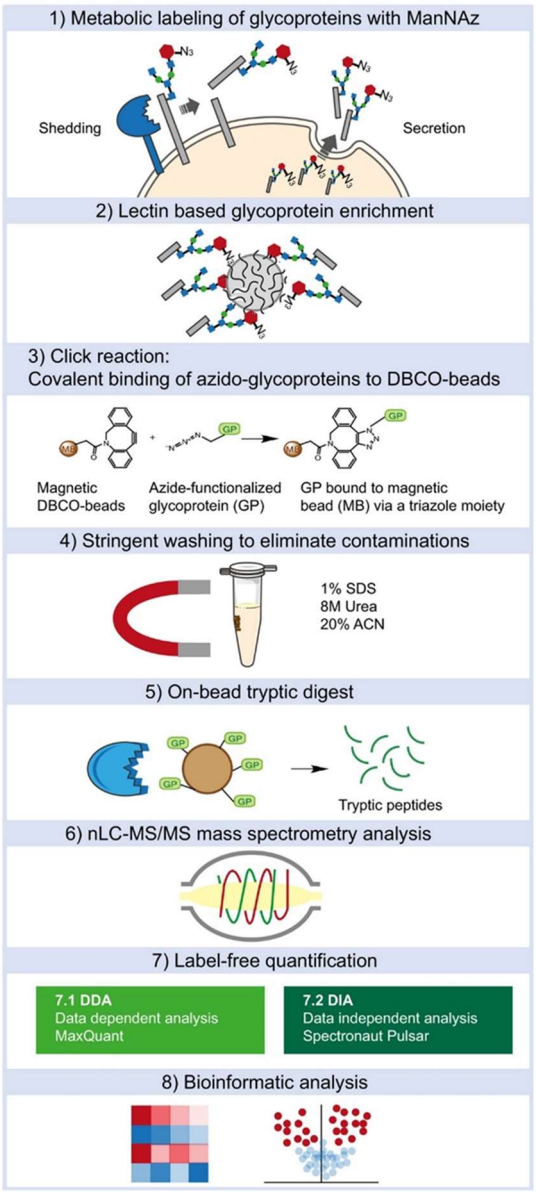

Figure 1.4. hiSPEC workflow for shed glycoprotein analysis. The cells are labeled with ManNAz, tagged through the click chemistry reaction, and the enriched glycopeptides were analyzed with LC-MS/MS. Reproduced from Tushaus, J.; Muller, S. A.; Kataka, E. S.; Zaucha, J.; Sebastian Monasor, L.; Su, M.; Guner, G.; Jocher, G.; Tahirovic, S.; Frishman, D.; Simons, M.; Lichtenthaler, S. F., An optimized quantitative proteomics method establishes the cell type-resolved mouse brain secretome. *EMBO J* 2020, 39 (20), e105693. Copyright retained by authors under the CC BY 4.0 license.

1.3.9 Signal boosting approach

The development of the TMT reagents allowed for the simultaneous identification and quantification of proteins in multiple samples. With this approach, peptides from different samples are labeled with different channels of the TMT reagents. The labeled peptides are then combined and analyzed with MS. The peptides from different samples appear as a single peak in MS1 spectra due to the same m/z. Once they are selected for MS2 analysis, the intensities of the reporter ions can be used for accurate quantification of peptides in different samples. This multiplex approach improves the reproducibility and shortens the analysis time. Furthermore, recent studies showed that the TMT labeling can improve the peptide signal intensity in MS1, facilitating the selection of low-abundance peptides for MS2 analysis. Notably, Budnik et al. employed this method for single-cell proteomics analysis (Single Cell ProtEomics by Mass Spectrometry; SCoPE-MS).[111] Another approach termed Boosting to Amplify Signal with Isobaric Labeling (BASIL) employed a similar approach for protein phosphorylation analysis.[112] In these approaches, a protein carrier sample, e.g., the greater number of cells in SCoPE-MS or the combination of proteins from the quantitation channels in BASIL, with a higher protein abundance compared with the quantitation samples was employed and labeled with one of the TMT channels (typically the last channel to avoid overlapping of the isotopic envelope).

Using the boosting approach, secreted proteins in clinical samples were studied.[113] Our lab recently combined the boosting approach with metabolic labeling using GalNAz to label secreted glycoproteins from cultured cells without serum starvation.[114] We compared different glycoprotein sources to use for the boosting channel and evaluated the effect of the boosting-to-sample ratios on the results. With this approach, over 200

glycoproteins were detected in a site-specific manner from different cell types, including monocytes, macrophages, and Hep G2 cells under serum-containing conditions.

1.4 Glycoprotein analysis in extracellular vesicles

Cells release several types of membrane-bound, extracellular vesicle (EV), including exosomes and microvesicles. Exosomes are intraluminal vesicles contained within the multivesicular body that, upon fusion with the plasma membrane, release their contents outside of the cells. Release of exosomes was once thought to be a method for waste disposal, but later studies showed that it can mediate intercellular communication through transferring molecules such as proteins, lipids, and RNAs to acceptor cells.[115-116] Exosome analysis has gained increased interest in recent years due to its roles in the immune system and the potential to serve as disease biomarkers and targeted drug delivery vehicles for therapeutics.[117]

MS analysis can identify exosomal proteins and their PTMs, including phosphorylation, ubiquitination, palmitoylation, sumoylation, and glycosylation, that may unveil the mechanisms of exosome formation and their biological significance.[118-119] Apart from the well-known proteins found in the exosomes such as tetraspanin CD63, certain glycoproteins, including MUC1, were highly enriched according to a recent data-independent MS analysis of the exosomes from non-small cell lung cancer carcinoma.[120] In a study by Chauhan et al., surface glycoproteins from the exosomes of myeloid-derived suppressor cells were oxidized and biotinylated for their enrichment with streptavidin beads, and the majority of the proteins identified can also be found on the cell surface of the parent cells.[121] Sialoglycoproteins were found to be enriched in the exosomes of ovarian

carcinoma cells. Several types of glycans were identified, including di-, tri-, and tetraantennary N-glycans, as well as the core fucose modification and high-mannose glycans. In diantennary glycans, bisecting GlcNAc was detected.[122] Sharma et al. compared the exosomes that were released from cancer cell lines with different invasiveness. It was found that the exosomes from the highly invasive cells, which contained several unique glycoproteins involved in cell migration, were correlated with greater cell migration than those from the less invasive ones.[123] In cells, glycosylation can regulate the trafficking of proteins to the cell surface. Similarly, glycosylation affects the trafficking of proteins to the exosomes. For protein EWI-2, the abundance was decreased when its complex N-linked glycan synthesis was limited to only the high-mannose ones.[119]

The exosomes contain specific molecules from their parent cells and, therefore, glycoproteins detected may serve as biomarkers for disease detection. Glypican-1, a cell-surface glycoprotein on the exosomes, was identified as a potential biomarker for early detection of pancreatic cancer using MS.[124] In patients with benign and metastatic pulmonary nodules, conventional diagnosis is difficult to distinguish them apart. The exosomes were isolated from the plasma of these patients, and label-free quantification by MS revealed that glycoproteins, such as fibrinogen beta chain (FGB) and fibrinogen gamma chain (FGG), were up-regulated in the exosomes from patients with metastatic pulmonary nodules.[125] In patients with non-small cell lung cancer, alpha-2-HS-glycoprotein (AHSG) and extracellular matrix protein 1 (ECM1) were upregulated compared with normal patients.[126]

Global analysis of glycoproteins in the exosomes may suffer from the high dynamic range of proteins, and thus effective enrichment methods can increase the coverage

especially for the low-abundance ones. For example, Bai et al. recently developed a hydrazide-based polymer that can homogeneously enrich N-glycoproteins from the exosomes, and the polymer can be recovered by raising the temperature.[127] Some protein markers such as aquaporin-2 and others related to renal diseases and blood pressure regulation were identified from the exosomes separated from urine.[128] The glycoproteins tetraspanins-1 and hemopexin were recently proposed as early biomarkers for T cell-mediated kidney transplant rejection.[129] Because of the amount of the exosomes needed for proteomic analysis, the large volume of urine may hinder the analysis. In one glycomic study, through a combination of miniaturized sample preparation and prefractionation, the volume of urine was decreased dramatically while comprehensive analysis of glycans was not compromised. The study confirmed the presence of sialylated glycans and several core fucosylations, as well as 16 mucin-type O-glycans and paucimannosidic glycans that were only reported in invertebrates and plants previously.[130]

Recently, a new type of non-membranous, extracellular nanoparticles termed exomeres was discovered trough asymmetric flow field-flow fractionation (AF4).[131] The exomeres are smaller and stiffer than the exosomes. Unique sets of proteins were found, including N-glycoproteins and those associated with the extracellular matrix, ER, mitochondrion, and cytoskeleton. Proteins involved in glycan processing and control of glycan-mediated protein folding, and sialoglycoproteins were both enriched in the exomeres, suggesting the possible roles in specific glycan recognition and glycosylation modulation of the recipient cells. Analysis with MALDI-TOF MS and LC-MS/MS revealed different N-glycan profiles between the exosomes and exomeres. Glycans from the exomeres contain α-2,3-linked and α-2,6-linked sialic acids while α-2,3-linked sialic

acids were found exclusively in glycans from small exosome vesicles. A later study showed that the enzyme β-galactoside α-2,6-sialyltransferase 1 (ST6Gal-I) was present on the membrane of the exomeres and can be transferred to the acceptor cells, causing hypersialylation on the cells.[132]

There are some major differences between different EVs, such as the size, density, biogenesis pathways, and protein markers. Ultracentrifugation or sucrose-gradient centrifugation is typically employed to isolate the exosomes based on their physical properties. This may lead to a mixed population of EVs due to their similar properties such as the overlapping size and density, which could cause different glycosylation profiles from the same sample.[133] Ideally, the identity or the presence of specific markers on different EVs needs to be confirmed before further characterization. The International Society for Extracellular Vesicles (ISEV) has updated the guidelines for exosome studies in 2018 (Minimal Information for Studies of Extracellular Vesicles, MISEV2018) and the term "exosome" may not be appropriate unless the origin of the vesicles can be verified. Otherwise, the term should be addressed as small EVs.[134] Databases such as ExoCarta contain information about proteins, RNAs, DNAs, and lipids that have been found in the exosomes and meet the minimum requirements from ISEV.[135-136] Other databases including information from extracellular vesicles (i.e. microvesicles, apoptotic bodies and membrane blebs) are Vesiclepedia and EVpedia.[137-138]

1.5 Conclusions

The promising progress and exciting development of novel methods in glycoproteomics have tremendously expanded the knowledge of protein glycosylation and

facilitated our understanding of extracellular glycoprotein functions. Because of the low abundance of many glycoproteins and the complexity of biological samples, it is imperative to enrich extracellular glycopeptides/glycoproteins prior to MS analysis. Several enrichment methods to analyze extracellular protein glycosylation are summarized in this chapter, including methods for surface, secreted, and extracellular vesicle glycoprotein analysis. Considering the importance of extracellular glycoproteins and the complexity of protein glycosylation, it is expected that the field of glycoproteomics will continue to thrive in the future. Global analysis of extracellular glycoprotein still requires the development of effective and innovative methods. Glycoproteins with a specific glycan represent a rich source to study the roles of abnormal expression of glycoproteins in of diseases such as cancer, and innovative methods are urgently needed to separate these glycoproteins for MS identifications and quantifications. New instrumentation with better sensitivity, resolution, and speed, and bioinformatic tools with better performance for interpreting MS spectra will not only advance the identification of glycoproteins, but also improve the throughput and the quantification accuracy. Furthermore, it will allow us to detect glycoproteins with extremely low abundance that often carry highly valuable information. Future glycoproteomics is likely to increase the depth of glycosylation analysis and to reveal the roles of glycoproteins in physiological and pathological processes, which will facilitate our understanding of the molecular mechanisms of human diseases and the discovery of glycoproteins as novel and effective biomarkers for disease detection and surveillance.

CHAPTER 2. SURFACE GLYCOPROTEOMIC ANALYSIS REVEALS THAT BOTH UNIQUE AND DIFFERENTIAL EXPRESSION OF SURFACE GLYCOPROTEINS DETERMINE THE CELL TYPE

Adapted with permission from Suttapitugsakul, S.; Ulmer, L. D.; Jiang, C.; Sun, F.; Wu, R., Surface glycoproteomic analysis reveals that both unique and differential expression of surface glycoproteins determine the cell type. Analytical Chemistry 2019, 91 (10), 6934-6942. Copyright 2019 American Chemical Society.

2.1 Introduction

Proteins located on the cell surface are normally modified with carbohydrates.[139] These surface glycoproteins play vital roles in nearly every extracellular event, including cell-cell communication, cell-matrix interactions, and cellular response to environmental cues.[140-141] Many surface glycoproteins function as ion channels and transporters for molecules across the plasma membrane, while others are receptors, such as G-protein-coupled receptors, that sense and mediate cellular responses to extracellular stimuli.[142-143] Enzymes and binding proteins located on the cell surface are also commonly glycosylated.[144]

Surface glycoproteins frequently reflect the developmental and diseased statuses of cells. A number of surface glycoproteins serve as disease biomarkers and for cell-type classification.[23, 145] Moreover, these surface glycoproteins are often the targets of macromolecular drugs such as antibodies or enzymes in the emerging immunotherapy field.[146] Comprehensive and site-specific analysis of cell-surface glycoproteins will aid in a better understanding of glycoprotein functions and cellular activities.

Immunophenotyping of surface glycoproteins has been performed using flow cytometry.[147] This technique, however, requires prior knowledge of the proteins of interest or the cell type. The availability and the specificity of antibodies and the low throughput restrict its applications. Additionally, flow cytometry provides limited information on the abundance of these cell-surface glycoproteins. Modern mass spectrometry (MS)-based proteomics provides a unique opportunity for global and site-specific analysis of proteins and their modifications.[11, 148-156] However, it is extraordinarily challenging to comprehensively and site-specifically analyze glycoproteins located only on the cell surface because of the following reasons. First, many glycoproteins occur in very low abundance, and their analysis is hampered by highly abundant proteins.[30, 157] Second, the heterogeneity of the glycans further complicates the analysis.[158-159] In addition, surface glycoproteins have to be effectively separated before MS analysis.

Previously, Zhang et al. developed an innovative MS-based method to identify surface N-glycoproteins by first oxidizing the glycans with $NaIO_4$, and the resulting carbonyl groups on the glycoproteins were used for the enrichment with hydrazide beads.[157] A few years later, Wollscheid et al. designed a beautiful Cell Surface-Capturing (CSC) method to identify and quantify cell surface N-glycoproteins. This was based on glycan oxidation with $NaIO_4$ and biocytin hydrazide tagging prior to the enrichment with streptavidin beads.[29] Despite the importance of glycoproteins on the cell surface, their comprehensive analysis is much understudied compared to global analysis of proteins.

In this work, we systematically studied glycoproteins on the surface of eight types of commonly used human cells. Surface glycoproteins were metabolically labeled with a functionalized sugar and then tagged with biotin through the strain-promoted copper-free

click chemistry reaction. Surface glycopeptides with biotin were selectively enriched and subsequently deglycosylated with PNGase F in heavy-oxygen water for site-specific analysis using MS. This approach allowed for global and site-specific identification of >2,000 N-glycosylation sites from >1,000 surface glycoproteins, with an average of 683 glycosylation sites and 354 surface glycoproteins per cell type. We also quantified glycoproteins using label-free quantification and discovered that only a small portion of the proteins are cell-specific while many were differentially expressed across the cell types. Furthermore, different groups of proteins were more highly expressed in one cell line than in the others, and served particular functions depending on the cell type. Benefiting from site-specific analysis, we explored the behaviors and occurrence of the glycosylation sites, including the solvent accessibility of the sites and the effect of protein structures on the sites. The current results lead to a better understanding of cell-surface glycoproteins and provide vital information in developing new biomarkers and drug targets.

2.2 Materials and methods

2.2.1 Cell culture

Human cell lines, including HEK293T, HeLa, Jurkat, K562, MCF7, and U266, were from the American Type Culture Collection (ATCC). HeLa and PANC1 cell lines were a generous gift from Professor M.G. Finn's lab. HEK293T cell line was from Dr. Gang Bao's lab. Cell lines were not authenticated. HEK293T, HeLa, HepG2, MCF7, and PANC1 cells were maintained in high-glucose Dulbecco's Modified Eagle's Medium (DMEM, Sigma-Aldrich) containing 10% fetal bovine serum (FBS, Corning). Suspension cell lines, including Jurkat, K562, and U266, were maintained in RPMI-1640 medium

(Sigma-Aldrich) containing 10% FBS. All cells were grown in a humidified incubator at 37 ºC with 5% carbon dioxide (CO_2).

2.2.2 Metabolic labeling and click-chemistry reaction

Adherent cells were cultured in high-glucose DMEM medium until the cells reached ~50% confluency. The cells were labeled with 100 µM N-azidoacetylgalactosamine-tetraacylated (Ac_4GalNAz, Click Chemistry Tools) in low-glucose DMEM medium (Sigma-Aldrich) with 10% FBS. Suspension cells were cultured in RPMI-1640 medium until the cell density was ~$7x10^5$ cells/mL as determined by hemocytometry and trypan blue staining and labeled similarly to the adherent cells. After 24 hours of metabolic labeling, the adherent cells were washed twice with Dulbecco's Phosphate Buffered Saline (DPBS, Sigma-Aldrich) while the suspension cells were centrifuged at 300*g* for 5 minutes to remove the medium and washed twice with DPBS similarly. Adherent cells were tagged with 100 µM water-soluble dibenzocyclooctyne (DBCO)-biotin (Click Chemistry Tools) in Cellstripper solution (Corning) for 1 hour in the humidified incubator. Suspension cells were labeled similarly except that DPBS was used instead of the Cellstripper solution. The reaction was quenched with 10 mM dithiothreitol (DTT, Sigma-Aldrich). The cell pellets were washed twice with ice-cold DPBS and kept on ice until the next steps.

2.2.3 Protein extraction and peptide purification

The cell pellets were incubated with a buffer containing 25 µg/mL digitonin (Sigma-Aldrich), 150 mM sodium chloride (NaCl, Sigma-Aldrich), 50 mM N-(2-hydroxyethyl)piperazine-N'-2-ethanesulfonic acid (HEPES, Sigma-Aldrich, pH=8.2), and

1 tablet/10 mL cOmplete ULTRA Tablets protease inhibitor cocktail (Roche) at 4 ºC for 10 minutes on an end-over-end rotator. The suspension was centrifuged at 2,000*g* for 10 minutes, and the supernatant was removed. An ice-cold lysis buffer containing 0.5% sodium deoxycholate (SDC, Sigma-Aldrich), 50 mM HEPES (pH=8.2), 150 mM NaCl, 20 units/mL universal nuclease for cell lysis (Pierce), and 1 tablet/10 mL cOmplete ULTRA Tablets protease inhibitor cocktail was added to the cell pellets. After the lysis at 4 ºC for 45 minutes on an end-over-end rotator, the suspension was centrifuged at 25,830*g* for 10 minutes. The supernatant was collected and reduced with 5 mM DTT at 56 ºC for 25 minutes and subsequently alkylated with 14 mM iodoacetamide (Sigma-Aldrich) for 30 minutes in the dark. The alkylation reaction was quenched by incubating with DTT to the final concentration of 5 mM in the dark for another 15 minutes.[160] Proteins were purified and pelleted with methanol/chloroform precipitation and digested with sequencing grade modified trypsin (Promega) at 37 ºC for 16 hours (enzyme:substrate ratio of ~1:100) in a buffer containing 5% acetonitrile (ACN, Sigma-Aldrich), 1.6 M urea (Sigma-Aldrich), and 50 mM HEPES (pH=8.2). The digestion was quenched by adding trifluoroacetic acid (TFA, Sigma-Aldrich) to the final concentration of 0.4%, and the pH was checked to be lower than ~2. The peptides were desalted using a Sep-Pak Vac tC18 cartridge (Waters) and dried in a vacuum concentrator.

2.2.4 Enrichment of cell-surface glycopeptides

Glycopeptides tagged with biotin were enriched with high-capacity NeutrAvidin agarose resin (Thermo Scientific) according to the manufacturer's protocol. The enriched peptides were eluted three times with 200 µL 8 M, pH=1.5 guanidine hydrochloride (Promega) at 56 ºC for 2 minutes each. The eluates were pooled, desalted, and dried in a

vacuum concentrator overnight. Glycopeptide deglycosylation was performed with 3 units of PNGase F (Sigma-Aldrich) in 40 μL of 40 mM ammonium bicarbonate (pH=9, Sigma-Aldrich) in heavy-oxygen water ($H_2{}^{18}O$, Isoflex) at 37 ºC for 3 hours with shaking. The reaction was quenched with formic acid (FA, Fisher Scientific) to the final concentration of 1%. The peptides were desalted with StageTips and eluted into three fractions with 20%, 50%, and 80% ACN containing 1% acetic acid (Sigma-Aldrich).[161] The eluates were dried again in a vacuum concentrator.

2.2.5 LC-MS/MS analysis and database searching

The peptides were dissolved in a solution containing 5% ACN and 4% FA and separated by a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific) with a microcapillary column containing C18 beads (Magic C18AQ, 3 μm, 200 Aº, 75 μm*16 cm) packed in-house. A total of ~1 μg of peptides was loaded into the column by a Dionex WPS-3000TPL RS autosampler (Thermostatted Pulled Loop Rapid Separation Nano/Capillary Autosampler). Peptides were separated by reversed-phase liquid chromatography (LC) using an UltiMate 3000 binary pump with 80-minute gradients of 4-25%, 10-38%, and 15-50% ACN containing 0.125% FA for the three fractions, respectively. The LC is coupled to an LTQ Orbitrap Elite Hybrid Mass Spectrometer (Thermo Scientific) with Xcalibur software (version 3.0.63). MS/MS analysis was performed with a data-dependent Top20 method.[104, 162] For each cycle, a full MS scan (resolution: 60,000) in the Orbitrap with 1 million automatic gain control (AGC) target was followed by up to 20 MS/MS in the LTQ for the most intense ions. Selected ions were excluded from further sequencing for 90 seconds. Ions with singly or unassigned charge

were not sequenced. Maximum ion accumulation times were 1,000 ms for each full MS scan and 50 ms for MS/MS scans.

Raw MS files were analyzed by MaxQuant (version 1.6.2.3).[163] MS spectra were searched against the human proteome database downloaded from UniProt containing common contaminants using the integrated Andromeda search engine.[164] Glycopeptides were searched separately for the identification experiments. All default parameters were left unchanged, except adding variable modification for asparagine deamidation (+2.9883 Da) for glycosylation site determination and 3 maximum missed cleavages. In the quantification experiments, all raw files were searched together with the three files from the same experiment grouped together. Label-free quantification was also enabled with the LFQ min ratio count of 1, the match-between-runs option was enabled, asparagine deamidation modification was used in protein quantification, and the iBAQ option was enabled. The false discovery rates (FDR) were kept at 0.01 at the peptide spectrum match, protein, and site decoy fraction levels.

2.2.6 Bioinformatic analysis

Data analyses were performed with Perseus[165] and Excel. Glycopeptides were filtered to only contain the canonical sequences (N-X-S/T) and non-canonical sequence (N-X-C), where X is any amino acid except proline, for N-linked glycosylation. Human membrane protein information was extracted from UniProt database: SL-9905 (single-pass type I membrane proteins), SL-9906 (single-pass type II membrane proteins), SL-9907 (single-pass type III membrane protein), SL-9908 (single-pass type IV membrane proteins), SL-9909 (multi-pass membrane proteins), and SL-9903 (peripheral membrane

proteins). For those whose membrane information is not available, further sequence analyses were performed using Phobius (phobius.sbc.su.se), which predicts the transmembrane and signal peptide regions of proteins.[166] SecretomeP (cbs.dtu.dk/services/SecretomeP) was used to further predict protein secretion through non-classical secretory pathways with the cutoff score of 0.6.[167]

Gene ontology-based enrichment analysis was performed on Gene Ontology Consortium website (http://www.geneontology.org). Fisher's exact test was used to calculate the P values and only those with $P<0.05$ were included. Residue solvent accessibility and structure were predicted using NetsurfP (version 1.1).[168] The structure (helix, strand, or coil) with the highest probability among the others was assigned a structure for the residue.

For the quantification experiments, the glycopeptide LFQ intensity was extracted from the peptides.txt table and limited to only glycopeptides (with the deamidation sites). The glycoprotein intensity was calculated by summing the peptide LFQ intensities together. The final LFQ intensity for each cell line was an average of the intensities between two biological duplicate experiments. iBAQ was used to estimate the absolute protein abundance ranking.[169] iBAQ intensity was calculated manually by dividing the summed glycopeptide intensity by the number of theoretical tryptic peptides, which was extracted from the proteinGroups.txt table. Shannon's entropy was calculated the same way as the previous report.[170] The entropy was calculated using the formula $H(S)=-\sum_t p(S_t)\ln p(S_t)$, where t is the protein index, $p(S_t)$ is the ratio of the protein LFQ intensity to the summed LFQ intensity of the protein. Because of the missing values, 1/8 was added to the raw LFQ intensity so that the natural log can be calculated.

Unsupervised hierarchical clustering was performed with Perseus. Euclidean distance was used to calculate the distance. Protein intensity was converted to a $\log_2$ scale before further analysis. For the heat map generation, missing values were imputed with a normal distribution (width=0.3, shift=1.8) before Z-score transformation.[165] ANOVA was performed with Perseus (S_0=0.5, Benjamini-Hochberg FDR=0.05). The proteins were filtered so that they contain at least 8 out of 16 valid values in order to reduce the effect of quantifying low-abundance surface glycoproteins.[171]

Protein interaction network was processed using Cytoscape.[172] Interaction information was extracted from STRING database with the high confidence cutoff (score=0.7).[173] Pathway analysis was performed with the Cluego plugin of Cytoscape.[174] All default parameters were used.

2.3 Results and discussion

2.3.1 Global analysis of cell-surface glycoproteins in human cells

Sugar analogs containing a biologically inert but chemically functional group have been proven to be powerful labeling reagents for glycoproteomic studies.[175-177] We previously demonstrated that labeling with Ac_4GalNAz resulted in the highest coverage of cell-surface N-glycoproteins compared with N-azidoacetylglucosamine-tetraacetylated (Ac_4GlcNAz) and N-azidoacetylmannosamine-tetraacetylated (Ac_4ManNAz)[47], and thus Ac_4GalNAz is used in this study. Cells incorporated GalNAz into the glycans on glycoproteins, including those located on the cell surface. These surface glycoproteins were selectively tagged through the strain-promoted, copper-free click chemistry reaction between the azido group and dibenzocyclooctyne (DBCO)-biotin in flask under very mild

conditions.[178] They are then enriched with NeutrAvidin beads at the peptide level, deglycosylated with PNGase F in $H_2{}^{18}O$ to generate a common tag, and analyzed with LC-MS/MS (Figure 2.1A).

Using this approach, a total of 1,047 glycoproteins and 2,172 N-glycosylation sites were identified with an average of 354 glycoproteins and 683 sites from each cell type (Figure 2.1B). The average posterior error probability of the peptide identification is 0.005. Compared with the previously reported results, including total glycoproteomic analysis[96, 179-180], we identified 349 new glycosylation sites. Protein occurrence analysis showed that the number of glycoproteins identified in only one cell line is the highest, and as the number of cell types increases, the occurrence decreases (Figure 2.1C). Biological duplicate experiments revealed that, on average, over 70% of glycoproteins and glycosylation sites were identified in both experiments, showing high reproducibility of the approach (Figure 2.2). The conditions for tagging cell surface glycoproteins are mild, which do not stimulate cellular response, or harm the cells because copper or oxidizing reagents are not employed. This allows site-specific quantification and dynamic studies of cell-surface glycoproteins.[47-48, 181] Even though we globally analyzed surface glycoproteins in cultured cells. Spiciarich et al. recently employed metabolic labeling with ManNAz for the identification of sialoglycoproteins from the proteomes of human prostate cancer and normal tissues. The authors identified 972 proteins from both samples with about 50% of the proteins localized on the plasma membrane.[43] This is very promising and shows the efficiency of metabolic labeling for tissue samples. Sugar analogs can also be fed to animals, such as zebrafish[182] and mouse[183]. Therefore, this method is applicable to study surface glycoproteins in tissue samples and model animals.

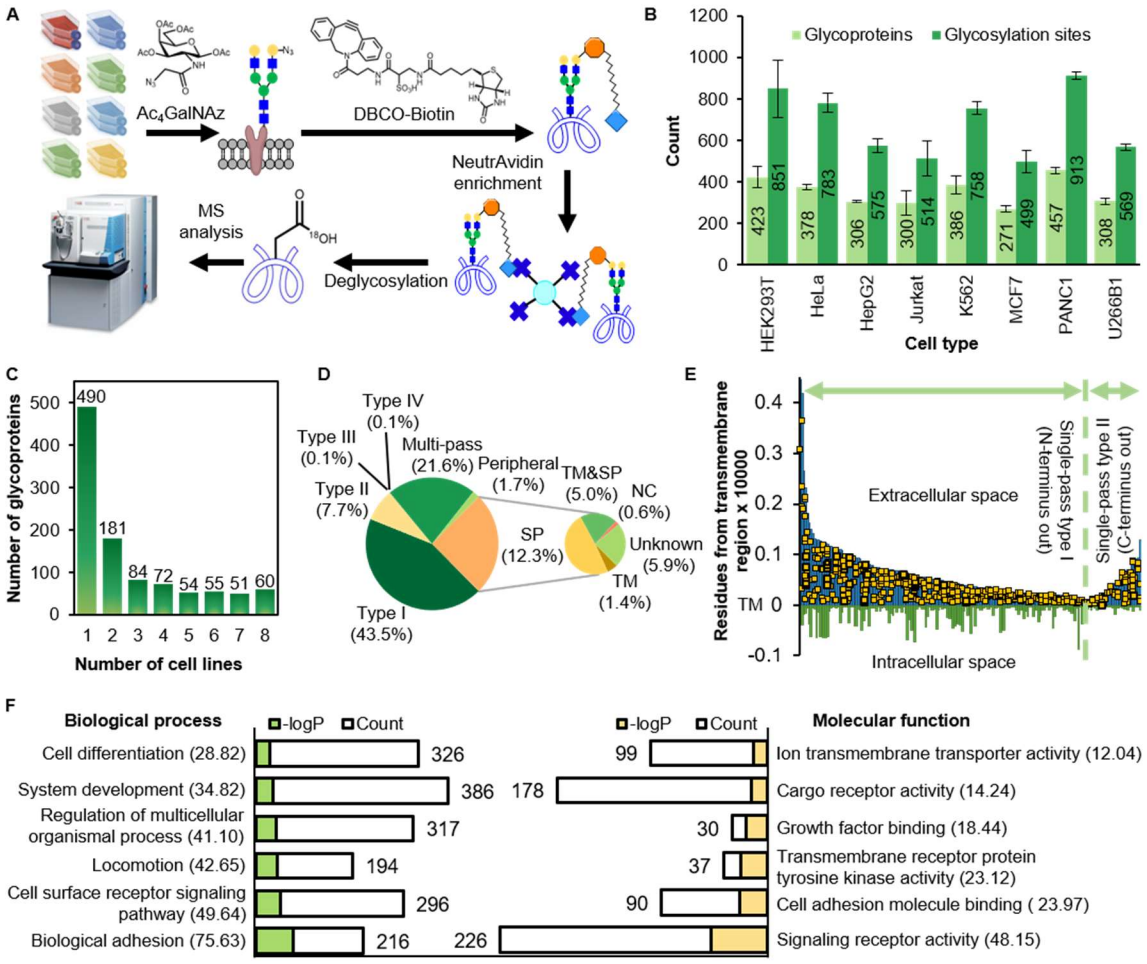

Figure 2.1. Overview of global and site-specific analysis of cell-surface glycoproteins from eight popular types of human cells. (A) A diagram showing the experimental procedure. (B) Numbers of cell-surface glycoproteins and glycosylation sites identified from each cell type. The error bars represent one standard deviation from two biological duplicate experiments. (C) Number of cell-surface glycoproteins identified from multiple cell types. (D) Types of the identified surface glycoproteins. Types I-IV for single-pass membrane protein types I-IV, TM for transmembrane region(s), SP for signal peptide, and NC if the protein enters the non-classical secretory pathway. (E) Single-pass types I and II surface glycoproteins from K562 cells are aligned against the transmembrane domain (TM). Yellow dots represent the identified glycosylation sites. (F) Protein clustering of all identified surface glycoproteins based on biological process and molecular function. The number in the parentheses show the -logP values.

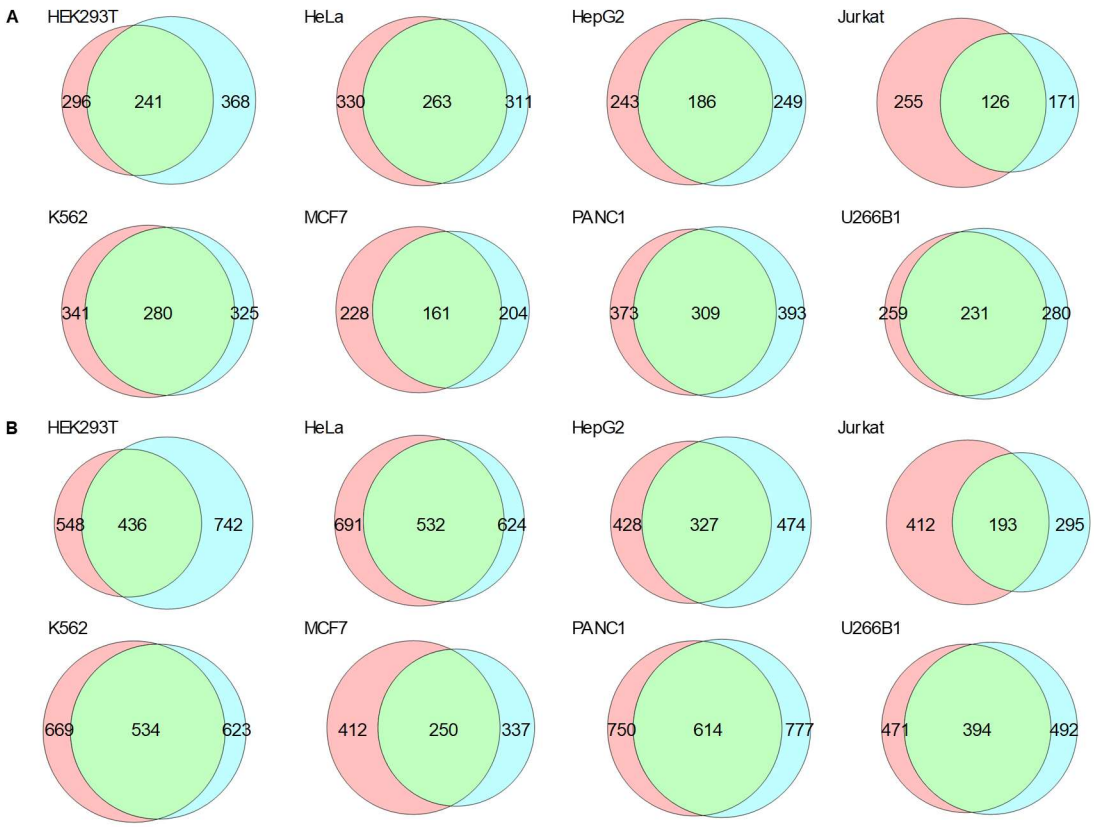

Figure 2.2. Reproducibility of the identification of cell-surface glycoproteins from biological duplicate experiments. (A) Overlap of cell-surface glycoproteins identified from each cell line. (B) Overlap of cell-surface glycosylation sites identified from each cell line.

2.3.2 Classification of identified surface glycoproteins

Among the identified surface glycoproteins (1,047), over 800 belong to membrane proteins ($P=4.51 \times 10^{-128}$) according to gene ontology analysis based on cellular component. Compared with UniProt subcellular location information, on average, 75% of the glycoproteins identified from all cell types are classified into single-pass types I-IV, multi-pass, and peripheral membrane proteins (Figure 2.1D). For those without membrane protein classification information available, Phobius was employed to predict if they have either a transmembrane (TM) region and/or a signal peptide (SP).[166] SecretomeP 2.0 was

also used to predict whether they may be secreted through the non-classical secretory (NC) pathways and then located on the cell surface.[167] Eventually, only 5.9% of the identified glycoproteins were left without information supporting their localization on the cell surface. These proteins may not be annotated or discovered at the cell surface yet. Another possibility is that their identifications could be due to non-specific binding of some peptides during the enrichment. In spite of that, the approach specifically targets proteins on the extracellular side because the cells were tagged with DBCO directly in the flask without affecting the cell integrity. These sites on types I and II single-pass membrane proteins from K562 cells are displayed as yellow dots in Figure 2.1E, with the X-axis representing the transmembrane region and the Y-axis showing how far away the glycosylation sites are from the transmembrane region. No glycosylation sites inside the cells were identified.

The biological functions of the identified proteins from GO enrichment analysis correspond very well with the known functions of cell-surface glycoproteins (Figure 2.1F), including biological adhesion ($P=2.36\times10^{-76}$), cell surface receptor signaling pathway ($P=2.30\times10^{-50}$), locomotion ($P=2.24\times10^{-43}$), and cell communication ($P=2.19\times10^{-30}$). Proteins with binding activities, such as growth factor binding ($P=3.64\times10^{-19}$), collagen binding (for cell-matrix adhesion, $P=5.23\times10^{-10}$), and calcium, copper, and chloride ion bindings were also enriched. Many membrane enzymes with signaling receptor activity ($P=7.06\times10^{-49}$) were identified, including those involved in the regulation of protein kinase B signaling, phosphatidylinositol 3-kinase signaling, and MAP kinase activity. Other less famous functions of cell surface glycoproteins, such as the regulation of cell size ($P=7.09\times10^{-15}$), the regulation of body fluid level ($P=3.24\times10^{-7}$), ossification ($P=5.86\times10^{-7}$), and learning or memory ($P=5.31\times10^{-5}$) were also found. Interestingly, we identified

proteins involved in DNA-binding transcription factor activity, such as vasculin and alpha-enolase, which were reported to localize in both the plasma membrane and the nucleus.[184-185] Therefore, some proteins without the membrane information could still be localized at the surface.

2.3.3 Distribution and occurrence of N-linked glycosylation sites on surface glycoproteins

Benefiting from the deglycosylation reaction with PNGase F in $H_2{}^{18}O$ to generate a mass tag on glycosylation sites, we confidently localized the sites with an average probability of 0.97. N-glycosylation has a well-known N-!P-S/T canonical sequences, and here, the N-!P-C motif was also included because of the previous reports.[186-187] The majority of the identified proteins contain 1 or 2 sites (Figure 2.3A), and some have many more such as 25 sites identified on LRP1 (prolow-density lipoprotein receptor-related protein 1). The protein is made of over 4,000 amino acid residues and has 75 N-!P-S/T/C motifs. The total number of glycosylation sites does not depend on the number of glycosylation motifs (Figure 2.4A). In our dataset, 39% of all glycosylation motifs were glycosylated. For example, both neural cell adhesion molecule 1 (L1CAM) and lymphocyte antigen 75 (LY75) have 22 glycosylation motifs but 19 and 2 glycosylation sites were identified, respectively. More than half of the identified glycosylation sites have the N-!P-T sequence despite the similar occurrence of the N-!P-S and N-!P-T motifs (Figure 2.3B), agreeing with the previous results.[186] Only 2% of the sites have the N-!P-C sequence when this motif occurs at 10% of the total motifs. As the protein length increases, the number of motifs also increases with an acceptable R^2 value of 0.71. However, this does not translate into a higher number of glycosylation sites (R^2=0.02) (Figure 2.4B).

Because the surface glycoproteome is relatively small, the number of glycosylation sites from a large-scale study by Xiao et al.,[96] which identified one of the largest experimental datasets for N-glycosylation, was also evaluated and the result is the same (Figure 2.4C).

The relatively large and hydrophilic N-glycans play critically important roles in the regulation of protein folding and structures. Accordingly, the location of the asparagine residues in each protein was predicted using NetsurfP algorithm.[168] It was found that the asparagine residues are, as expected, exposed to the solvent with only a small fraction buried inside proteins (0.79 and 0.21, respectively) (Figure 2.3C). A closer look into these residues shows that those containing a glycosylation motif are even more exposed to the solvent than those without one (0.87 and 0.76, respectively). This observation also applies to the glycosylated residues. We also determined the solvent accessibility for each of the N-!P-S/T/C sequons. The fractions of N-!P-S and N-!P-T exposed to the solvent are very similar (0.87 and 0.89), compared to 0.80 for N-!P-C. Overall, the fractions of the solvent accessibility for the glycosylated sites are similar to those for the motifs (Figure 2.4D).

We then investigated the predicted structure of these sites in proteins using the results from NetsurfP. Because of its structural role, N-glycosylation sites are usually located on loops and turns.[188] The majority of the asparagine residues are on coils (loops and turns), and fewer on helices and strands, respectively (Figure 2.4E). The fractions of the identified sites and the motifs are very similar for each structure. For the helix structure, the fractions of identified sites and residues with the glycosylation motifs are half of those without the motif and the total, but for the strand structure they are twice, even though both are ordered structures. Surprisingly, a higher fraction of the N-!P-C motifs is located on coils than the N-!P-S and N-!P-T motifs, but a lower fraction is actually glycosylated

(Figure 2.3D). Glycosylation with the N-!P-C motif is, however, more preferred at the helices even though the fraction of helical N-!P-C is smaller than that of coiled N-!P-C. The prediction was compared with the solved structures of some proteins and the results agree very well.

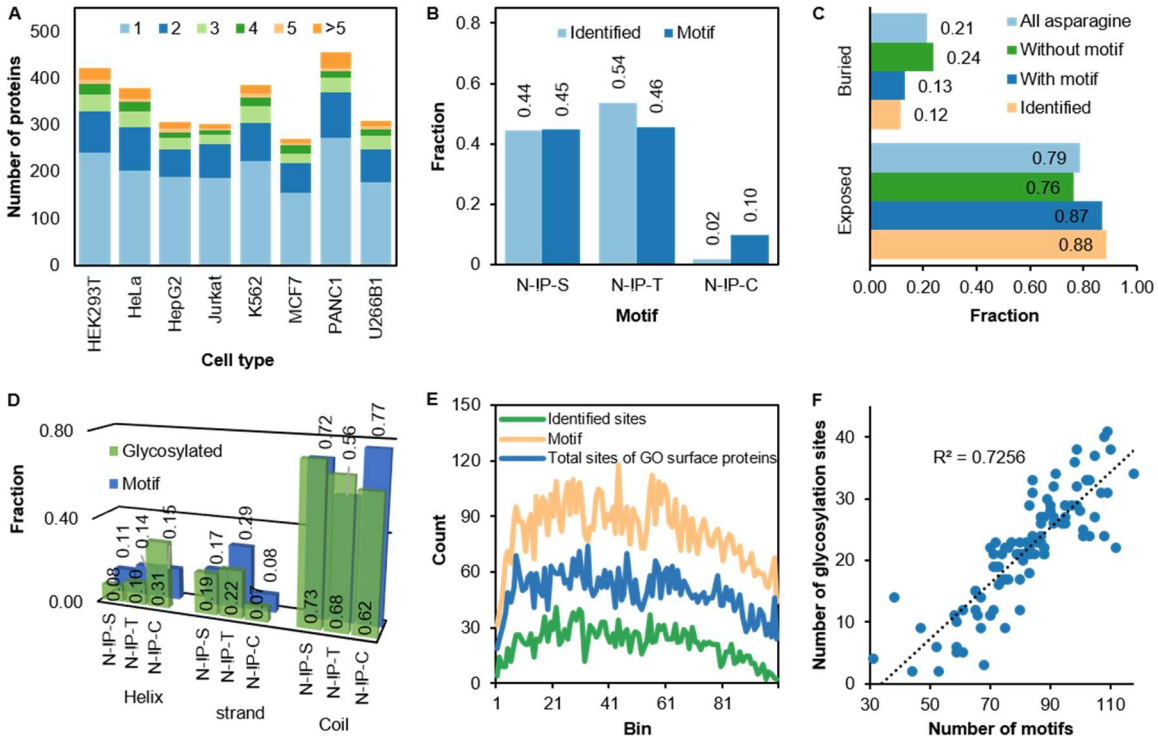

Figure 2.3. Site-specific analysis shows the distribution and occurrence of N-glycosylation sites and the motifs of cell-surface glycoproteins. (A) Number of protein glycosylation sites from each cell type. (B) Relative abundance of the glycosylation motifs and the identified glycosylation sites. (C) Solvent accessibility prediction of all asparagine residues. (D) Distributions of the predicted structure at each N-glycosylation motif and glycosylation site. (E) The occurrence of the identified glycosylated sites, the glycosylation motifs, and total sites from GO surface proteins extracted from UniProt when each protein length is divided into 100 bins. (F) The correlation between the number of identified glycosylation sites and the number of the motifs in each bin.

To further investigate the occurrence of glycosylation sites along the protein length, each protein length was divided into 100 bins, and the number of glycosylation sites in each bin was counted (Figure 2.3E). Generally, fewer identified glycosylation sites on the

N- and C-termini were observed. For the N-termini, it may be due to the inability of the accepting site near the signal peptide sequence to reach the active site of oligosaccharyltransferase (OST) before the signal peptide is cleaved.[189] For the C-termini, it may be because of the steric hindrance from the secondary structures formed in the near-complete nascent polypeptide, or the inefficiency of N-linked glycosylation when the polypeptide translation is terminated before the glycosylation accepting residue reaches the active site of OST.[190-191] Not only is the frequency of glycosylation sites lower at both termini, but also the number of the glycosylation motifs themselves are different in each bin and lower at both termini. A similar pattern of the glycosylation motifs and the glycosylation sites was observed (Figure 2.3E). Even though we showed earlier in Figure S2B that the number of glycosylation sites does not depend on the protein length, the correlation between the number of glycosylation sites and the number of the motifs is reasonably high with R^2 of 0.73 when their position along the protein length is considered (Figure 2.3F). We compared this with proteins annotated with GO surface proteins (Figure 2.3E) and those from Xiao et al.,[96] and also found a similar correlation (Figure 2.4F and G).

Some glycopeptides might be present but were not identified by the mass spectrometer. For example, tryptic glycopeptides that are too short or too long may not be identified. Other post-translational modifications on the same glycopeptide may also hinder its identification. Therefore, some peptides with a particular glycosylation site could be missing from the analysis. Nevertheless, our approach is very sensitive as 87 proteins containing only one glycosylation motif (maximally one glycosylation site) were identified. It should also be noted that these sites are pooled from all cell types, so some

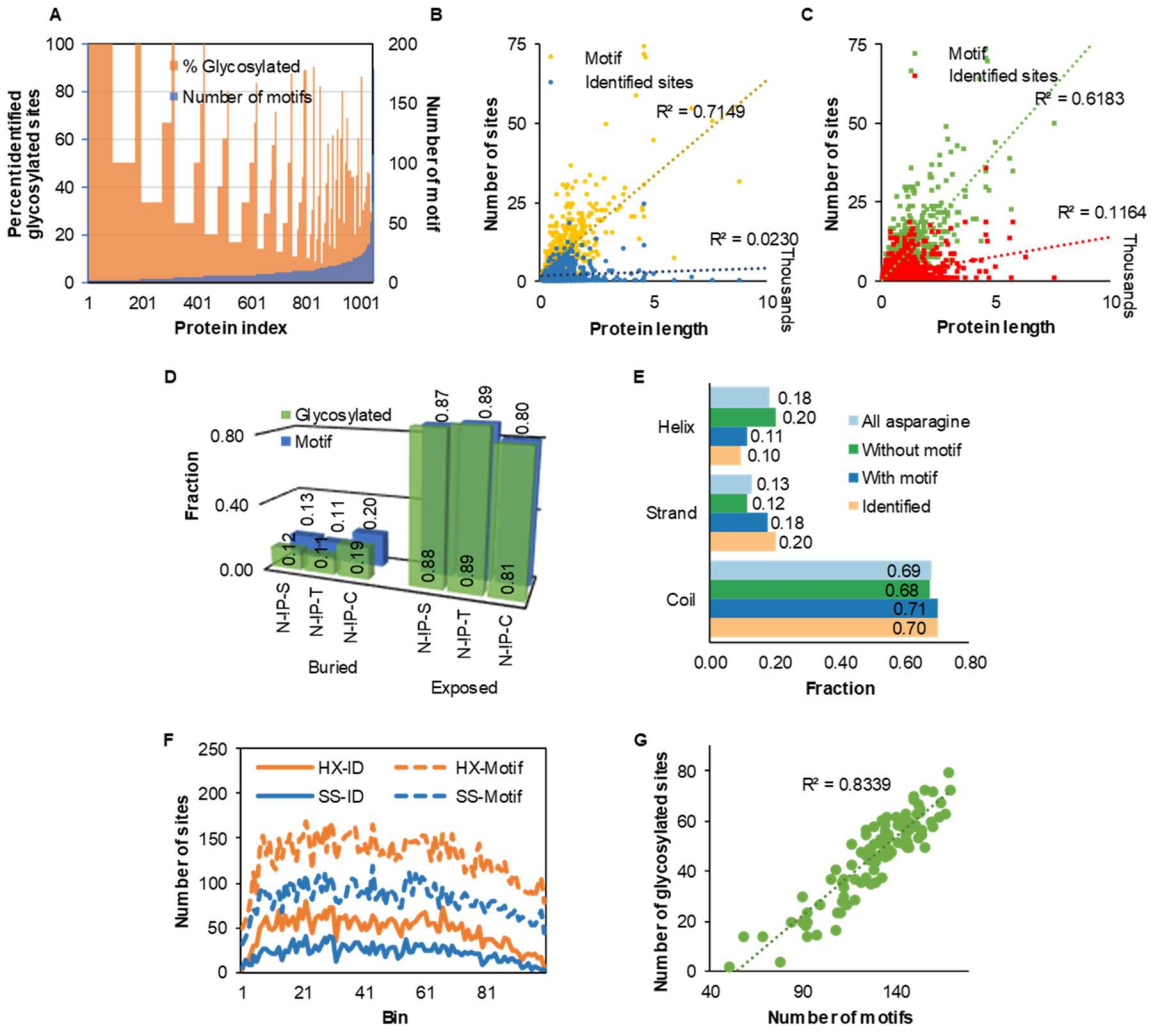

Figure 2.4. Site-specific analysis of surface protein glycosylation. (A) The number of glycosylation motifs, including all N-!P-S/T/C motifs, of all identified proteins and their percentages of the identified glycosylation sites. (B) The correlation of the protein length and the numbers of glycosylation motifs and identified sites from this work. (C) The correlation of the protein length and the numbers of glycosylation motifs and identified sites from Xiao et al. (Ref. 31). (D) Distribution of the solvent accessibility at each N-linked glycosylation motif. (E) Structure prediction of all asparagine residues from the identified surface glycoproteins. (F) Glycosylation motifs and site distribution when each protein length is divided into 100 bins from this work (SS) compared to those identified from Xiao et al., 2018 (G) The correlation between the number of glycosylated sites and the number of motifs in each bin for the data set from Xiao et al., 2018.

sites may not be found in a specific cell type. The coverage of surface glycoproteins may be further improved. For example, previously, different sugar analogs were used for

labeling different subsets of surface glycoproteins such as ManNAz for sialoglycoproteins,[47, 175, 182, 192] and thus a combination of sugar analogs may further increase the coverage. Many surface glycoproteins are also present in low abundance. Therefore, more cells used in the experiments will allow us to more effectively identify low-abundance glycoproteins on the surface.

2.3.4 Label-free quantification of cell-surface glycoproteins

The identification of cell-surface glycoproteins from eight types of cells does not provide quantitative information. Therefore, we quantified these cell-surface glycoproteins between cell types with label-free quantification (LFQ) using MaxQuant.[163] Based on the LFQ results, 784 surface glycoproteins were quantified. The reproducibility of LFQ quantification was first evaluated between two biological duplicate experiments because the abundance of cell-surface glycoproteins is intrinsically low compared to intracellular proteins, which may affect the quantification precision between runs. The $\log_2$ intensities from both runs are in an acceptable agreement with an average R^2 of 0.81 (Figure 2.5A). Proteins with zero intensities were excluded because the log values cannot be determined. Conversely, a comparison between different cell types showed a weaker correlation with an average R^2 of 0.45 (Figure 2.5B and Figure 2.6A). These results differ from the previous proteome experiments where the correlations from biological replicate experiments are very similar (0.83) while the correlation between the results from different cell lines were much higher (0.74).[171] These results clearly demonstrate that the method is reasonably reproducible, and compared to intracellular proteins, surface glycoproteins are much more unique to the cell type. The identified proteins are categorized into globally expressed and cell-specific surface glycoproteins. Here, we define globally expressed surface

glycoproteins as those quantified across all cell lines without any missing values while the cell-specific ones are quantified in a single cell type. Shannon's entropy was also applied to the quantified proteins to show the expression of each protein across the cell types.[170] Generally, a higher entropy value means a more uniform expression across cell types or that the protein is quantified with valid abundance values in more cell types, while a lower entropy indicates differential expression of the protein or it is expressed in particular cell types (Figure 2.6B and Figure 2.6C). We quantified 104 globally expressed surface glycoproteins (Figure 2.5C), with Shannon's entropy in the range of 0.71-2.04. These proteins are involved in biological processes such as adhesion, cell-surface receptor signaling pathway, and cell migration. Four-hundred and eighty-three proteins were found in 2-7 cell lines. Surprisingly, a total of 197 proteins were cell-specific, corresponding to an average of 25 proteins or 7% of all quantified proteins per cell type (examples are included in Table 2.1). The proteins have the Shannon's entropy values in the range of 1.76×10^{-4} – 8.26×10^{-8}. The number is different from that in Figure 2.1C because the match-between-runs option was enabled during the quantification, and some proteins were assigned to more than one cell line. Protein clustering analysis did not yield much valuable information about these proteins for a specific cell type due to the low number of proteins. Nevertheless, the functions enriched from HEK293T-specific proteins are consistent with relevant biological processes to kidney cells including kidney development and nervous system development.

The absolute abundance of these surface glycoproteins was also estimated using intensity-based absolute quantification (iBAQ). The $\log_2$-transformed intensity plots show a normal S-shaped distribution and varies within five orders of magnitude (Figure 2.6D).

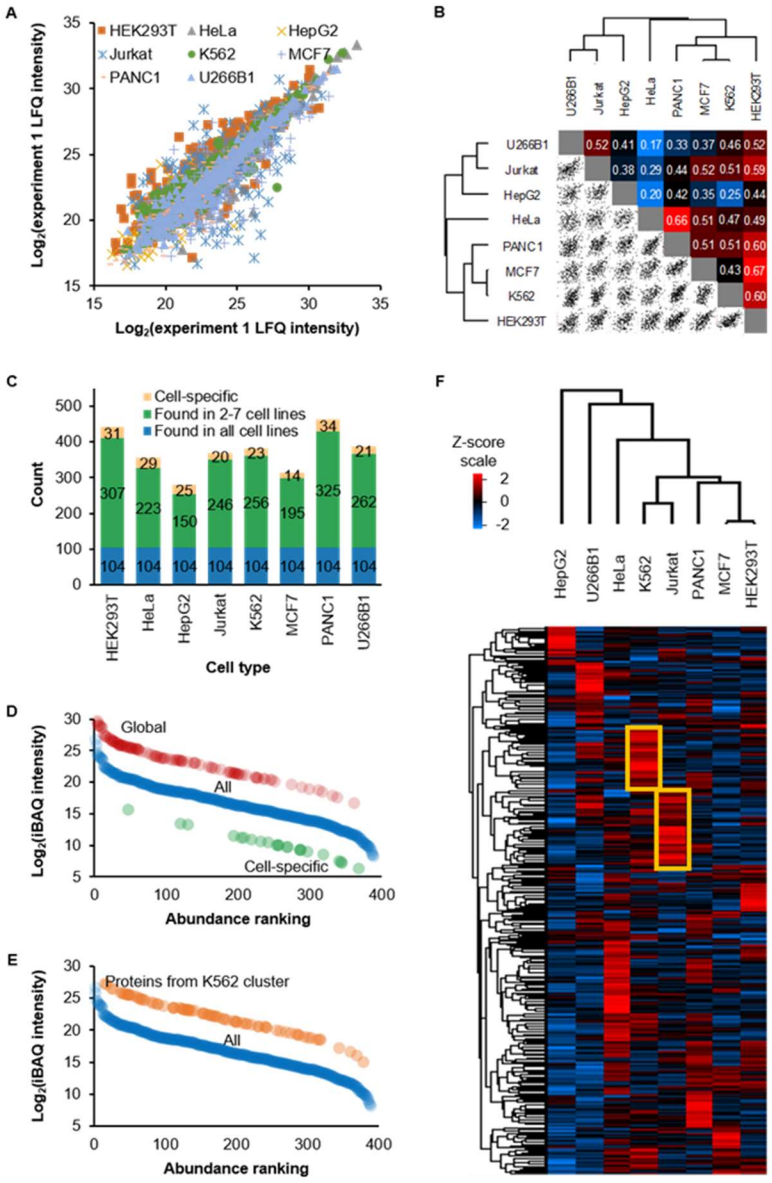

Figure 2.5. Label-free quantification of surface glycoproteins. (A) Correlation of LFQ intensity from biological duplicate experiments of each cell type. (B) Correlation and hierarchical clustering of LFQ intensity between cell lines. R^2 values are displayed in the figure. The log_2-transformed average LFQ intensity of the two duplicate experiments were used when calculating the correlation. (C) The numbers of proteins that are cell-specific, are in 2-7 cell lines, and are globally expressed from each cell line from the quantification with LFQ. (D) Ranking of absolute protein abundance by iBAQ from K562 cells. Data points from global and cell-specific proteins were shifted by +5 and -5, respectively, to clearly show their positions. (E) A similar plot as Figure 2.5D. However, the iBAQ intensity of proteins in the yellow square of K562 cells of Figure 2.5F are plotted as orange circles against total proteins from K562. Data points were shifted by +5 to more clearly show their positions. (F) A Z-score transformed heat map of log_2 LFQ protein intensity showing relative protein expression of surface glycoproteins. Missing values were imputed with a normal distribution (width=0.3, shift=1.8).

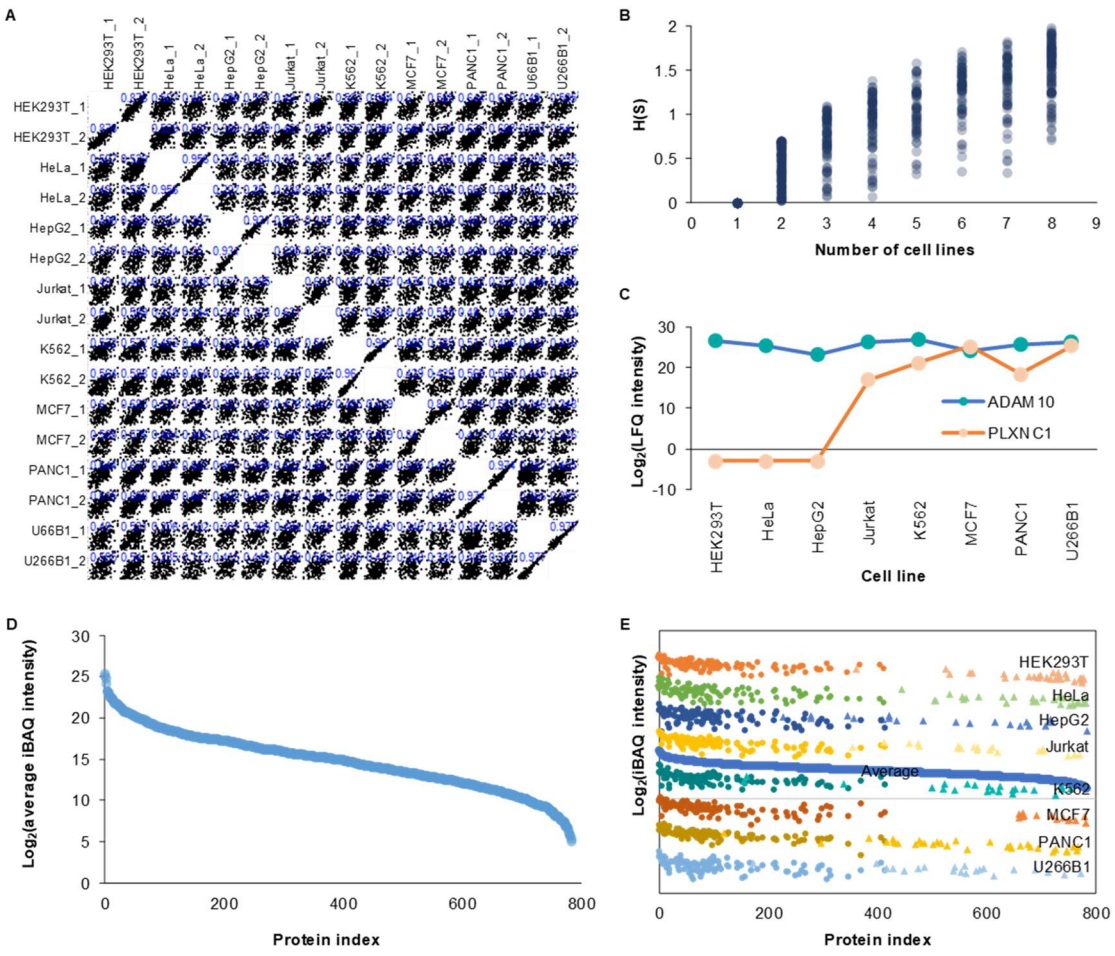

Figure 2.6. Correlation of label-free quantification of cell-surface glycoproteins. (A) The correlation of log2-transformed LFQ intensity from biological duplicate experiments or between cell lines. Pearson correlation is shown on the plot. Zero values were omitted. (B) Distribution of Shannon's entropy (H(S)) of proteins found in different number of cell lines. (C) Example of two proteins with different Shannon's entropies and their LFQ intensities across the cell lines after missing value imputation. ADAM10 has the entropy of 1.89 while PLXNC1 has the entropy of 0.83. (D) Average iBAQ intensity of proteins from all cell lines. (E) Abundance ranking of cell-specific and globally expressed proteins from all cell lines, similar to Figure 2.5D. The log2-tranformed intensities for each cell line were offset by different values so the plot can visibly be seen. The data points from the same cell line are shown in the same color. Those that are globally expressed are shown in circles while those specific to the cell line are shown in triangles.

Surface glycoproteins from each cell type are categorized as high or low abundance if their intensities fall within the fourth or the first quartiles, respectively. The 104 glycoproteins

that are globally expressed have higher absolute abundance while the cell-specific ones are present in relatively low abundance. The estimated absolute abundance of 77% globally-expressed proteins fall into the third and fourth quartiles while 83% of those K562-specific proteins are in the first and second quartiles (Figure 2.5D). There are also 12 high-abundance glycoproteins that were quantified in all cell types. These high-abundance proteins are mostly transporters or adhesion molecules, including integrin beta-1 (ITGB1), basigin (BSG), and neutral amino acid transporter (SLC1A5). No common low-abundance proteins were found in all cell types.

We then performed a hierarchical clustering analysis with the LFQ intensities to compare the surface glycoprotein expression between cell types. The heat map shows differential expression of proteins across the cell types (Figure 2.5F). ANOVA test was also performed and the expression of over half of the proteins were statistically different in at least one cell type (65% without missing value imputation and 71% with the imputation). Because of the intrinsic low abundance of cell-surface glycoproteins, it is expected that many values would not pass the cutoff for statistical significance, thus the percentages may be underestimated. The first, closest expression pattern from hierarchical clustering analysis arose from MCF7 and HEK293T while the second was from Jurkat and K562 cells. There are also groups of glycoproteins that are more abundant in a specific cell type than in the others. These glycoproteins are responsible for particular processes regarding the cell type. For example, those from K562 cells (in the yellow frame of Figure 2.5F) are involved in the processes such as neutrophil degranulation, while those from Jurkat include positive regulation of T cell activation. Interestingly, the estimated absolute abundances of these proteins span a wide range in that cell line, as shown in Figure 2.5E

for K562 cells, but their relative abundances are higher than the corresponding ones in the other cell lines if expressed.

Overall, the current results demonstrate that the majority of commonly identified surface glycoproteins across the cell types usually have higher absolute abundance. Most of these proteins are necessary for normal cellular functions and cell survival. In biomedical research, cell lines are often chosen based on specific protein expression on the surface that are appropriate for the experiment. We observed, however, that only a small portion of quantified surface glycoproteins from each cell type are cell-specific, and their absolute abundances are quite low. While we cannot disregard that these cell-specific proteins define the cell type, the difference among the cell types further arises from their differential expression. These cell-specific proteins cannot be excluded in other cell types. If the detection limit is lower and some of these proteins may be detected in another different cell type, but the relative abundance is still different. In this case, quantification information will be more meaningful. In addition, proteins responsible for specific functions of the cell type are expressed at any absolute abundance, but the relative abundances of these proteins are normally higher than in other cell types. It also illustrates that not only the expression but also the abundance of the protein of interest should be taken into consideration when choosing a cell type for a specific experiment.

Table 2.1. Examples of cell-specific surface glycoproteins. The LFQ intensity is shown for each protein in each cell line. See the method section for the calculation of Shannon's entropy.

Major cell type	UniProt ID	Name	Annotation	Site	Shannon's entropy	HEK293T	HeLa	HepG2	Jurkat	K562	MCF7	PANC1	U266B1
HEK293T	P21802	FGFR2	Fibroblast growth factor receptor 2	N318	4.80E-06	3.3E+6	0	0	0	0	0	0	0
	P23471	PTPRZ1	Receptor-type tyrosine-protein phosphatase zeta	N105	2.64E-05	5.4E+5	0	0	0	0	0	0	0
HeLa	P42262	GRIA2	Glutamate receptor 2	N256	1.85E-06	0	9.0E+6	0	0	0	0	0	0
	P55285	CDH6	Cadherin-6	N399, N437	2.74E-05	0	5.2E+5	0	0	0	0	0	0
HepG2	P05534	HLA-A	HLA class I histocompatibility antigen, A-24 alpha chain	N110	6.43E-06	0	0	2.4E+6	0	0	0	0	0
	P07307	ASGR2	Asialoglycoprotein receptor 2	N102, N170	3.16E-07	0	0	5.8E+7	0	0	0	0	0
Jurkat	Q6GTX8	LAIR1	Leukocyte-associated immunoglobulin-like receptor 1	N69	4.18E-06	0	0	0	3.8E+6	0	0	0	0
	P06127	CD5	T-cell surface glycoprotein CD5	N116, N241	5.02E-07	0	0	0	3.6E+7	0	0	0	0

Table 2.1. Continued

U266B1		PANC1		MCF7		K562							
Q96FE7	Q02223	Q04900	Q02297	Q96NY8	Q8IZF3	Q6UWB1	P04629						
PIK3IP1	TNFRSF17	CD164	NRG1	NECTIN4	ADGRF4	IL27RA	NTRK1						
Phosphoinositide-3-kinase—interacting protein 1		Sialomucin core protein 24	Pro-neuregulin-1, membrane-bound isoform	Nectin-4	Adhesion G protein-coupled receptor F4	Interleukin-27 receptor subunit alpha	High affinity nerve growth factor receptor						
N66	N42	N146	N164	N281	N61, N250, N686	N76	N67, N188, N262						
2.32E-06	4.17E-06	6.31E-05	7.92E-06	4.66E-06	3.52E-05	1.26E-05	1.59E-07						
0	0	0	0	0	0	0	0						
0	0	0	0	0	0	0	0						
0	0	0	0	0	0	0	0						
0	0	0	0	0	0	0	0						
0	0	0	0	0	0	1.2E+6	1.2E+8						
0	0	0	0	3.4E+6	4.0E+5	0	0						
0	0	2.1E+5	1.9E+6	0	0	0	0						
7.1E+6	3.8E+6	0	0	0	0	0	0						

2.3.5 Classification of cell types by cluster of differentiation

As of December 2018, 451 cluster of differentiation (CD) molecules were listed on UniProt (https://www.uniprot.org/docs/cdlist), among of which 396 are proteins. Here, we identified a total of 155 CD proteins with an average of 76 proteins per cell type. With LFQ intensity and the MBR option enabled, we quantified 148 CD proteins. Twenty-nine CDs were globally expressed in all cell lines, and most of these proteins function in the response to stimulus process. On average, twenty-six of the CD proteins were cell-specific, such as CD7 protein that was identified and quantified only in Jurkat cells.

Similar to the surface glycoproteome results, we performed hierarchical clustering analysis and observed differential expression with specific groups of proteins being more abundant in specific cell types (Figure 2.7A). For example, proteins highlighted in the yellow box in the figure are highly expressed in Jurkat cells. These are proteins involved in specific T-cell processes, such as CD3D, CD5, and CD6. Despite the use of CDs as markers for a specific cell type, we noticed that some CD proteins can be expressed in other types of cells. For example, CD28 was detected in both Jurkat and U266B1 (Figure 2.7B), and the relative abundance of CD28 in U266B1 is even greater than that in Jurkat cells. Similarly, CD6 was also detected in both cell lines. The estimated absolute abundance of CD28 in both types of cells is high, and that of CD6 is high in Jurkat, but is low in U266B1 (Figure 2.8). Another example is CD56, a phenotype marker for natural killer cells, was also found in HEK293T, HeLa, Jurkat, and U266B1 at different expression levels (Figure 2.7B). Previous studies found that CD56 could be expressed in different cell types, including T cells, dendritic cells, and monocytes.[193] There are also some published cases where a specific CD molecule was discovered in other different cell lines.[194-195]

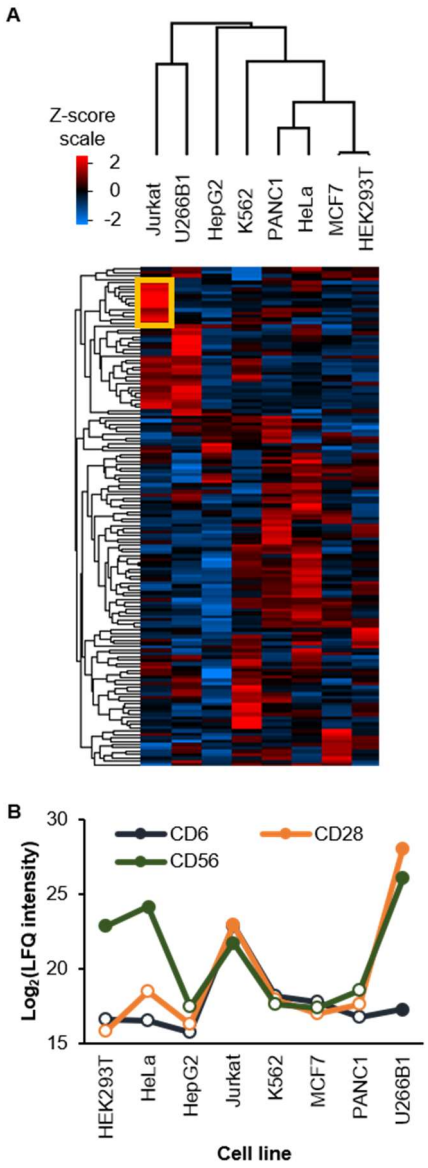

Figure 2.7. Cluster of differentiation (CD) proteins are differentially expressed in different cell types. (A) A heat map with hierarchical clustering showing relative expression of CD proteins. Missing values were imputed similar to that in Figure 2.5F. (B) Relative expression of CD6, CD28, and CD56. Missing values were imputed and are indicated by blank data points.

With the differential expression of CD proteins, the abundances of these CD molecules might need to be taken into consideration when using them to classify cell types. There is an effort to determine the expression of CD molecules through the CDmaps

project.[23] A combination of surface CD molecule identification with their abundances may help increase the accuracy of the classification considering that many of CD proteins are expressed with different abundances in different cell types.

2.3.6 Cell-surface glycoprotein interactions and pathway analyses

To explore the connection network among the quantified proteins, surface glycoprotein interactions with high confidence score were extracted from STRING database.[173] The network is complexed with several modes of interactions between these proteins including binding, catalysis, reaction, inhibition, and post-translation modification (Figure 2.8A). Proteins in the integrin family have the highest degree of interactions with other proteins, with ITGB1 interacting with the greatest number of proteins, including enzymes such as receptor-type tyrosine-protein phosphatase C, adhesion molecules such as neural cell adhesion molecule 11, and transporters such as basigin. Integrin beta-1 also interacts with other proteins in the integrin family, including ITGAV, ITGA1, ITGA2, ITGA3, and ITGA5, that are globally expressed. Integrin proteins are well-known surface proteins that play important roles at the cell interface, including acting as adhesion molecules and receptors. In fact, we identified most members of proteins in the integrin alpha family, and all proteins from the integrin beta family were identified, from ITGB1 to ITGB 8. Epidermal growth factor receptor (EGFR), a receptor tyrosine kinase with several roles in protein signaling, is another protein interacting with some integrin proteins and several other groups of proteins, such as those in the ephrin family, and transferrin receptor protein 1 (Figure 2.8A).

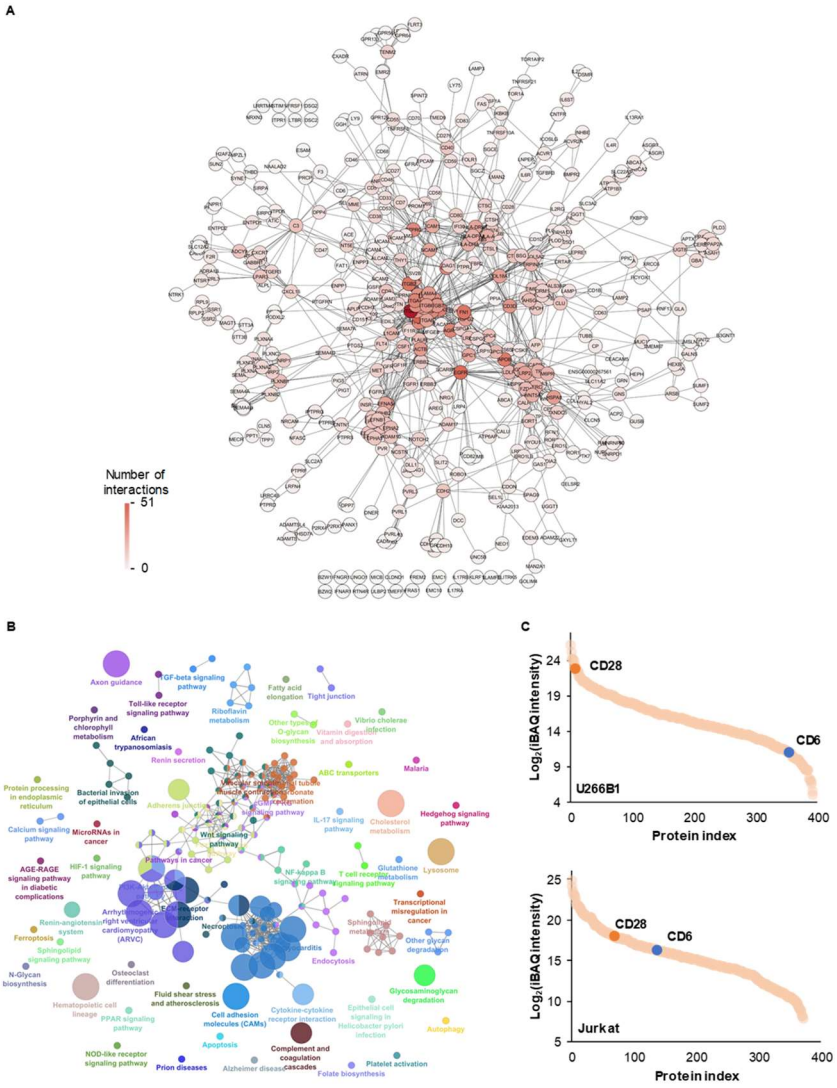

Figure 2.8. Cell-surface glycoprotein interaction and CD proteins. (A) Cell-surface glycoprotein interaction of all quantified surface glycoproteins with the data extracted from STRING. Those without any interactions were omitted. Data was processed by Cytoscape. (B) KEGG pathway analysis of all quantified surface glycoproteins. Data were processed with ClueGO plugin on Cytoscape. Default parameters were used. The size of each node corresponds to the significance of the term (a larger node means that the corrected P value is lower). The term with the highest significance is labeled on the plot. (C) Estimated abundance ranking of CD28 and CD6 from U266B1 and Jurkat cells.

Different interaction networks also arose for each cell line, with proteins in the integrin family at the center of the network (Figure 2.9A for HEK293T cells). About 40% of the quantified proteins do not have any interactions with the others. It is highly probable

that these proteins interact with others outside of the current list or their interactions have not been reported yet. However, even with 100 additional proteins extracted from outside of our dataset, some proteins, such as choline transporter-like protein 1, do not show any interactions despite that they are globally expressed in all these cell lines.

Some cell-surface glycoproteins, especially receptors, regulate signal transductions and are involved in many pathways. We assigned these quantified surface glycoproteins into 213 KEGG pathway annotations (Figure 2.8B). As expected, most of these proteins are annotated as cell adhesion molecules (CAMs). With the specific surface glycoprotein expression patterns, several pathways arise depending on the function of that cell type. For example, those from the axon guidance process were identified from HEK293T cells, a neuronal-originated cell line (Figure 2.9B), while proteins involved in the leukocyte trans-endothelial migration pathway were found in Jurkat cells (Figure 2.9B). Several signaling pathways were also identified, with 51 surface glycoproteins involved in the PI3K-Akt signaling pathway and 23 proteins in the RAP1 signaling pathways, for example. Other pathways only contain a few surface glycoproteins. These numbers are small compared to the phosphoproteome since the majority of signal transduction pathway is conducted through protein phosphorylation, and the numbers of phosphoproteins, kinases, and phosphatases far exceed the number of cell-surface glycoproteins.[170]

2.4 Conclusions

Cells are normally covered with glycans, and almost all proteins on the surface are glycosylated. Surface glycoproteins are essential for cells and regulate nearly every extracellular event, and aberrant protein glycosylation on the cell surface is directly related

to human diseases. Compared to global analysis of proteins, comprehensive analysis of surface glycoproteins is much more understudied despite their importance. It is extraordinarily challenging because of the low abundance of many surface glycoproteins, the heterogeneity of glycans, and the requirement of selective separation of glycoproteins only located on the cell surface. In this work, we characterized cell-surface glycoproteins from eight types of commonly used human cells. The distribution and occurrence of N-glycosylation sites were systematically investigated, and it was found that protein secondary structures have dramatic influence on N-glycosylation sites. For the sites with the motif N-!P-C, about one third of them are located on helix structures while those with the motif N-!P-S/T prefer strand structures over helix structures. Quantification results revealed that besides cell-specific surface proteins, the relative expression of surface glycoproteins also contributes to the uniqueness of each type. Our results suggested that it is better to consider multiple surface glycoproteins including their abundances for cell classification, rather than a single CD protein normally used in conventional methods. Global analysis of cell-surface glycoproteins facilitates a better understanding of protein glycosylation and cellular properties, and their quantitative analysis may lead to the identification of important surface glycoproteins as effective disease biomarkers and drug targets

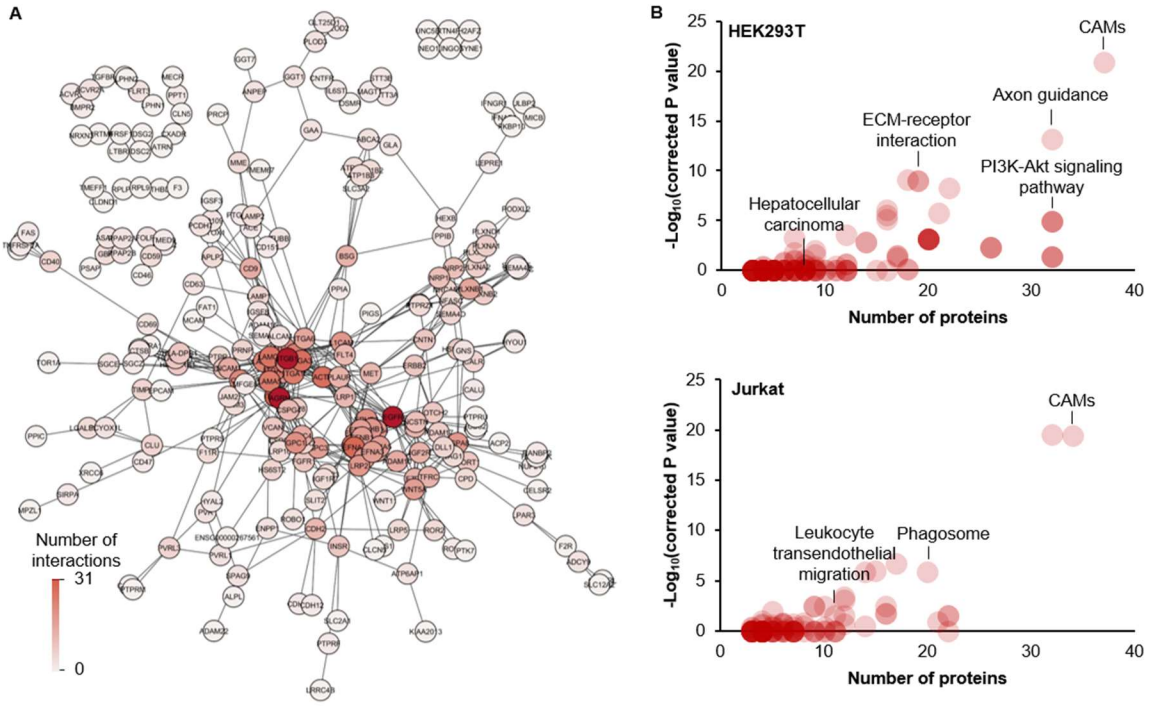

Figure 2.9. Cell-surface glycoprotein interactions and their roles in biological pathways. (A) Interaction network of surface glycoproteins from HEK293T cells extracted from STRING database. Those without any interactions are not shown. (B) KEGG pathway analysis shows enriched pathways in HEK293T (top) and Jurkat (bottom) cells, respectively.

CHAPTER 3. TEMPORAL ANALYSIS OF SURFACE GLYCOPROTEINS REVEALS DISTINCT RESPONSES OF MONOCYTES AND MACROPHAGES TO INFECTION

3.1 Introduction

The immune system protects the body from pathogens such as bacteria and viruses that can cause adverse health effects. Innate immunity is the first line of defense against these pathogens that acts in an immediate and non-specific manner. This involves physical and chemical barriers such as the epithelium and the low stomach pH. Additionally, leukocytes including macrophages and natural killer cells that may be recruited to the infection site once these pathogens penetrate into the body.[196-198] Adaptive immunity, however, is the delayed and specific response to the infections, during which B and T cells are critical.[199-200] Cell-surface glycoproteins regulate a number of cellular activities in the immune system by sensing the presence of these pathogens and functioning as receptors or transporters for cellular communication to recruit different immune cells to the infection location.[201-202] The innate and adaptive immune systems are also interconnected by antigen-presenting cells, including macrophages and dendritic cells that have major histocompatibility complex (MHC) proteins on their surface.[203]

Gram-negative bacteria contain lipopolysaccharides (LPS) that are released from the cell wall into the extracellular environment when the cell integrity is disrupted.[204] LPS

binds to pattern recognition receptors (PRRs) on the cell surface, particularly those in the toll-like receptor (TLR) family through pathogen-associated molecular pattern (PAMP) recognition. LPS triggers signaling cascades that activate transcription factors including NF-κB, AP-1, and IRF3. This promotes the pro-inflammatory response and could eventually result in the secretion of cytokines such as tumor necrosis factors, interleukins, or chemokines that attract other cells into the vicinity.[205-206] Monocytes and macrophages are important cells in the innate immune system that respond to LPS, and thus several cell-culture models such as THP-1 and U937 human monocytes or RAW264.7 mouse macrophages have been widely used for *in vitro* studies to avoid the donor-to-donor variation issue with peripheral blood mononuclear cells (PBMCs).[207-211] The two human monocyte models can be differentiated into dendritic cells or M0 macrophages and further polarized into M1 pro-inflammatory or M2 anti-inflammatory subtypes.[209, 212] Previous studies have already reported the whole proteomic and transcriptomic analyses of these immune cells in response to LPS,[208, 210, 213] while others focused on secreted proteins released after the treatment using antibody-based methods such as ELISA.[212, 214] Despite the importance of surface glycoproteins, their behaviors during immune response processes remain to be explored particularly in a time-resolved and site-specific manner, which can pinpoint the transient and important events that may not be observed in a typical study.

In this study, combining metabolic labeling, bioorthogonal chemistry, and multiplexed mass spectrometry (MS)-based proteomics, we investigated the time-resolved and site-specific responses of surface glycoproteins on THP-1 monocytes and macrophages to LPS and during the monocyte-to-macrophage differentiation. Differential remodeling of the surface glycoproteomes was observed among the cells during these processes,

including the expression of new glycoproteins to the surface and the removal/internalization of existing surface glycoproteins that participate in different biological processes. Interestingly, the abundance of some surface glycoproteins was transiently altered before being restored to the normal state, which would not be observed in an experiment without the time factor. Nonetheless, some glycoproteins were minimally affected. A comparison of both cell types revealed that surface glycoproteins that responded similarly are involved in general functions of surface proteins, such as cell-cell adhesion, but those that responded differently participate in the immune response process. The results suggested that the priming of the cells during the differentiation process may be responsible for the different responses in monocytes and macrophages to LPS. Apart from previously reported markers, novel surface glycoproteins in the immune response process were also detected. Furthermore, selective and site-specific protein glycosylation was observed in these systems. This work results in a better understanding of the important roles of cell-surface glycoproteins in the immune response during the bacterial infection and potentially the identification of surface glycoproteins as disease biomarkers and drug targets.

3.2 Materials and methods

3.2.1 Cell culture, monocyte-to-macrophage differentiation, and cell-surface glycoprotein labeling

All reagents were purchased from Sigma-Aldrich unless noted otherwise. THP-1 cells (ATCC) were maintained in RPMI-1640 medium (Gibco) containing 10% fetal bovine serum (Corning) and 1% penicillin-streptomycin in a humidified incubator with 5%

carbon dioxide at 37 ºC. Monocyte-to-macrophage differentiation was performed similar to the previously reported protocol.[208] Briefly, the cells were grown until the density reached 10^6 cells/mL and then treated with 100 ng/mL phorbol 12-myristate 13-acetate (PMA) in the normal medium for 48 hours. Adherent cells were washed with PBS buffer and rested in the normal medium without PMA for 24 hours. For the identification experiments, surface glycoproteins were labeled similar to the method reported previously.[215-216] Briefly, THP-1 monocytes at the cell density of $8x10^5$ cells/mL or the differentiated M0 macrophages were labeled with 100 μM N-azidoacetylgalactosamine-tetraacylated (Ac_4GalNAz, Click Chemistry Tools) in the normal medium in the humidified incubator. After 24 hours, the cells were tagged with 100 μM dibenzocyclooctyne (DBCO)-biotin (Click Chemistry Tools) in PBS for 1 hour at 37 ºC. For the differentiated macrophages, Cellstripper solution (Corning) was employed instead of trypsin to detach the cells. The reaction was quenched with 100 mM dithiothreitol (DTT) in ice-cold PBS, centrifuged at 300*g* for 4 minutes, and kept on ice until the next step.

3.2.2 Time-resolved quantification of surface glycoproteins in response to LPS and during the PMA-induced differentiation

Time-resolved experiments were performed for the treatment of THP-1 monocytes with LPS, the treatment of M0 macrophages with LPS, and the differentiation of monocytes to M0 macrophages. For each experiment, biological duplicate experiments were performed. To account for the difference in Ac_4GalNAz incorporation efficiency among glycoproteins and the effect of cell growth, control experiments were also included by treating the cells with PBS (for the LPS experiments) or DMSO (for the PMA experiments). The LPS-treatment experiments were performed by pre-labeling the cells

with 100 μM Ac_4GalNAz 12 hours before the LPS treatment, which is considered as a 0-hour time point. THP-1 monocytes or the differentiated macrophages were treated with 1 μg/mL LPS from *E. coli* O111:B4. After 0, 1, 3, 6, 12, and 24 hours, the cells were collected and tagged with 100 μM DBCO-biotin similarly to the identification experiment. For the experiment with M0 macrophages, the pre-labeling took place 12 hours after the start of the resting period and 12 hours before the LPS treatment similar to the monocyte experiment. For the differentiation experiment, monocytes at the density of 10^6 cells/mL were treated with 100 ng/mL PMA and tagged with DBCO-biotin after 0, 3, 6, 12, 24, and 48 hours. The reaction was quenched with 100 mM DTT in ice-cold PBS, centrifuged at 300*g* for 4 minutes, and stored at -80 °C until cells from different time points were obtained.

3.2.3 Cell lysis, protein digestion, and peptide purification

The cell pellets were treated with a buffer containing 25 μg/mL digitonin, 150 mM sodium chloride (NaCl), 50 mM N-(2-hydroxyethyl)piperazine-N'-2-ethanesulfonic acid (HEPES, pH=8.8), and 1 tablet/10 mL cOmplete ULTRA tablets protease inhibitor cocktail (Roche) at 4 °C for 10 minutes on an end-over-end rotator. The suspension from each sample was centrifuged at 2,000*g* for 10 minutes, and the resulting pellet was lysed with an ice-cold lysis buffer containing 0.5% sodium deoxycholate, 50 mM HEPES (pH=8.8), 150 mM NaCl, 20 units/mL universal nuclease for cell lysis (Pierce), and 1 tablet/10 mL cOmplete ULTRA tablets protease inhibitor cocktail at 4 °C for 1 hour on an end-over-end rotator. The lysates were centrifuged at 25,830*g* for 10 minutes and the supernatants were collected. Proteins were reduced with 5 mM DTT at 56 °C for 25 minutes, alkylated with 14 mM iodoacetamide for 30 minutes in the dark, and quenched by incubating with DTT to the final concentration of 5 mM in the dark for another 15 minutes. Proteins were

purified and pelleted with the methanol/chloroform precipitation and digested with sequencing grade modified trypsin (Promega) at 37 ºC for 16 hours (enzyme:substrate ratio of ~1:100) in a buffer containing 5% acetonitrile (ACN), 1.6 M urea, and 50 mM HEPES (pH=8.8). Once the digestion is completed, the pH of the mixture was adjusted to <2 using trifluoroacetic acid (TFA). Peptides were desalted with a Sep-Pak Vac tC18 cartridge (Waters) and dried in a vacuum concentrator.

3.2.4 Glycopeptide enrichment, TMTsixplex labeling, and PNGase F treatment

Glycopeptides containing biotin were enriched with the high-capacity NeutrAvidin agarose resins (Thermo) according to the manufacturer's protocol. The enriched peptides were eluted three times with 200 µL 8 M, pH=1.5 guanidine hydrochloride (Promega) at 56 ºC for 2 minutes each. The eluates were pooled, desalted, and dried in a vacuum concentrator overnight. For peptides from the time-resolved experiments, BCA assay (Pierce) was performed with peptides from each time point to normalize the total protein abundance before the enrichment. The peptides were dried after the enrichment, resuspended in a buffer containing 200 mM HEPES (pH=8.5) with 40% ACN, and labeled with each channel of the TMTsixplex reagents. The reaction was performed at room temperature with shaking for 1 hour and subsequently quenched with 8 µL of 5% hydroxylamine hydrochloride in 200 mM HEPES (pH=8.5) at room temperature with shaking for 15 minutes. The labeled peptides from all time points were pooled, purified, and dried overnight.

Glycopeptides were treated with 3 units of PNGase F in 40 µL of 40 mM ammonium bicarbonate (pH=9) in heavy-oxygen water ($H_2{}^{18}O$) at 37 ºC for 3 hours with

shaking. The reaction was quenched with formic acid (FA) to the final concentration of 1.0%. The peptides were desalted with the StageTip method and eluted into three fractions with 20%, 50%, and 80% ACN containing 1% acetic acid.[161] The eluates were dried again in a vacuum concentrator before LC-MS/MS analysis.

3.2.5 LC-MS/MS analysis

The peptides were dissolved in a solution containing 5% ACN and 4% FA and loaded into a microcapillary column packed in-house (Magic C18AQ, 3 µm, 200 Å, 75 µm x 16 cm). The peptides were separated with reverse-phase liquid chromatography on a Dionex UltiMate 3000 UHPLC system (Thermo) using 112-minute gradients of 3-22%, 6-30%, and 12-45% or 3-14%, 4-17%, and 8-24% of ACN with 0.125% formic acid for the 20%, 50%, and 80% ACN fractions from the identification and time-resolved quantification experiments, respectively. The UHPLC was connected on-line to an LTQ Orbitrap Elite hybrid mass spectrometer (Thermo). For each cycle, a full MS spectrum was recorded in the Orbitrap at the resolution of 60,000 with $1x10^6$ automatic gain control (AGC). In the identification experiments, collision-induced dissociation (CID) was used due to the faster speed and higher sensitivity, allowing for the better coverage of surface glycoproteins from both monocytes and macrophages. For the quantification experiments including the treatment with LPS and PMA-induced differentiation, higher-energy collision dissociation (HCD) was employed to accurately quantify glycopeptides based on the TMT reporter ions in the low mass-to-charge range. The TMT labeling enabled us to quantify glycopeptides from several time points simultaneously, increasing the throughput and the quantification accuracy. For the identification experiments, peptides were fragmented in the LTQ using a data-dependent Top20 method; for each cycle, a full MS

scan in the Orbitrap was followed by up to 20 MS/MS in the LTQ for the most intense ions. Selected ions were excluded from further sequencing for 90 seconds. Ions with singly or unassigned charge were not sequenced. Maximum ion accumulation times were 1,000 ms for each full MS scan and 50 ms for MS/MS scans. For the quantification experiments, peptide fragmentation was performed by higher-energy collisional dissociation (HCD) with the normalized collisional energy of 40% using a Top15 method, i.e., up to 15 MS/MS of the most abundant peaks were recorded in the Orbitrap cell at the resolution of 15,000.

3.2.6 Database searching, data filtering, and bioinformatic analysis

The raw files were converted into an mzXML format and searched using the SEQUEST algorithm (version 28).[217] The spectra were matched against the human proteome downloaded from UniProt (www.uniprot.org). The peptide mass tolerance was 20 ppm, and the fragment ion mass tolerance was 1.0 Da for the identification experiment (detected in the LTQ cell) or 0.025 Da for the quantification experiment (detected in the Orbitrap cell). The maximum number of missed cleavages was 3 and the maximum number of differential modifications was 4. The following amino acid modifications were used: static modifications of +57.0215 Da for the carbamidomethylation of cysteine and +229.1629 Da for the TMT labeling at lysine and the N-terminus (only for the quantification experiment), and differential modification of +2.9883 Da for the glycosylation site and +15.9949 Da for the oxidation of methionine. The target-decoy method was employed to estimate the false discovery rates (FDRs) of glycopeptide and glycoprotein identifications.[218] Linear discriminant analysis was applied to control the quality of glycopeptide identification using multiple parameters such as XCorr, ppm, peptide length, and charge state. The FDR for glycopeptide identifications was controlled

to <1%, and an additional filter was also applied to control the FDR to <1% at the protein level. The dataset was limited to only glycopeptides or glycoproteins when calculating the FDRs at both levels.

Reverse hits and common contaminants were removed from the dataset. The confidence of the glycosylation localization was performed using a method called ModScore, which is similar to Ascore for protein phosphorylation.[219] Glycopeptides with the score greater than 13 have the localization confidence of >95% while those with the score greater than 19 show the confidence of >99%. For the quantification experiments, the correction parameter from Thermo was used to adjust the TMT reporter ion intensities. The signal-to-noise (S/N) ratios were used for peptide-level quantification. Glycopeptides with S/N ratios less than 5 in any channel were removed from further quantification. The summed-based S/N of glycopeptides were employed for glycoprotein quantification. The glycoprotein abundance was normalized to that of the 0-hour time point and between the control and treatment groups. If the protein was detected in both biological duplicate experiments, the average ratio is used. Only those detected in both the treatment and control groups were reported because their abundances can be normalized for the incorporation efficiency of Ac_4GalNAz and the elimination of the possible effect from cell growth. For site-specific quantification, glycopeptides were filtered to those only containing one glycosylation sites and have a ModScore of at least 13. Depending on the experiment, proteins were clustered with the Gene Ontology Resource (www.geneontology.org) or the Database for Annotation, Visualization and Integrated Discovery (DAVID).[220] Changes of surface glycoprotein abundances were classified into four types based on the following criteria: A) Proteins that increased their abundances over time. The fold change of at least

one time point should be greater than 1.5-fold and did not decrease to lower than half of the maximum abundance or below 1.5. All data points should be greater than 0.67; B) Proteins that abruptly increased their abundances and then returned to normal. The fold change of at least one time point should be greater than 1.5-fold and decreased to lower than half of the maximum abundance or below 1.5. All data points should be greater than 0.67; C) Proteins that decreased their abundances. The fold change of at least one data point should be lower than 0.67. Other data points should be lower than 1.5; D) No change. The fold change is within the 0.67-1.5 range. Others are considered as not determinable (N/D). Ingenuity Pathway Analysis (IPA; Qiagen) was performed using the default parameters to predict the upstream regulators and downstream effects, as well as the regulator effects. Since IPA considers each time point (not the change type, i.e., A, B, C, D, or N/A), the cutoff for up- or down-regulation is >1.5 or <0.67, respectively. Only the significant results are included in Table S5.

3.3 Results and discussion

3.3.1 Site-specific identification of glycoproteins on the cell surface of monocytes and macrophages

Modern mass spectrometry-based proteomics provides a possibility to globally analyze proteins and their modifications.[148, 221-227] To target surface glycoproteins, we need to selectively separate them prior to MS analysis due to their low abundance compared with cytosolic proteins.[228-231] We systematically investigated the dynamics of glycoproteins on the surface of THP-1 monocytes and macrophages in response to LPS in time-resolved and site-specific manners by combining metabolic labeling, bioorthogonal

chemistry, and multiplexed proteomics (Figure 3.1A and Figure 3.1B).[176, 181, 215, 232] THP-1 cells were labeled with 100 μM N-azidoacetylgalactosamine-tetraacylated ($Ac_4GalNAz$). Surface glycoproteins containing the functional azido group were then tagged with 100 μM dibenzocyclooctyne (DBCO)-biotin while the cells remained intact. After cell lysis, protein extraction, and digestion, the tagged glycopeptides from surface glycoproteins were selectively enriched using NeutrAvidin beads, followed by a deglycosylation reaction with PNGase F in $H_2{}^{18}O$ and MS analysis. The glycopeptides from the quantification experiments were also labeled with the TMT reagents to determine the protein dynamics.

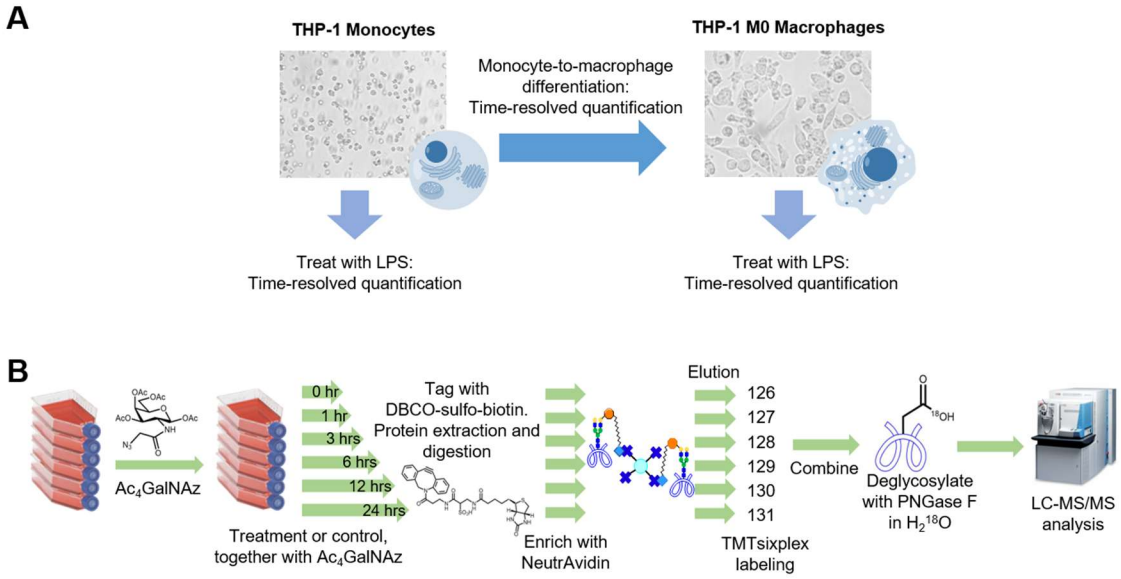

Figure 3.1. Experimental procedure for global quantitation of the dynamics of surface glycoproteins. (A) The experiments with THP-1 monocytes and macrophages, including the LPS treatment and the differentiation induced by PMA. (B) The general time-resolved experiment scheme to determine the dynamics of cell-surface glycoproteins after the treatment.

Overall, 764 cell-surface glycoproteins and 2,205 unique glycopeptides were detected from all identification and quantification experiments (Figure 3.2A). Over 400 glycoproteins and 1,000 glycopeptides were found commonly in both monocytes and

macrophages. Gene ontology (GO)-based clustering of all identified glycoproteins based on the biological process showed that immune system process ($P=4.99x10^{-57}$), cell adhesion ($P=4.93x10^{-51}$), cell-surface receptor signaling pathway ($P=3.84x10^{-41}$), and transmembrane transport ($P=9.60x10^{-11}$) are highly enriched, which is in excellent agreement with the known functions of cell-surface glycoproteins (Figure 3.2B). Based on cellular component, over 600 glycoproteins are annotated to the membrane part ($P=5.27x10^{-116}$) while ~500 are on the plasma membrane ($P=9.71x10^{-101}$), demonstrating that the current experiments are effective for specific detection of glycoproteins on the cell surface. The reproducibility of the results from biological duplicate experiments is reasonably high, such as those from the identification experiments with monocytes (Figure 3.2C). Some glycoproteins were found only in a specific cell type despite the identical genome backgrounds of THP-1 monocytes and macrophages. Clustering of 180 proteins identified specifically in monocytes revealed highly enriched processes such as humoral immune response mediated by circulating immunoglobulin ($P=9.30x10^{-10}$) and regulation of complement activation ($P=3.2x10^{-8}$) (Figure 3.2D). For 107 proteins found only in macrophages, those involved in the regulation of cytokine secretion ($P=1.30x10^{-8}$), response to molecules of bacterial origin ($P=2.20x10^{-7}$), and regulation of adaptive immune response ($P=8.00x10^{-4}$) are highly enriched (Figure 3.2D).

Among important groups of proteins in the immune response process, several in the TLR family were identified. Only TLR2 was detected in monocytes while TLR1, TLR2, TLR4, and TLR6 were found in macrophages. Nonetheless, TLR2 was the only protein quantified in the time-resolved experiments. Other important proteins were also detected, such as MHC protein classes I and II for antigen presentation (including HLA-A,

HLA-B, HLA-E, HLA-F, HLA-DRA, and HLA-DRB1) and Fc receptors for antigen recognition (including FCAR, FCER2, FCGRT, and FCGR1A). A total of 158 clusters of differentiation (CD) glycoproteins were detected. Only seven were found exclusively in monocytes, including killer cell immunoglobulin-like receptor 3DL1 (KIR3DL1; CD158e). Twenty-three CDs were identified exclusively in macrophages, such as CD180 antigen that cooperates with MD-1 and TLR4 during LPS binding. Some of these may be useful for future cell type classification.

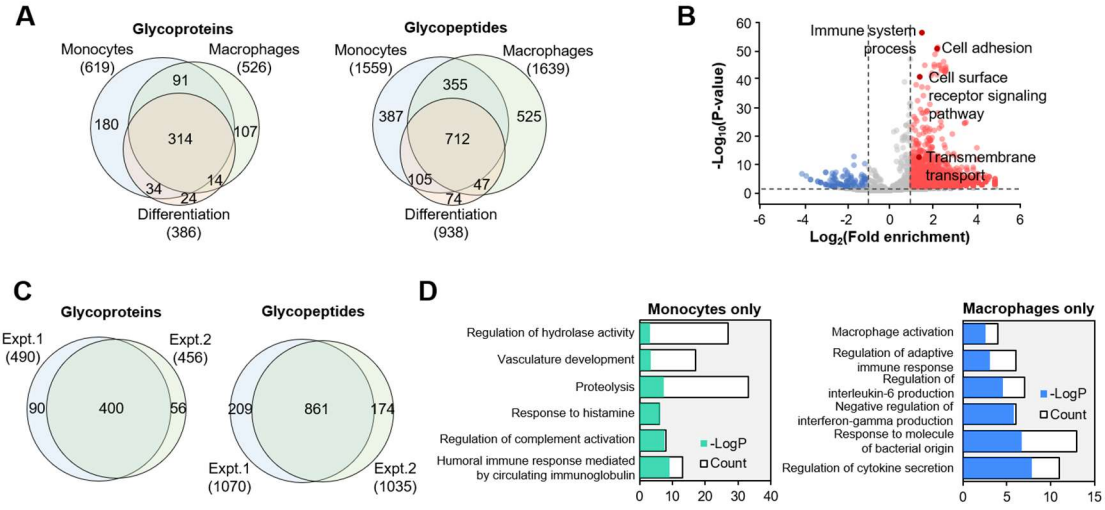

Figure 3.2. Glycoproteins identified from all experiments and some functional analysis results. (A) The numbers of all detected glycoproteins and glycopeptides. (B) GO clusters of the identified glycoproteins based on the biological process. (C) Overlaps of glycoproteins and glycopeptides on monocytes from the duplicate identification experiments. (D) Clustering of surface glycoproteins found exclusively in monocytes (n=180) or macrophages (n=107). The x-axis represents the number of surface glycoproteins in each cluster or the -log(P) value, and both share the same x-axis. The P-values for protein clustering were determined using DAVID.

3.3.2 Time-resolved changes of cell-surface glycoproteins on monocytes in response to LPS

For THP-1 monocytes, we detected >1,500 unique glycopeptides from >600 glycoproteins in both the identification and quantification experiments (Figure 3.3A). In

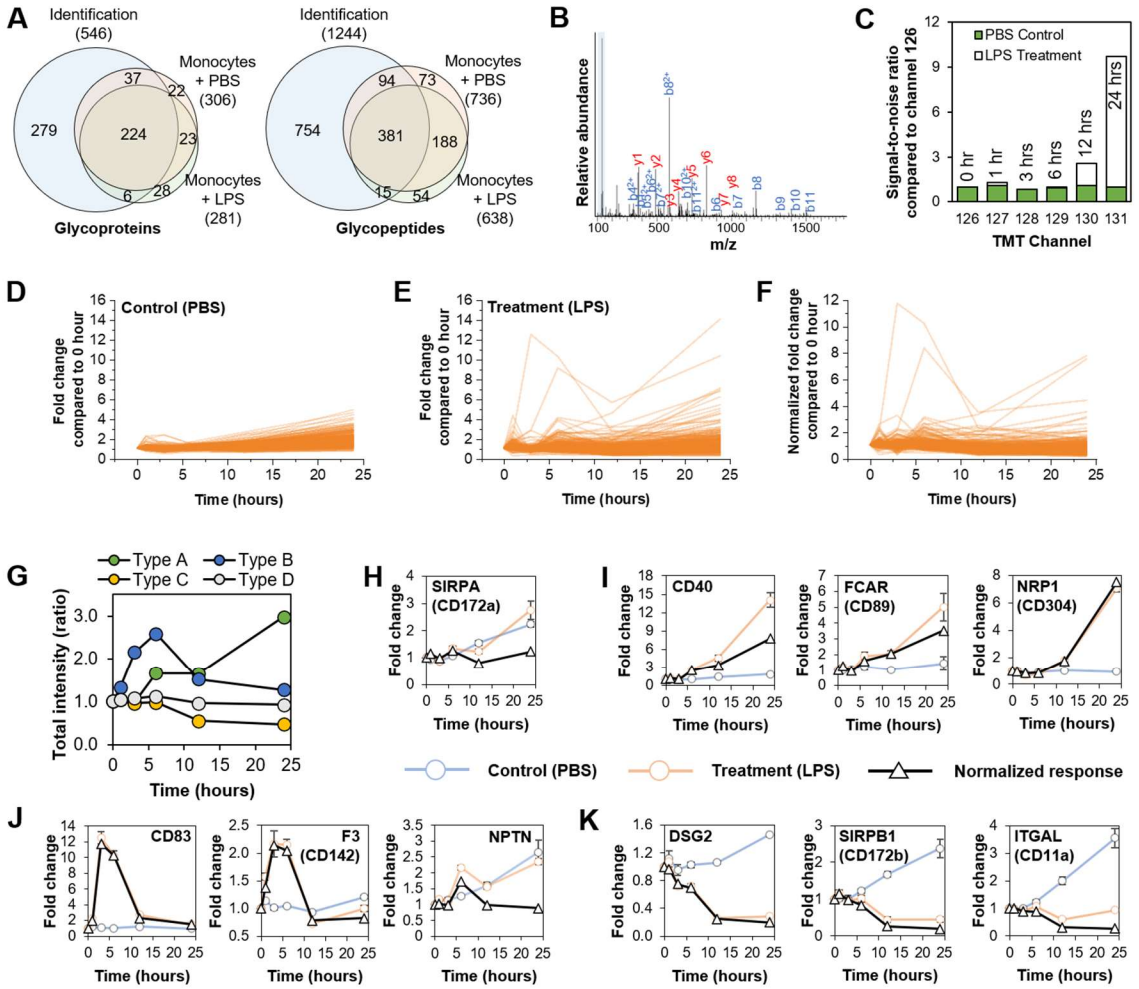

Figure 3.3. Quantitation of cell-surface glycoproteins on monocytes in response to LPS. (A) Overlaps of glycoproteins and glycopeptides from the identification and quantitation experiments. (B) An example MS/MS spectrum of a quantitated glycopeptide (see text). The highlighted blue region shows the TMT reporter ions. (C) The TMT ratios of the peptide from B) demonstrating the difference between the control (green) and the treatment (blank) groups. (D-F) Time-resolved plots of surface glycoprotein abundances over 24 hours from the PBS-treated (control, D) or LPS-treated (E) cells, and the normalized abundances (F). (G) Total intensities of glycoproteins that changed their abundances in four different patterns. The plot shows glycoproteins that increased their abundances gradually over time (Type A), increased abruptly and returned to normal (Type B), decreased (Type C), and were not affected (Type D). (H-K) Example glycoproteins from Type D (H), Type A (I), Type B (J), and Type C (K). The raw and normalized abundances are displayed. The error bars represent one standard deviation.

the quantification experiments, the cells were pre-labeled with Ac_4GalNAz for 12 hours to allow sufficient labeling of the glycoproteins before the treatment with 1 μg/mL LPS or PBS (as a control). The cells were collected at six time points over 24 hours and the surface glycopeptides from different time points were extracted. Eventually, the glycopeptides were labeled with the sixplex TMT reagents and analyzed with LC-MS/MS (Figure 3.1B). In one example, the glycopeptide RGPECSQN#YTTPSGVIK (# refers to the glycosylation site) from neuropilin-1 (NRP1) was confidently identified and quantitated with the XCorr score for the peptide identification of 5.3 and the ModScore for the glycosylation site localization of 1000 (Figure 3.3B). While the abundance of this glycopeptide remained relatively the same across 24 hours in the control group, its abundance in the cells treated with LPS increased rapidly at 12 and 24 hours (Figure 3.3C).

In the control experiments, the abundances of cell-surface glycoproteins generally increased over time due to the cell growth (Figure 3.3D), in contrast to the LPS-treated groups where the abundances of some surface glycoproteins changed dramatically (Figure 3.3E). The abundances at each time point from the treatment group were then normalized by the abundances of the corresponding surface glycoproteins from the PBS-treated group, resulting in 247 glycoproteins in the final dataset with high reproducibility (Figure 3.3F, Figure 3.4A, and Figure 3.4B). The quantitated glycoproteins were classified into four types based on their abundance changes (Figure 3.3G). Type A (12 glycoproteins, 5%) increased their abundance gradually and did not return to the normal level as before the treatment. These glycoproteins are involved in the regulation of inflammatory response, the regulation of programmed cell death, and cell development. One example is CD40, which is typically expressed on B cells and binds to the CD40L ligand on T cells under

inflammatory conditions.[233] The protein gradually increased its abundance to >7 times at 24 hours after the treatment (Figure 3.3I). Type B (18 glycoproteins, 7%) also increased their abundance abruptly but returned to the normal level. These glycoproteins are involved in protein transport and ion sequestering (especially calcium), and cytokine production (IL-2, IL-4, IL-10, and IFN-γ). One Type B glycoprotein that stood out in the profile plot is CD83, which also has the highest abundance compared with other glycoproteins at 3 hours (Figure 3.3J). CD83 is a marker for mature dendritic cells and involved in antigen presentation.[234] The previous report showed that the protein can be pre-formed inside immune cells and exported to the surface once stimulated with LPS.[235] Type C (92 glycoproteins, 37%) decreased their abundance throughout the treatment time, with processes including extracellular matrix organization, positive regulation of cell proliferation, interleukin production (IL-5 and IL-13), and cell-cell adhesion enriched among proteins in the group. These proteins, such as integrin alpha-L (ITGAL; CD11a; Figure 3.3K), may be internalized and degraded by cells or shed into the extracellular environment by ectodomain shedding.[236]

While the response from the three glycoprotein types correspond with the known responses from cells to bacterial infection,[197] we observed several glycoproteins, such as CD172a (Figure 3.3H), that were minimally affected by the treatment. These Type D glycoproteins (64 glycoproteins, 26%) may not participate in the response towards LPS or may be regulated through their interactions with other molecules rather than changing their abundances. Additionally, some glycoproteins did not meet the criteria for any above patterns, but instead their abundances fluctuated over the treatment period, and hence the type cannot be assigned (Figure 3.5B). Other example proteins are displayed in Figure 3.6

and Table 3.1). Some of these proteins, to our knowledge, have never been linked to the cellular response towards LPS and may play important roles in the immune response process, such as adipocyte plasma membrane-associated protein (APMAP) that has arylesterase activity. The protein increased its abundance abruptly to 1.8-fold at 3 hours after the LPS challenge before returning to normal.

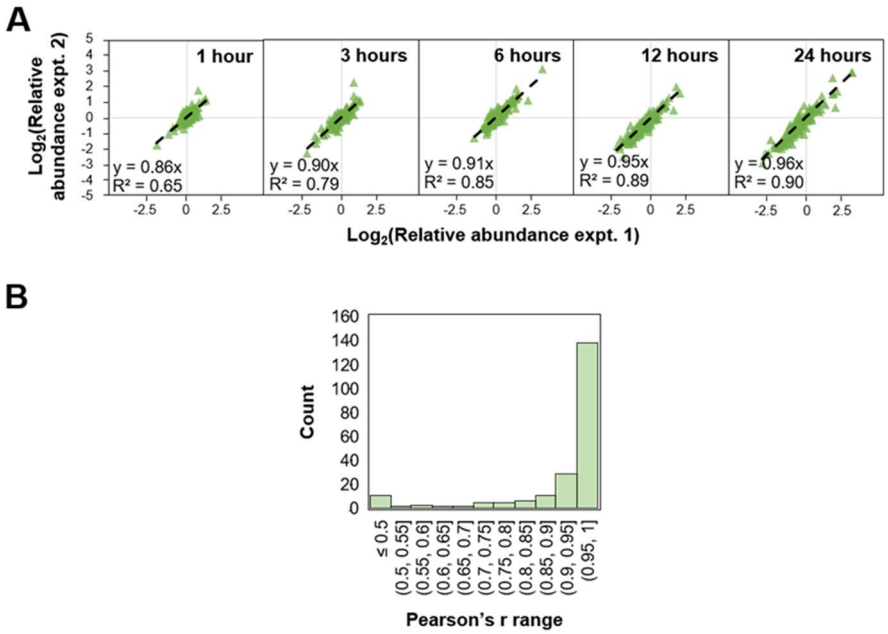

Figure 3.4. Reproducibility of the quantification of cell-surface glycoproteins on monocytes in response to LPS. (A) Comparison of the normalized abundance at each time point from both experiments. The correlation of the normalized abundance increased over time. The correlation after 1 hour is relatively low (R^2=0.65), which may be due to the low abundances of labeled glycopeptides. The correlation increases after each time point and reaches R^2=0.90 at 24 hours. (B) Alternatively, the correlation of all six time points was compared from two biological duplicate experiments simultaneously using Pearson's correlation and about two thirds of the proteins quantified (66%) have the Pearson's R values greater than 0.95.

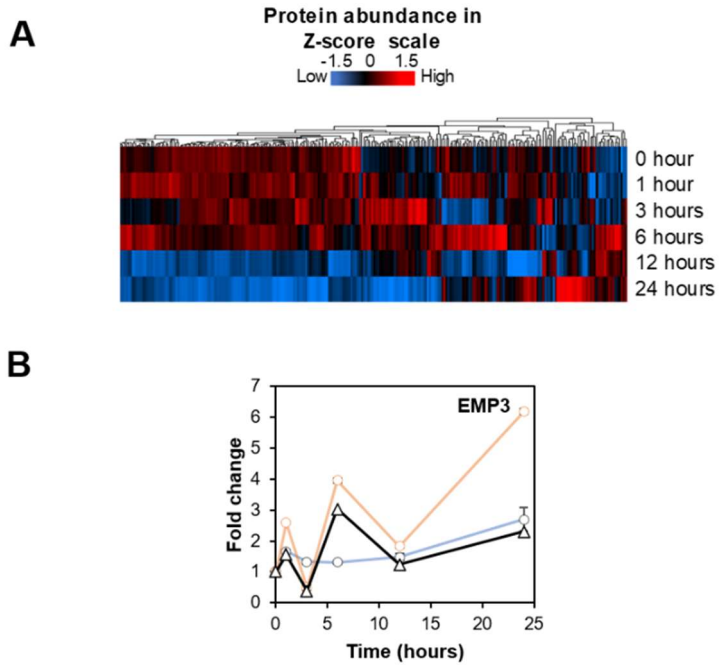

Figure 3.5. Quantification of cell-surface glycoproteins on monocytes in response to LPS. (A) heat map based on hierarchical clustering with log_2- and Z-score-transformed normalized abundance showing changes of cell-surface glycoproteins during the LPS treatment. (B) In some cases, the response type is inconclusive such as the signal from EMP3 that fluctuates over time. The control (PBS), treatment (LPS), and normalized intensities are shown similar to Figure 3.3.

3.3.3 Quantification of the dynamics of cell-surface glycoproteins on macrophages in response to LPS

With M0 THP-1 macrophages, >1,600 glycopeptides were detected from >500 surface glycoproteins (Figure 3.7A). Similar to the experiment with monocytes, the differentiated macrophages were labeled with Ac_4GalNAz, treated with either 1 µg/mL LPS or PBS, and collected over 24 hours for surface glycoprotein analysis (Figure 3.7B and Figure 3.7C). Eventually, 223 glycoproteins were quantitated commonly in the control and treatment experiments (Figure 3.7A, Figure 3.7D, and Figure 3.8A). The quantitated glycoproteins were classified into the four response types as above, including 24 Type A (11%), 36 Type B (16%), 68 Type C (30%), and 56 Type D (25%) glycoproteins. Type A

glycoproteins function in the nitric oxide metabolic process, negative regulation of programmed cell death, and positive regulation of reactive oxygen species biosynthetic process. Type B glycoproteins are involved in the positive regulation of adaptive immune response, sequestering of ions, including calcium similar to that of monocytes, detection of bacterium, and positive regulation of chemotaxis. Antigen processing and presentation of endogenous peptide antigen were also enriched in this group. Those that decreased their abundances (Type C) participated in cell migration and the regulation of cell-cell adhesion. Some examples of glycoproteins that significantly changed their abundances on the surface

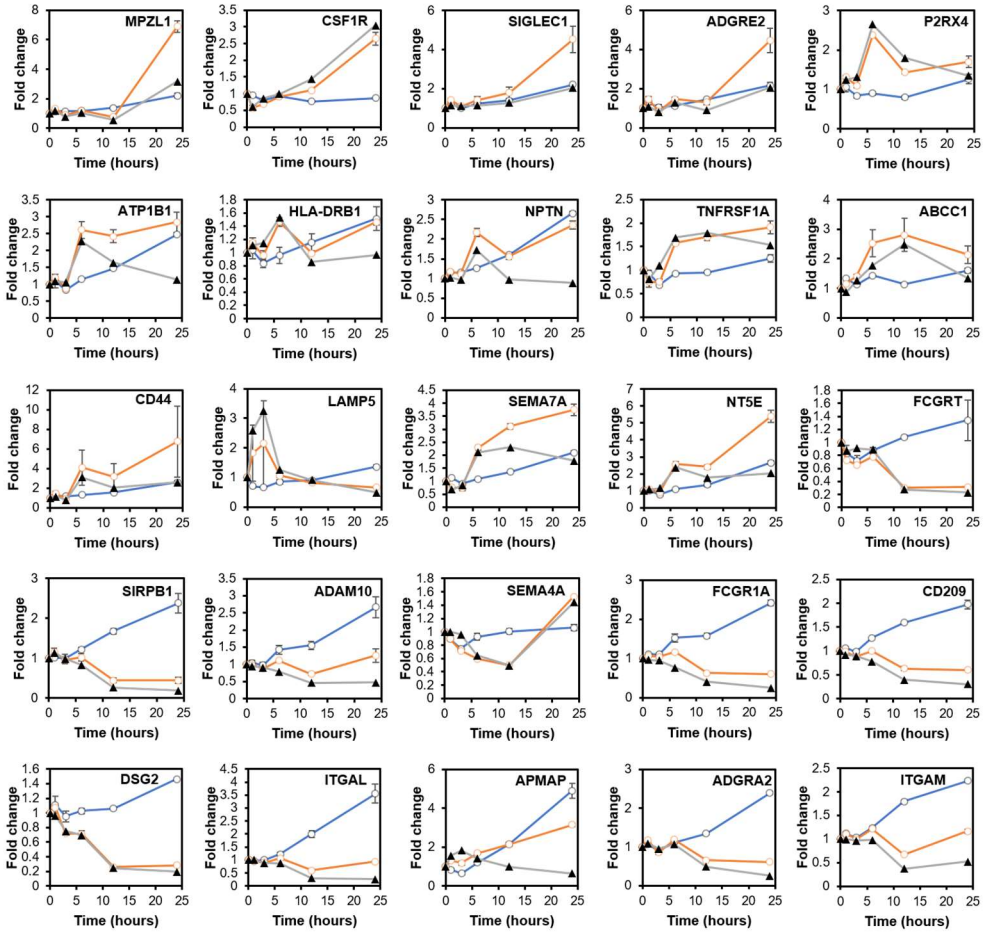

Figure 3.6. Example quantified cell-surface glycoproteins on monocytes in response to LPS. The abundance changes of glycoproteins in the control (PBS) and treatment (LPS) experiments, and the normalized abundance changes are shown similar to Figure 3.3.

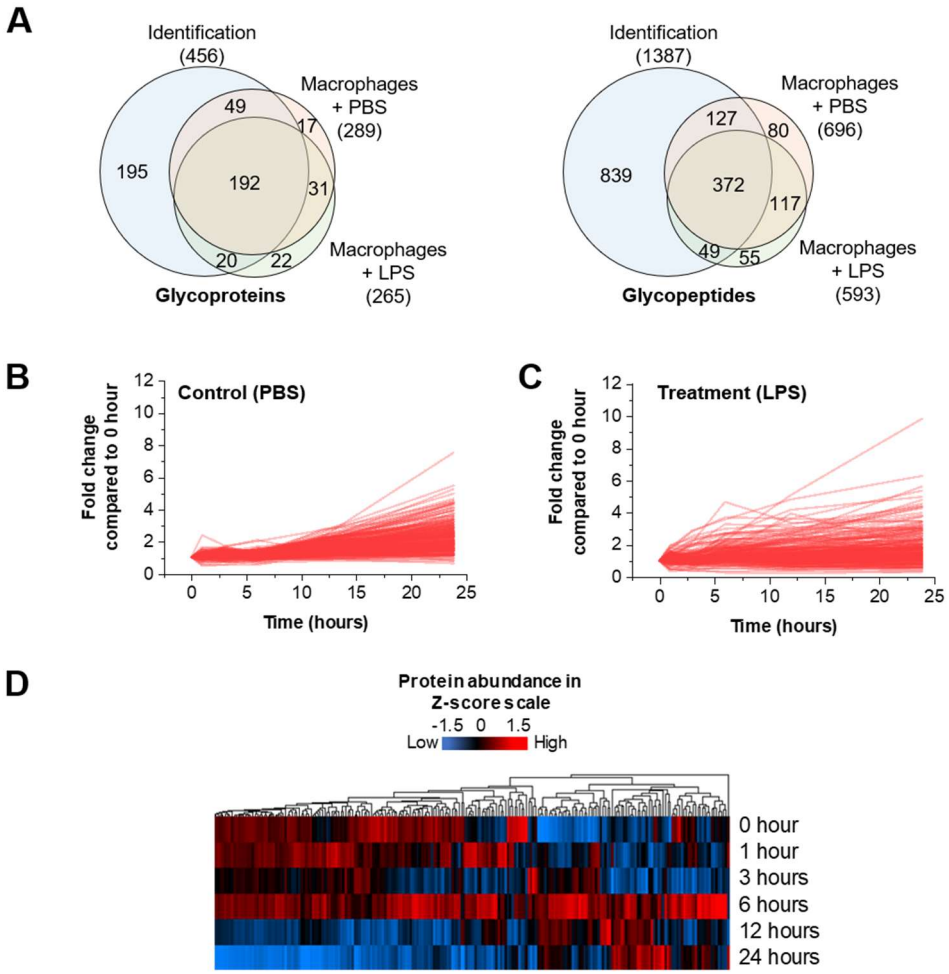

Figure 3.7. Quantification of cell-surface glycoproteins on macrophages in response to LPS. (A) Overlaps of glycoproteins and glycopeptides from the identification and quantification experiments (control and treatment). (B-C) Profile plots for surface glycoprotein abundances on macrophages treated with PBS (B) or LPS (C). (D) A heat map based on hierarchical clustering with log_2- and Z-score-transformed normalized abundance showing changes of cell-surface glycoproteins from macrophages treated with LPS.

include tumor necrosis factor receptor superfamily member 9 (TNFRSF9; CD137; Type B), which reached its highest abundance at 6 hours after the treatment and is consistent with two previous MS-independent studies (Figure 3.8B).[237-238] HLA-A, a protein that presents antigens to the cell surface, had a peak abundance at 12 hours, suggesting the activation or connection of the innate and adaptive immune systems. T-lymphocyte

activation antigen CD80, a marker for the M1 subtype, increased its abundance on the cell surface, showing that the LPS treatment might polarize them toward the M1 pro-inflammatory state as reported previously.[239] It was not detected in the control experiment, likely due to the higher expression induced by LPS, with an increased abundance over 13 times after 24 hours (Figure 3.9). Other quantified glycoproteins are included in Figure 3.10 and Table 3.1. Similar to monocytes for glycoproteins that have never been associated with the response to LPS, APMAP increased its abundance to 2.7-fold after 6 hours in response to LPS.

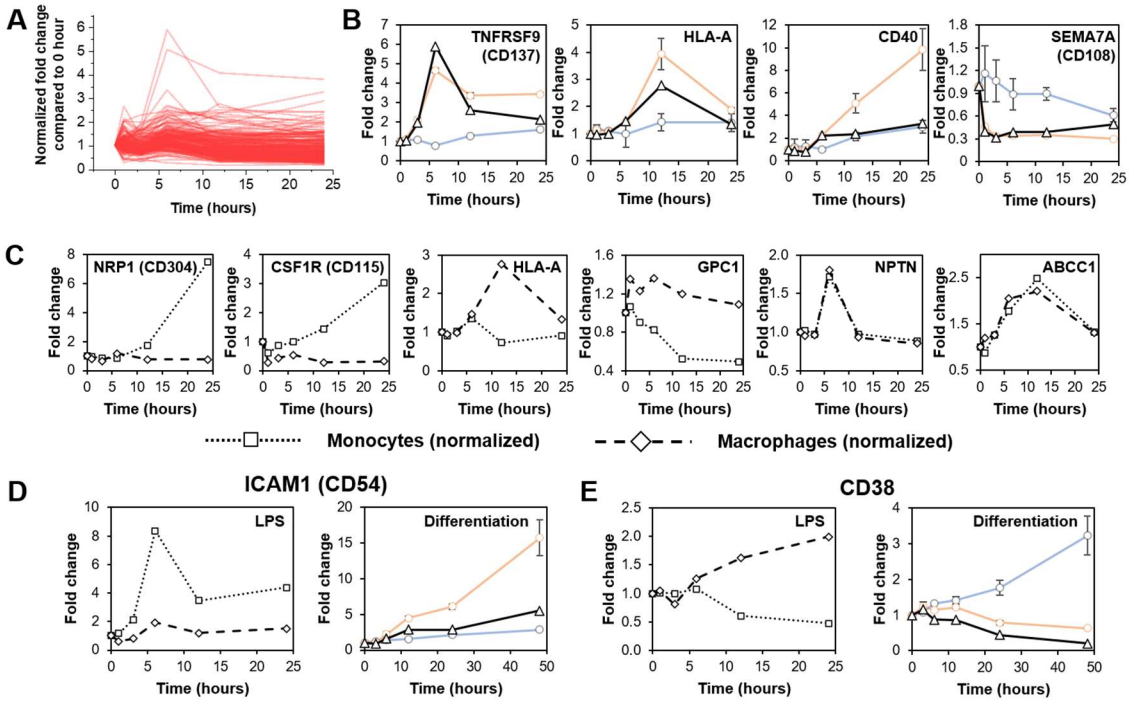

Figure 3.8. Quantitation of cell-surface glycoproteins on macrophages in response to LPS. (A) Normalized abundances of cell-surface glycoproteins from macrophages in response to LPS. (B) The abundance changes of example surface glycoproteins. The control (PBS), treatment (LPS), and normalized intensities are shown similar to Figure 3.3. The error bars represent one standard deviation. (C) Comparison of the abundance changes of some surface glycoproteins on monocytes and macrophages in response to LPS. (D-E) Abundance changes of ICAM1 (D) and CD38 (E) on the cell surface showing the priming of the cells during the monocyte-to-macrophage differentiation for their specific functions.

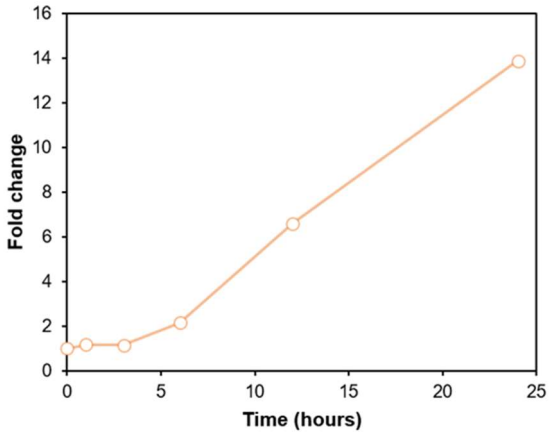

Figure 3.9. Abundance of CD80 on macrophages treated with LPS without normalization with the abundance from cells treated with PBS.

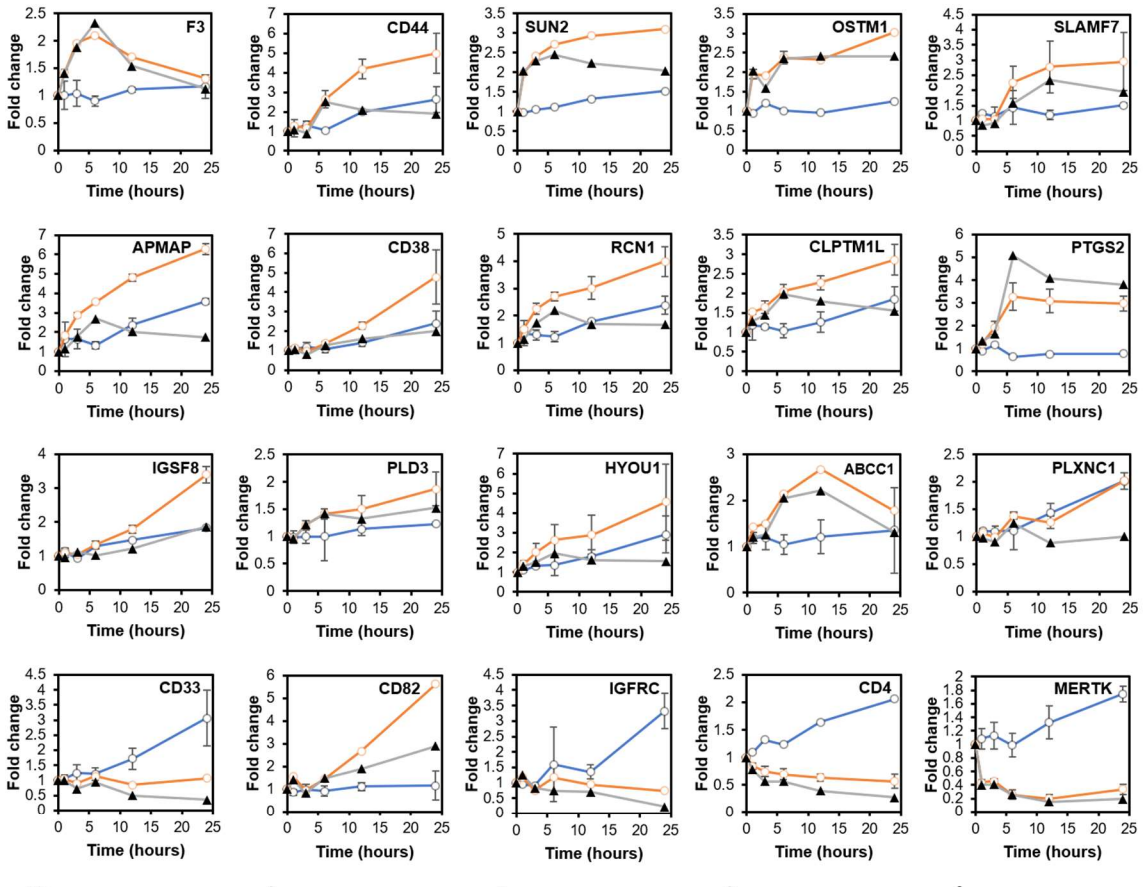

Figure 3.10. Quantification of example cell-surface glycoproteins on macrophages in response to LPS. The abundance changes of glycoproteins in the control (PBS) and treatment (LPS) experiments, and the normalized abundance changes are shown similar to Figure 3.3.

3.3.4 Comparison of surface glycoprotein responses from monocytes and macrophages to LPS

With the similar treatment in both THP-1 monocytes and macrophages, we then compared the responses at the surface glycoprotein level of the cells to LPS. A total of 116 glycoproteins were quantified commonly in both cell types with a determined response type (A, B, C, or D, but not N/A). Among those, 61 glycoproteins (53%) have the same response type. These proteins are involved in general processes such as heterotypic cell-cell adhesion and cell mobility. On the contrary, the responses in 55 glycoproteins (47%) participating in the regulation of defense response and detection of bacterium are different. Nevertheless, some glycoproteins involved in the leukocyte activation and innate immune response may have the same or different response type, such as NRP1 and CD115 where the abundances in monocytes increased gradually while those in macrophages stayed relatively similar or decreased, respectively.

We then compared the extent of the responses from both cell types. Some glycoproteins responded with the similar extent, and hence the same response type, in both cell types such as NPTN (Type B) and multidrug resistance-associated protein 1 (ABCC1; Type B) (Figure 3.8C). Several glycoproteins responded with different extents in the two cell types but followed the findings reported previously such as HLA-A, which has a greater abundance in macrophages after the LPS challenge (Type B in macrophages) but is not affected (Type D) in monocytes, corresponding well with previous reports regarding the antigen-presenting capability of macrophages.[207] Surprisingly, some glycoprotein abundances did not change as expected under the LPS treatment. For example, intercellular adhesion molecule 1 (ICAM1; CD54) was shown to be up-regulated in several previous

studies after the LPS treatment or under inflammatory conditions.[240-242] In this study, its abundance increased by over 8.4 times in monocytes but only 1.9 times in macrophages at 6 hours (Figure 3.8D). On the other hand, the abundance of ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1 (CD38) decreased in monocytes but increased in macrophages (Figure 3.8E).

Previous reports demonstrated that the differentiation of monocytes into macrophages prepared the cells for their functions in the immune processes. For example, the PMA treatment, which was used here to differentiate monocytes into macrophages, was found to increase the abundance of bound NF-κB in the cytoplasm (inactive form). Once the cells were challenged with LPS, more free NF-κB (active form) migrated into the nucleus, resulting in enhanced secretion of TNF-α and greater response to LPS.[243] Hence, we further applied our method to study the differentiation (discussed further in the next section), which revealed that the PMA treatment could also lead to the remodeling of the surface glycoproteome and could also prime the cells for the response in macrophages. For ICAM1, its surface abundance on monocytes is low before the LPS treatment and increased to >8 times after six hours so that they can participate in the response to LPS. The differentiation of monocytes into macrophages causes the abundance of ICAM1 to increase gradually to >5 times after 48 hours. Due to the already high surface glycoprotein abundance, macrophages may not need to express more ICAM1 to the surface as occurred in monocytes, and thus its abundance changes in macrophages after the LPS treatment are relatively lower compared with those in monocytes.

3.3.5 Cell-surface glycoprotein remodeling during the monocyte-to-macrophage differentiation

To better understand the different responses in THP-1 monocytes and macrophages to LPS, we further quantitated the surface glycoproteome changes during the monocyte-to-macrophage differentiation. As reported previously,[208] the differentiation was induced with phorbol myristate acetate (PMA; details in the supplemental information), and the surface glycoproteins were quantified as above. More than 900 glycopeptides from >300 surface glycoproteins were detected in the differentiation and the control experiments, with over 200 glycoproteins quantitated commonly in both experiments (Figure 3.11). Type A glycoproteins (27 glycoproteins, 12%) function in the response to lipoprotein particle, positive regulation of cytokine production, cell differentiation, and negative regulation of programmed cell death, indicating that the cells were transformed into the phenotype to respond to the infection. Type B proteins (28 glycoproteins, 12%) are involved in the regulation of ion transport, receptor internalization, and wound healing. Other glycoproteins (Type C, 117 glycoproteins, 51%) that decreased their abundances are involved in processes such as protein maturation and positive regulation of leukocyte proliferation. Lastly, the abundance of 32 glycoproteins (Type D, 14%) remained the same throughout the differentiation (Figure 3.12A and Figure 3.12B). Based on previously reported markers for the monocyte-to-macrophage differentiation, the current results clearly demonstrated that macrophages were obtained through the differentiation, including integrin alpha-M (ITGAM; CD11b), integrin alpha-X (ITGAX; CD11c), and ICAM1 that increased their abundances, CD280 and integrin alpha-4 (ITGA4; CD49d) that decreased their abundances, and CD36 and CD44 that increased their abundances abruptly at 6 hours (Figure 3.12C).[209, 244-245] Other quantitated surface glycoproteins are shown in Figure 3.13.

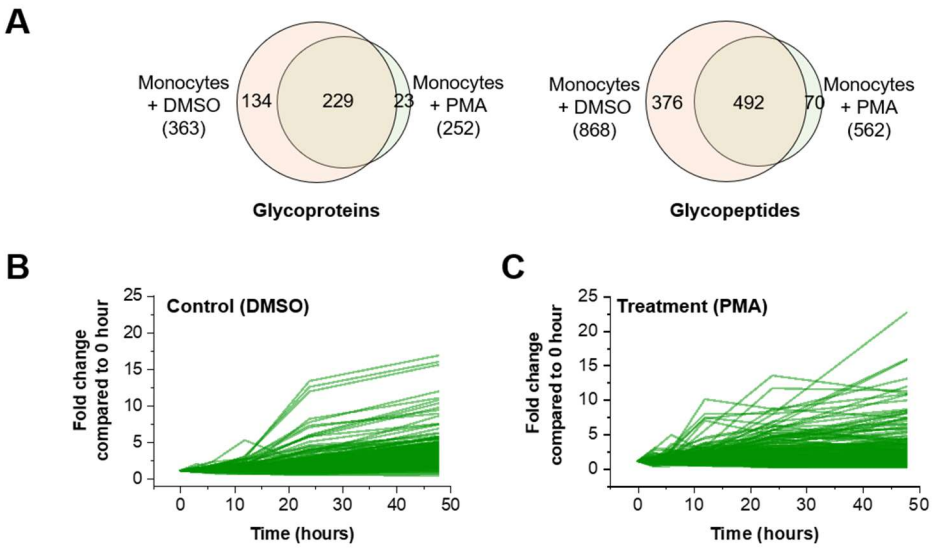

Figure 3.11. Quantification of cell-surface glycoproteins during the monocyte-to-macrophage differentiation. (A) Overlaps of glycoproteins and glycopeptides. (B-C) Profile plots for surface glycoproteins on monocytes treated with DMSO (B) or PMA (C).

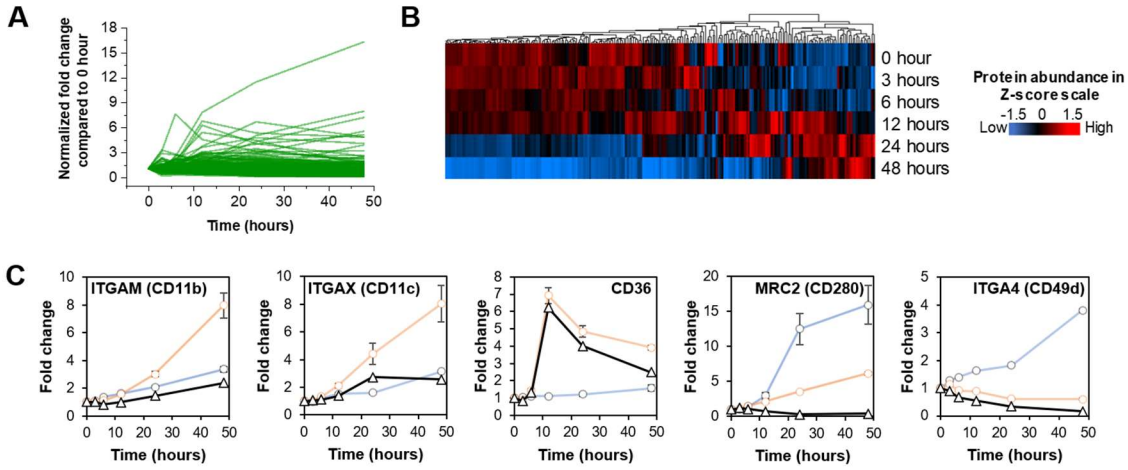

Figure 3.12. Surface glycoprotein analysis during the monocyte-to-macrophage differentiation. (A) Normalized abundances of cell-surface glycoproteins during the monocyte-to-macrophage differentiation. (B) A heat map based on hierarchical clustering with log2- and Z-score-transformed normalized abundances showing changes of cell-surface glycoproteins during the differentiation over time. (C) Surface glycoprotein markers for the monocyte-to-macrophage differentiation. The control (PBS), treatment (LPS), and normalized intensities are displayed similar to Figure 3.3. The error bars represent one standard deviation.

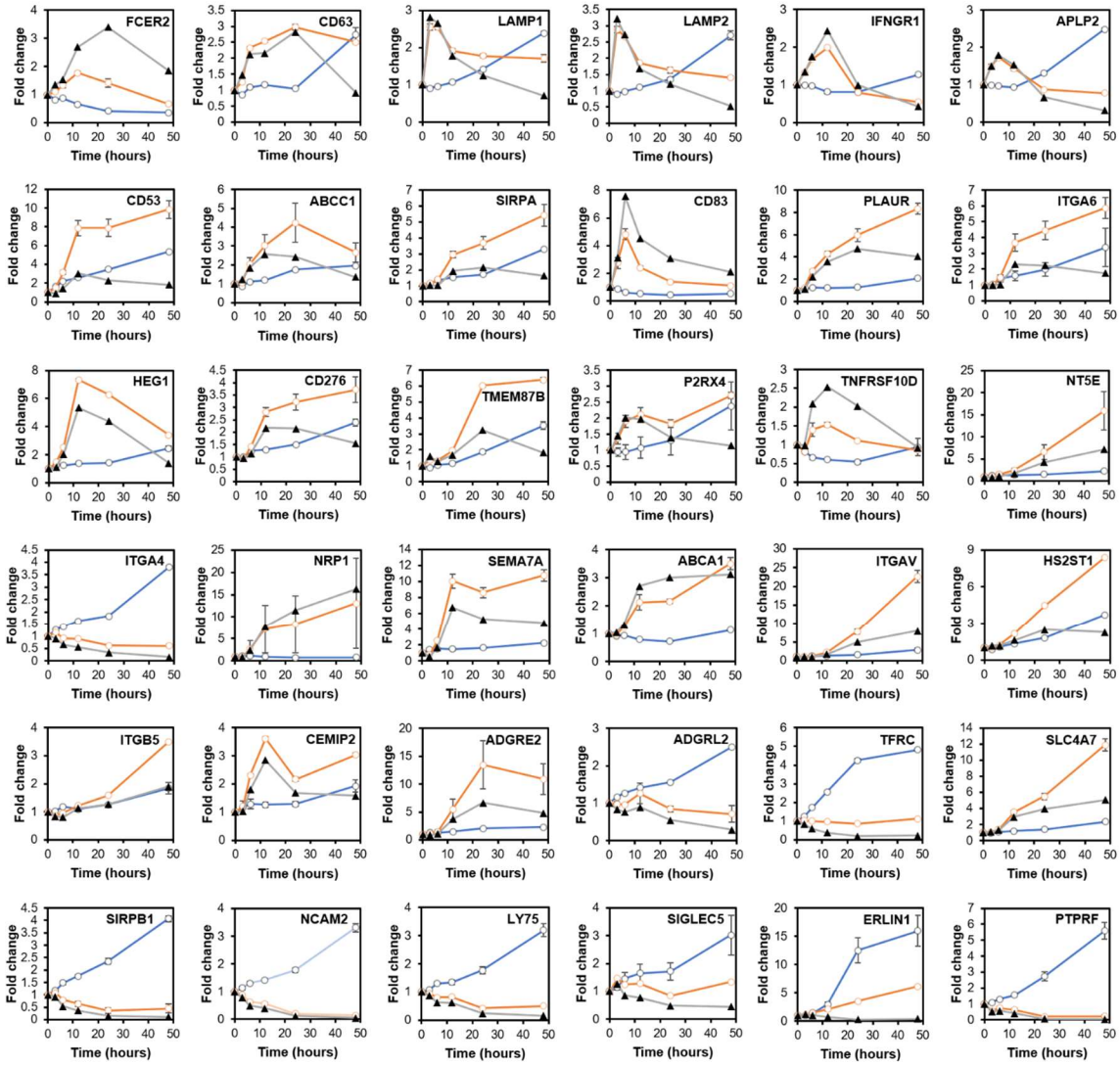

Figure 3.13. Example quantified cell-surface glycoproteins during the monocyte-to-macrophage differentiation. The control (PBS), treatment (PMA), and normalized intensities are shown similar to Figure 3.3.

3.3.6 Upstream regulators and downstream influences of regulated surface glycoproteins

Through global and time-resolved quantitation of surface glycoproteins in cells treated with LPS, we found the abundance changes of many surface glycoproteins including the transient changes. Potential upstream regulators of these regulated

glycoproteins and their downstream influences were further explored through Ingenuity Pathway Analysis (IPA) (Figure 3.14),[246] and the results should aid in a better understanding of the roles of these surface glycoproteins in the immune response. The LPS treatment clearly triggers the TLR4 signaling pathway as seen by the activation of NF-κB, MAPK3 (ERK1), MAPK1 (ERK2), and p38 MAPK in both cell types. This resulted in changes in surface glycoproteins that may signal and recruit other types of immune cells into the vicinity, as shown in the downstream effects including chemotaxis, infection of cells, interaction of mononuclear leukocytes, and infection of T-lymphocytes. Nevertheless, the IPA results demonstrate some different outcomes of the LPS challenge among the two cell types. The majority of regulators and downstream influences involved in the immune response were predicted at 12 hours in monocytes, and the production of reactive oxygen species was not predicted as a possible downstream effect in monocytes. This response was found to occur as fast as 6 hours in macrophages, which is possibly the result of the differentiation that primes the cell for the response to bacteria. Particularly at this time point, the transcription factors EGR1, SP1, RELA, and FOS, enzymes NOS2 and KRAS, and other proteins involved in the inflammatory process and response to molecules of bacterial origin are predicted to be activated, which subsequently resulted in the expression of more ICAM1, CD40, CD44, and CD55 on the cell surface as observed in the current study.[247-249] This eventually causes downstream effects including cell survival, synthesis of reactive oxygen species, and invasion of cells (Figure 3.14). At 6 hours after the LPS treatment, the majority of upstream regulators and downstream influences were also predicted to be in effect in macrophages similar to the maximum response time in a separate transcriptomic study.[208]

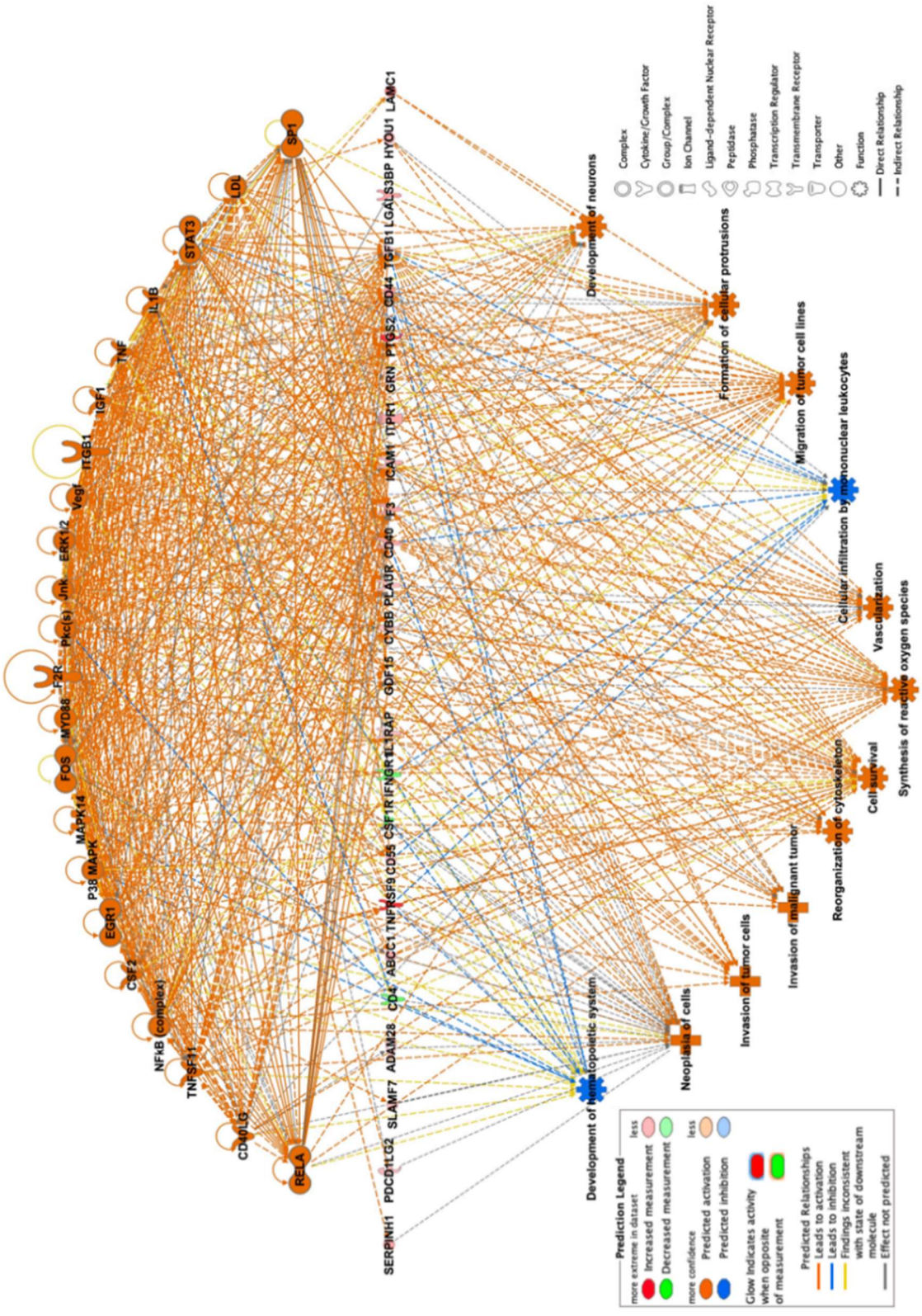

Figure 3.14. IPA analysis of upstream regulators and downstream effects of the regulated surface glycoproteins on macrophages treated with LPS for 6 hours.

For the PMA-induced monocyte-to-macrophage differentiation, most upstream regulators were predicted to be in effect at 12 hours. PMA was among the upstream regulators as expected. Known proteins involved in the differentiation of leukocytes as upstream regulators including F2, PAX5, and LDB1 were predicted. The IPA analysis demonstrates that the regulated surface glycoproteins could lead to downstream effects including cell attachment, reorganization of cytoskeleton, activation of antigen-presenting cells, and cell movement of phagocytes, indicating the change of phenotypes from monocytes to macrophages.

3.3.7 Site-specific analysis of cell-surface glycoproteins in response to different treatments

A total of 1,869 sites with the N-X-S/T/C motif were detected from all experiments, with several hundred sites identified exclusively in each cell type (Figure 3.15B). Distinct responses of different glycosylation sites on the same proteins were observed (Figure 3.15A and Figure 3.16). For instance, glycosylation on N229 of granulocyte-macrophage colony-stimulating factor receptor subunit alpha (CSF2RA) increased to 4.9 times in monocytes after the LPS treatment while that on N195 and N223 remained relatively similar throughout the treatment (Figure 3.15A). Some glycoproteins did not show site-specific changes, such as N150 and N522 on NRP1, which may result from the increased protein abundance on the surface. For ICAM1, glycosylation on N267 increased rapidly after 6 hours (16.7 times) compared with that on N145 and N183 (Figure 3.15C). The differentiation resulted in an increased glycosylation level on the site 267 by 21.9 times, and only 4.0 and 2.2 times for the other two sites, while the glycosylation level in macrophages for these sites are similar and lower than those from monocytes, suggesting

that N267 may be crucial for ICAM1 to function properly during the response. Different ICAM1 proteoforms can arise from alternative splicing, and specific glycosylation event further increases the complexity of the proteoforms.[250] For ICAM1, it was reported to increase its binding affinity with lymphocyte function-associated antigen 1 (LFA-1).[251-252] Thus, systematic and site-specific analysis of protein glycosylation facilitates a better understanding of glycoprotein functions and cellular immune responses.

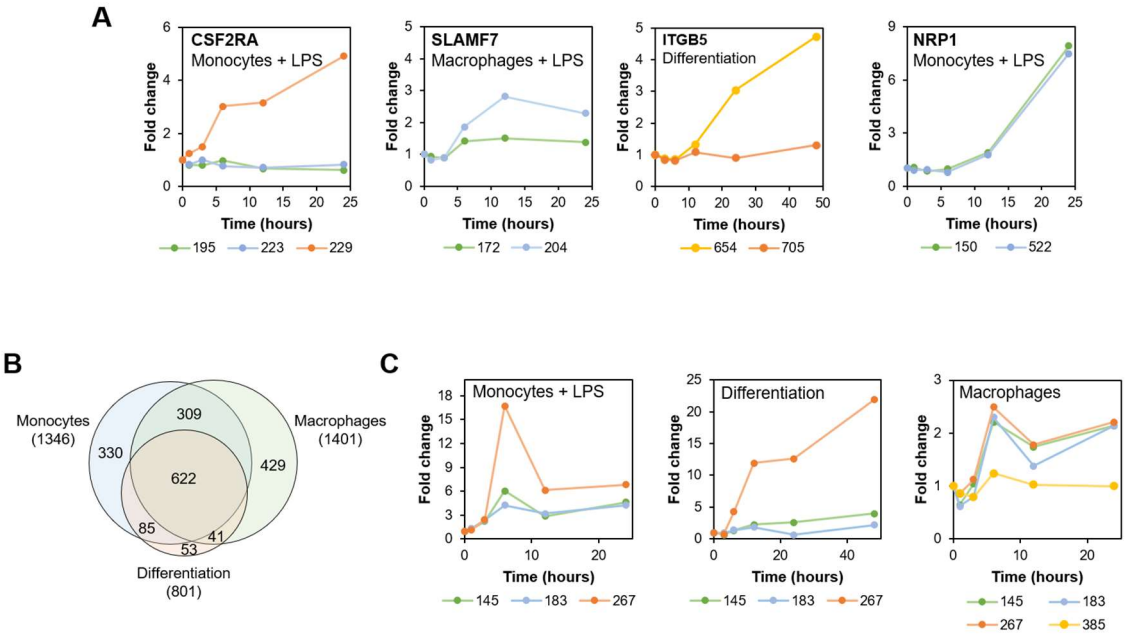

Figure 3.15. Site-specific quantitation of surface glycoproteins. (A) Examples of glycoproteins with multiple quantified glycosylation sites showing site-specific changes. The normalized fold change is displayed in the plots with the glycosylation sites annotated below the plot. (B) Overlap of glycosylation sites identified. (C) Site-specific glycosylation changes of ICAM1 from different processes in both monocytes and macrophages.

Bacterial infection can lead to inflammation,[253] fever, endotoxemia, septic shock,[254] liver damage,[255] and auto-immune diseases such as Crohn's[256-257] and Rheumatoid arthritis.[258] LPS has also been associated with diabetes[259-260] and neurological diseases including Parkinson's and Alzheimer's.[261-262] Depending on the environment and cell

types, various surface glycoproteins have an altered expression after the infection, as demonstrated in this work. Moreover, the expression changes of some surface glycoproteins such as receptors or transporters may also indicate specific infection within the body.[263-264] A previous report showed that the inflammatory monocyte subset is a valuable biomarker for human inflammatory diseases, including cardiovascular diseases, and may provide a potential therapeutic target for inflammatory monocytosis.[265] The up- or down-regulation of surface glycoproteins are crucial for the cellular response towards LPS/inflammation. Inhibition or promotion of these glycoproteins, the pathways regulating their expression, or the transcription factors responsible for their translation will inhibit the inflammation and affect the production of downstream cytokines.[266] The unprecedented and valuable information about the cell surface glycoprotein changes in the immune response can deepen our understanding of glycoprotein functions and cellular activities. More studies will allow us to identify some of them as disease biomarkers, which may be further developed as a clinical assay, and targets for drug development.

3.4 Conclusions

Cell-surface glycoproteins play crucial roles in the immune system. Prolonged inflammation has been related to autoimmune diseases such as Crohn's and Rheumatoid arthritis. The discovery of key surface glycoproteins that play roles in the process provides information that will lead to their discovery as drug targets that could suppress inflammation. Here, we comprehensively and site-specifically quantified the dynamics of glycoproteins only on the surface of monocytes and macrophages in response to LPS. The quantification results demonstrate cell-surface glycoprotein changes in monocytes and macrophages during the response, including the up-regulation of surface glycoproteins by

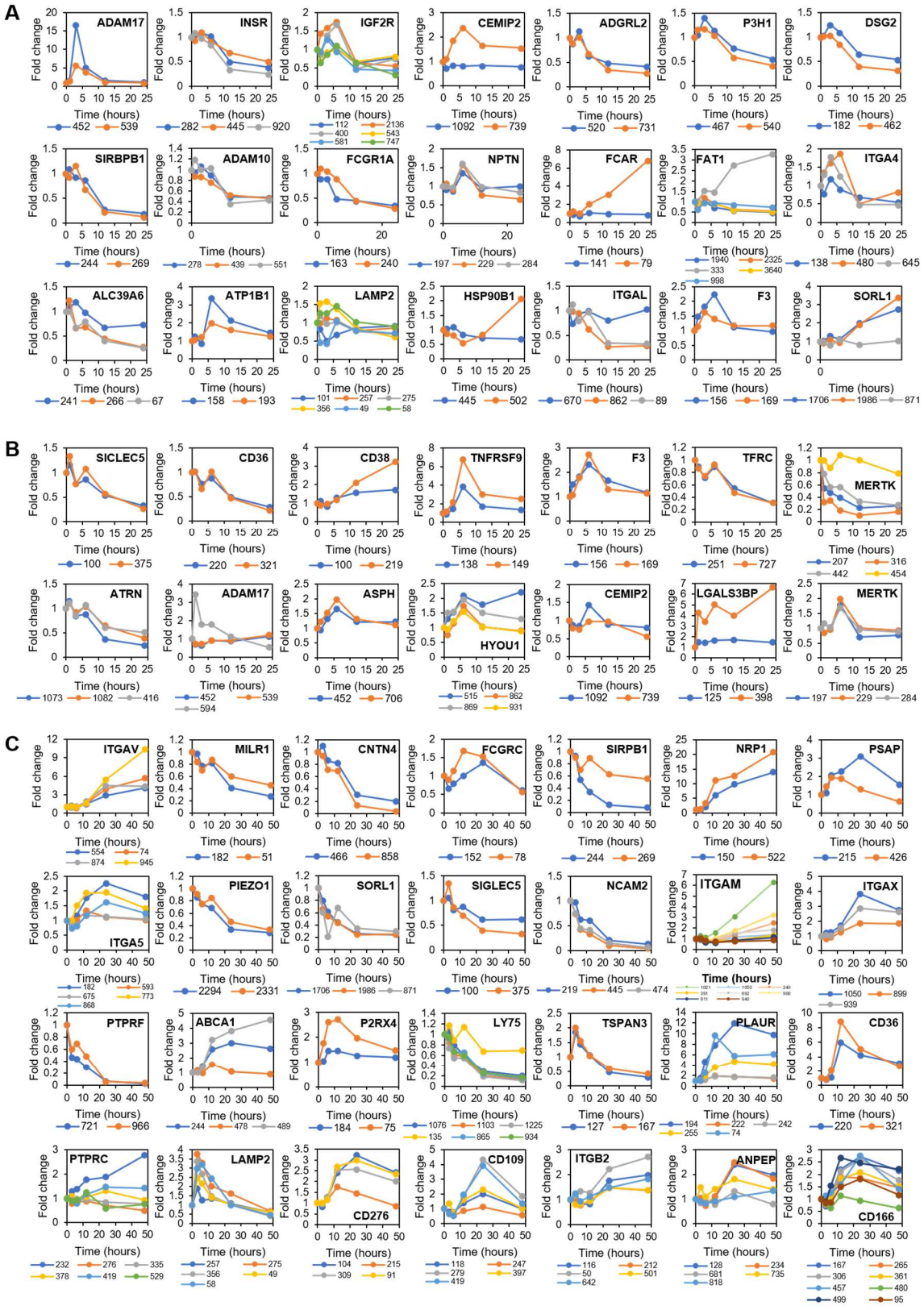
A
ADAM17
INSR
IGF2R
CEMIP2
ADGRL2
P3H1
DSG2
SIRBPB1
ADAM10
FCGR1A
NPTN
FCAR
FAT1
ITGA4
ALC39A6
ATP1B1
LAMP2
HSP90B1
ITGAL
F3
SORL1
B
SICLEC5
CD36
CD38
TNFRSF9
F3
TFRC
MERTK
ATRN
ADAM17
ASPH
HYOU1
CEMIP2
LGALS3BP
MERTK
C
ITGAV
MILR1
CNTN4
FCGRC
SIRPB1
NRP1
PSAP
ITGA5
PIEZO1
SORL1
SIGLEC5
NCAM2
ITGAM
ITGAX
PTPRF
ABCA1
P2RX4
LY75
TSPAN3
PLAUR
CD36
PTPRC
LAMP2
CD276
CD109
ITGB2
ANPEP
CD166
Fold change
Time (hours)

Figure 3.16. Site-specific quantification of cell-surface glycoproteins. (A) Site-specific quantification of surface glycoproteins on monocytes treated with LPS. (B) Site-specific quantification of surface glycoproteins on macrophages treated with LPS. (C) Site-specific quantification of surface glycoproteins during the monocyte-to-macrophage differentiation. Example quantified surface glycoproteins are shown, and the glycosylation sites are annotated below each figure.

gradually increasing their abundances over time or abruptly elevating their abundances before returning to the normal level. The responses are found to be different for both monocytes and macrophages. Besides the well-documented glycoprotein changes in response to LPS, we also identified some new surface glycoproteins participating in the immune response such as APMAP, TSPAN3, and IGSF8. Moreover, the current work provides site-specific information regarding protein glycosylation changes during the LPS treatment. The quantification of glycoproteins in the differentiation experiment from monocyte to macrophage revealed that the different responses were at least partly attributed to the priming of monocytes during the differentiation. Systematic investigation of the dynamics of surface glycoproteins results in a better understanding of glycoprotein functions and cellular immune responses, leading to the identification of surface glycoproteins as biomarkers and drug targets.

Table 3.1. Example surface glycoproteins quantitated from THP-1 monocytes and macrophages in response to LPS or in the monocyte-to-macrophage differentiation.

Experiment	UniProt ID	Gene Symbol	Type	Annotation
Monocytes	Q9BZZ2	SIGLEC1	A	Sialoadhesin
	Q99571	P2RX4	B	P2X purinoceptor 4
	P55899	FCGRT	C	IgG receptor FcRn large subunit p51
	Q15262	PTPRK	D	Receptor-type tyrosine-protein phosphatase kappa
Macrophages	P28907	CD38	A	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1
	P13726	F3	B	Tissue factor
	Q9UKQ2	ADAM28	B	Disintegrin and metalloproteinase domain-containing protein 28
	P20138	CD33	C	Myeloid cell surface antigen CD33
	P15260	IFNGR1	C	Interferon gamma receptor 1
Differentiation	O14786	NRP1	A	Neuropilin-1
	P18084	ITGB5	A	Integrin beta-5
	Q9Y6M7	SLC4A7	A	Sodium bicarbonate cotransporter 3
	Q01151	CD83	B	CD83 antigen
	P33527	ABCC1	B	Multidrug resistance-associated protein 1
	P02786	TFRC	C	Transferrin receptor protein 1

CHAPTER 4. QUANTIFICATION OF THE PROTEOME AND SURFACE GLYCOPROTEOME DYNAMICS DURING THE PARTIAL EPITHELIAL-TO-MESENCHYMAL TRANSITION (EMT) STATE

4.1 Introduction

Epithelial-to-mesenchymal transition (EMT) refers to the process where epithelial cells lose their cell-cell adhesion and polarity and transform through biochemical and physical processes into mesenchymal cells that are more mobile, more invasive, and more resistant to apoptosis.[267] The process is integral for biological processes including embryonic development, wound healing, and cancer metastasis.[268-273] EMT is characterized by several core changes, including skeleton remodeling, loss of apical-basal cell polarity, cell-cell adhesion weakening, cell-matrix adhesion remodeling, cell individualization, establishment of front-back polarity, and basement membrane invasion.[274] Protein makers for the transition include the loss of E-cadherin and the increased N-cadherin on the cell surface, the increased cytoskeletal vimentin and beta catenin, the increased fibronectin and laminin 4 and the decreased laminin 1 in the extracellular matrix.[275] Several inducers for the process have been found in the literature, including epidermal growth factor (EGF) and tumor necrosis factor (TNF).[276] The most notorious one would be transforming growth factor beta (TGF-β). TGF-β initiates the transition through surface TGF-β type I and type II receptors that subsequently activate the SMAD pathway. This eventually results in the regulation of transcription factors such as SNAIL and ZEB, and the transformation of epithelial to mesenchymal cells.[277]

While EMT was previously thought to be the transition between the two states, recent studies have shown that the partial EMT state is metastable and can be observed experimentally.[278-281] Zhang et al. demonstrated that the process occurs through a stepwise activation of two double-negative feedback loops, i.e., between SNAIL1 and miR34 and between ZEB1 and miR-200, following the cascading bistable switches model and resulting in the observable partial EMT state.[282] Interestingly, the authors showed that cells in the full mesenchymal state cannot revert to the partial or epithelial state, but those in the partial EMT state can revert back to the epithelial state after a reculture experiment, showing that the full mesenchymal state might not occur *in vivo* due to the requirement that cells must be able to propagate again. This was supported by other studies, such as the work from Aiello et al. showing that carcinoma cells migrate as clusters containing cells in the partial state, instead of individual cells in the full mesenchymal state.[283]

Mesenchymal cells are not easily observed *in vivo*. Several groups have created devices to capture these cells.[284-285] Nevertheless, many cell lines have been shown to undergo EMT *in vitro*.[286-290] Several studies employ transcriptomics analysis, but the knowledge at the proteomics scale, especially of the partial EMT state, is very limited.[291-294] Several post-translational modifications (PTMs) are involved in EMT, including phosphorylation, glycosylation, ubiquitination, hydroxylation, and acetylation.[295-299] Surface glycoproteins play essential roles in the regulation of nearly every extracellular event.[300-301] This includes the cadherin proteins that have been used as surface markers for the transition mentioned previously. Yet, the global view of cell-surface glycoprotein changes during the transition, especially at the partial state, has never been reported.

In this work, we study the EMT process in A549 cells treated with TGF-β in a time-resolved manner. The proteomics analysis revealed proteins that are involved in the epithelial, mesenchymal, as well as partial states. Comparing with transcriptomics results, several novel partial EMT drivers were discovered, including JUNB that was found to be the regulator of the partial EMT state. Surface glycoproteomics results showed several surface glycoproteins that are involved in the transition, but the involvement of sialic acids on the surface currently needs to be explored further. The results will lead to a better understanding of the EMT process and their use as markers for cellular development and cancer metastasis.

4.2 Materials and methods

4.2.1 Cell culture and EMT induction

All reagents were purchased from Sigma unless noted otherwise. A549 cells (ATCC) were maintained in high-glucose DMEM medium supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin at 37 ºC in a humidified incubator with 5% CO_2. The cells were passaged equally to 10 T25 flasks for the proteomics experiment or T175 flasks for the surface experiment and grown until the density reached ~80%. The cells were treated with 4 ng/mL human transforming growth factor beta 1 (hTGF-β1; Cell Signaling Technology) for 0, 12, 24, 48, and 96 hours in biological duplicate experiments. The medium was replaced every two days. At each time point, the cells were washed with ice-cold PBS, scraped in ice-cold PBS, collected, pelleted, and stored at -80 ºC until the next steps.

4.2.2 Whole proteomics analysis

In the time-resolved proteomics experiments, the cells were thawed on ice. The lysis buffer containing 0.5% sodium deoxycholate (SDC), 150 mM sodium chloride (NaCl), 50 mM N-(2-hydroxyethyl)piperazine-N'-2-ethanesulfonic acid (HEPES, pH=8.8), and 1 tablet/10 mL cOmplete ULTRA tablets protease inhibitor cocktail (Roche) was prepared and chilled on ice until use. The cells were lyzed at 4 ºC with end-over-end rotation for 1 hour. The lysates were clarified by centrifugation at 25,830*g* for 10 minutes and the debris were discarded. The proteins were reduced with 5 mM dithiothreitol at 56 ºC for 30 minutes, cooled down to room temperature, and alkylated with 14 mM iodoacetamide in the dark for 30 minutes. The proteins were purified using the methanol/chloroform precipitation method. The pellets were dried and dissolved in a digestion buffer containing 50 mM HEPES, 1.6 M urea, and 5% acetonitrile (ACN). The proteins were digested using sequencing-grade modified trypsin (Promega) at the enzyme:protein ratio of 1:100 overnight at 37 ºC with shaking. The mixtures were acidified to quench the digestion with trifluoroacetic acid (TFA, EMD Millipore) to 0.4% and the pH was checked to be lower than 2. The mixtures were clarified by centrifugation and the peptides were desalted using a Sep-pak tC18 cartridge (Waters). The peptides were dried using a speedvac.

For the quantification, about 100 µg of the peptides from each time point (measure using the bicinchoninic assay) was dissolved in 10 µL ACN and 33 µL 200 mM pH=8.8 HEPES. Each of the 10-plex TMT reagents (Thermo) was dissolved in 41 µL anhydrous ACN. The peptides were labeled with the reagent as followed: 126-0 hour, expt. 1; 127N-12 hour, expt. 1; 127C-24 hour, expt. 1; 128N-48 hour, expt. 1; 128C-96 hour, expt. 1; 129N-0 hour, expt. 2; 129C-12 hour, expt. 2; 130N-24 hour, expt. 2; 130C-48 hour, expt.

2; 131- 96 hour, expt. 2. The labeling proceeded for 1 hour at room temperature. The reaction was quenched using 5 µL of 8% hydroxylamine hydrochloride solution at room temperature for 15 minutes. The peptides were combined, dried using the speedvac, desalted with the Sep-pak cartridge, and dried again using the speedvac.

The TMT-labeled peptides were fractionated with high-pH chromatography using an Agilent 1260 Infinity HPLC and an XBridge C18 (3.5 µm, 4.6x250 mm) column (Waters). Buffer A was 10 mM ammonium formate (pH=10) in water and buffer B was 10 mM ammonium formate (pH=10) in 90% acetonitrile. The peptides were fractionated and collected during an 80-minute gradient of 5% to 70% buffer B and the flow rate of 0.7 mL/minute. The peptides were collected every 2 minutes, consolidated into 25 fractions, and dried using in a vacuum concentrator. The dried peptides were purified with StageTip and dried before LC-MS/MS analysis.[161]

4.2.3 Enzymatic labeling of surface glycoproteins

For the surface experiment, the cells were treated with TGF-β similar to the proteomics experiment and collected at 0, 12, 24, 48, 72, 96 hours. At each time point, the cells were treated with 250 mU/mL neuraminidase (sialidase) in PBS at 37 ºC for 30 minutes. The cells were then scraped and pelleted. Glycans of proteins on the cell surface were oxidized using a solution containing 5% FBS, 50 U/mL galactose oxidase (Innovative Research), and 40 U/mL horseradish peroxidase at 37 ºC with end-over-end rotation for 1 hour. The cells were pelleted and stored at -80 ºC until the next steps.

Once all the cells are collected, they were thawed on ice. The cells were treated with a solution containing 150 mM NaCl, 50 mM HEPES pH=8.8, 25 µg/mL digitonin and

1 tablet/10 mL cOmplete ULTRA tablets protease inhibitor cocktail on ice for 10 minutes. The cells were centrifuged at 2,000*g* for 10 minutes and the cell pellets were lyzed with the same lysis buffer in the proteomics experiment. The lysates were clarified by centrifugation. One hundred microliters of hydrazide beads (Thermo) were washed with water twice and added to each lysate. Aniline was also used as a catalyst at the concentration of 10 mM. The enrichment proceeded at 4 ºC with end-over-end rotation. After 24 hours, the beads were washed twice with a solution containing 8 M urea, 0.4 M ammonium bicarbonate, and 10% sodium dodecyl sulfate (SDS). The proteins were reduced on-beads with 5 mM DTT at 37 ºC, 45 minutes with end-ever-end rotation, and subsequently alkylated with 14 mM iodoacetamide at room temperature for 30 minutes in the dark. The beads were washed with the washing solution four more times with a 5-minute incubation on the end-over-end at each wash. The beads were resuspended in the digestion buffer and the proteins were digested with trypsin similarly.

The next day, the beads were washed with 1) 80% ACN, 0.1% TFA, 2) wash solution, 3) dimethyl formamide, and 4) 0.1 M ammonium bicarbonate, 5 times each. The enriched peptides were eluted using a solution containing 0.2 M methoxylamine hydrochloride, 1.5 M NaCl, and 0.1 M aniline in 0.1 M sodium acetate solution pH=4.5 at 37 ºC for 30 and 60 minutes. The eluants were pooled, desalted with the Sep-Pak cartridge, and dried using the speedvac. The dried peptides were subsequently labeled with the TMT 6-plex reagents as followed: 126-0 hour; 127-12 hours; 128-24 hours; 129-48 hours; 130-72 hours; 131-96 hours. After the combined peptides are purified, they are deglycosylated using 3 U PNGase F in $H_2{}^{18}O$ for 3 hours with shaking. The peptides were purified with

StageTip into three fractions using 20%, 50%, and 50% ACN with 1% FA and dried before LC-MS/MS analysis.

4.2.4 LC-MS/MS analysis

The peptides were analyzed with an LTQ Orbitrap Elite Hybrid Mass Spectrometer (Thermo). The peptides were first separated with reverse-phase chromatography on a Dionex UltiMate 3000 UHPLC system (Thermo) using a 112-minute gradient of 4-17% of ACN with 0.125% formic acid for all fractions of the proteomics experiment, and 3-14%, 4-17%, and 8-24% for the three fractions of the surface experiment. The microcapillary column used was packed in-house (Magic C18AQ, 3 μm, 200 Å, 75 μm*16 cm). MS/MS analysis was performed using a Top15 method. For each cycle, a full MS spectrum was recorded in the Orbitrap at the resolution of 60,000 with 1 M automatic gain control (AGC). Peptide fragmentation was performed using higher-energy collisional dissociation (HCD) with a normalized collisional energy of 40%. Up to 15 MS/MS were performed for the 15 most abundant peaks with an isolation width of 1.2 m/z and recorded in the Orbitrap cell at the resolution of 30,000 for the proteomics experiment or 15,000 for the surface glycoproteomics experiment and $2x10^5$ AGC. Selected ions were excluded from further sequencing for 90 seconds. Ions with singly or unassigned charge were not sequenced. Maximum ion accumulation times were 1,000 ms for each full MS scan and 50 ms for one MS/MS scans.

4.2.5 Data analysis and bioinformatics

The raw files were converted into the mzXML format and searched using the SEQUEST algorithm (version 28).[217] The spectra were matched against the human protein

sequences downloaded from UniProt database (www.uniprot.org). The peptide mass tolerance was 20 ppm. And The fragment ion mass tolerance was 0.025 ppm. The maximum number of missed cleavages was 3 and the maximum number of differential modifications was 4. The following amino acid modifications were used; static modifications of +57.0215 Da for carbamidomethylation of cysteine and +229.1629 Da for TMT labeling at lysine and the N-terminus, and differential modification of +15.9949 Da for oxidation of methionine. For the glycoproteomics experiment, additional differential modifications of +2.9883 Da on asparagine were included for the N-glycosylation sites. The target-decoy method, where each protein sequence is listed in both forward and reverse orders, was used to estimate the false discovery rates (FDRs) of peptide and protein identifications. Linear discriminant analysis was used to control the quality of peptide identification using parameters such as XCorr, ppm, peptide length, and charge state. The FDR for peptide identification was controlled to <1%, and an additional filter was performed at the protein level to control the protein identification FDR to <1%. For the glycopeptide analysis, the dataset was limited to only glycopeptides and glycoproteins when calculating the FDRs at both levels.

Data were processed with Microsoft Excel and Perseus.[165] Contaminants and falsely identified peptides were removed from the raw results. The signal-to-noise (S/N) ratios of each of the eight TMT channels were recorded and corrected according to the manufacturer's instruction for the quantification experiment. The TMT peak match tolerance was 0.003 Da and the nearest m/z peak in that tolerance window was chosen for each reporter ion. Peptides with S/N less than 5 were removed. The signals from all channels were normalized using the protein abundance from tubulin beta chain (TUBB),

β-actin (ACTB), and glyceraldehyde 3-phosphate dehydrogenase (GAPDH). The protein abundance changes were eventually calculated from the median TMT S/N of all peptides for each protein, and the final protein intensity was from the average of the two replicates. Gene-ontology (GO)-based clustering of proteins was performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID).[302] For the glycoproteomics experiment, a probabilistic-based algorithm called Ascore was used to determine the confidence of the glycosylation site localization.[219] Ascore of at least 19 shows a well-localized site with 99% confidence while the score of at least 13 shows 95% confidence.

4.3 Results and discussion

4.3.1 Morphology of A549 cells during the epithelial-to-mesenchymal transition

Studies have reported that A549 cells undergo EMT *in vitro* after TGF-β treatment.[303-306] We first verified the transition by observing the morphology under microscopy. A549 cells were cultured and treated with PBS as a control or with 4 ng/mL TGF-β over four days. The medium was replaced every two days to keep the concentration of TGF-β high. The morphology difference between the two groups can be observed as soon as only 24 hours after the treatment (Figure 4.1). As time progressed, cells in the control group remained rounded and packed tightly with other cells. On the contrary, cells in the TGF-β-treated group are more elongated and appeared more spread in the flask. The morphology observed in the experiment correlate with the previously reported changes in cells undergoing the EMT process.[307]

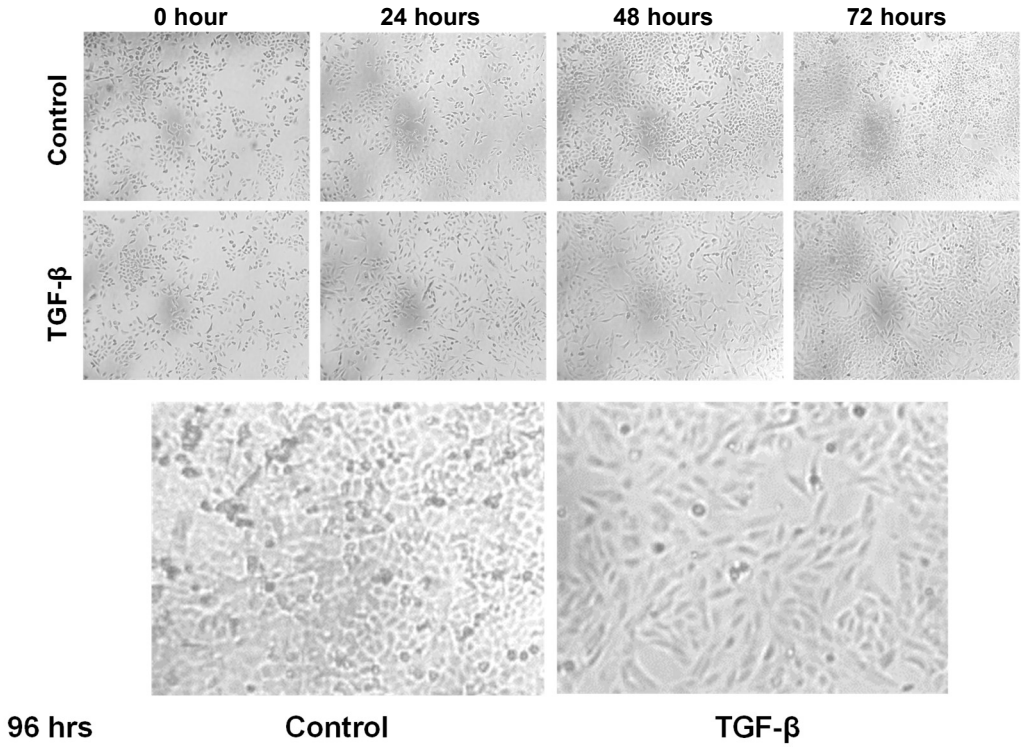

Figure 4.1. Images of A549 cells treated with PBS (control) or TGF-β over 4 days under light microscope.

4.3.2 Proteomics analysis of A549 cells during the epithelial-to-mesenchymal transition

A549 cells treated with 4 ng/mL TGF-β were collected after 0, 12, 24, 48, and 96 hours. The proteins were extracted and digested, and the resulting peptides were quantified from biological duplicate experiments using the TMT 10-plex (Figure 4.2). We identified a total of 6,317 proteins from 50,075 unique peptides. After protein filtering, 6,294 proteins were quantified across the ten samples at the five time points. With ANOVA, we found 3,895 proteins with a significant abundance changes in at least one of the time points. Principle component analysis shows that the two biological duplicates at each time point cluster as time progresses (Figure 4.3A). The abundance of proteins over 96 hours of the EMT process is shown in the heatmap after $\log_2$ and Z-score transformation (Figure 4.3B). Hierarchical clustering of the protein expressions demonstrated that the proteins clustered

into three major groups, including those that decreased their abundance throughout the transition (designated as epithelial, N=608), increased their abundance before decreased to the normal level (designated as partial, N=267), decreased their abundance throughout the transition (designated as mesenchymal, N=3020). Clustering of epithelial proteins revealed that they participate in ribosome biogenesis (N=49, P=2.5E-20) and mitotic cell cycle (N=73, P=1.2E-13). For partial proteins, they participate in the protein localization to membrane (N=26, P=3.3E-12). The number of proteins in the mesenchymal cluster is too high to be clustered by DAVID.

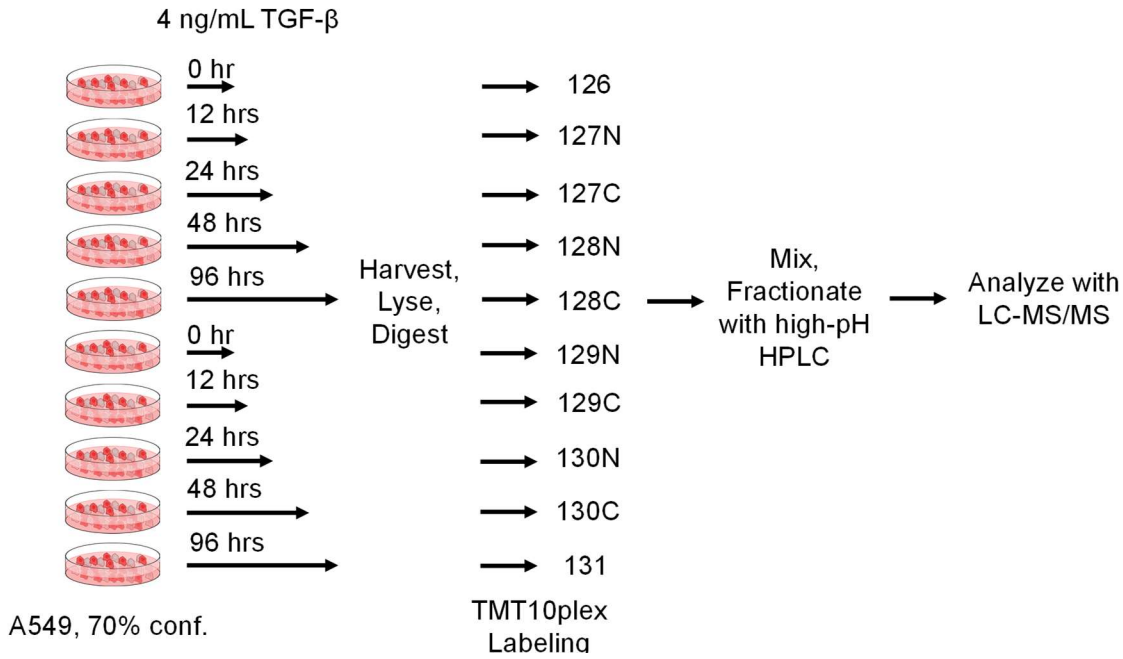

Figure 4.2. Experimental workflow for proteomics analysis of A549 cells during the epithelial-to-mesenchymal transition.

To confirm that the EMT occurred in our experiment, we investigated the abundance changes of several protein markers for the transition that have been reported previously in the literature. This includes mothers against decapentaplegic homolog 3 (SMAD3), vimentin (VIM), fibronectin (FN1), and cadherin-2 (CDH2) (Figure 4.4).[308]

The abundance changes of these proteins showed the transition of the cells during EMT in our study.

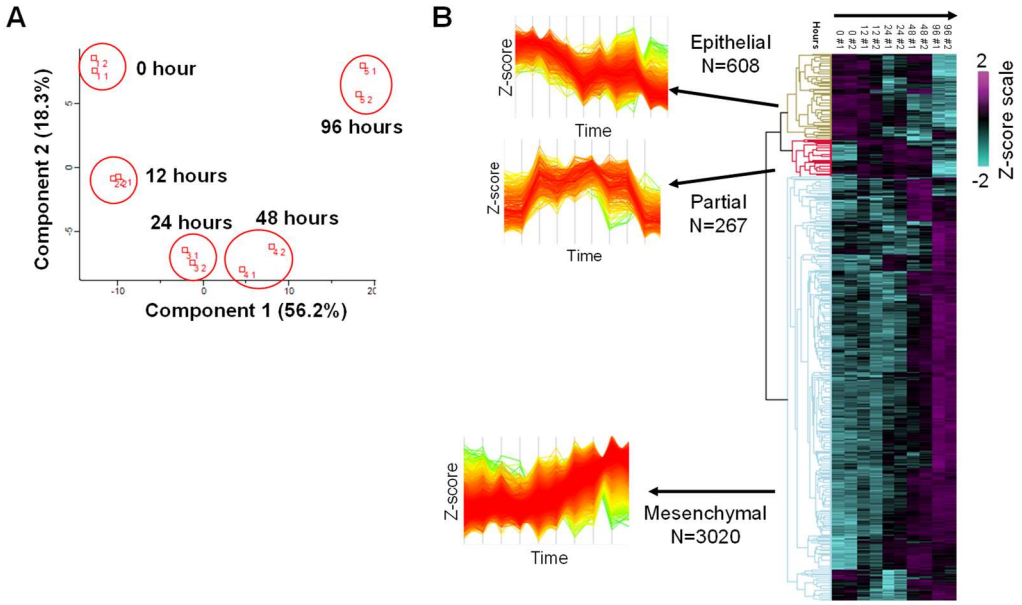

Figure 4.3. Proteomics quantification in cells during EMT. (A) PCA analysis of the protein abundance at the five time points. (B) Heat map of the protein abundance (log_2- and Z-score-transformed). The insets show the abundance profiles of proteins in the epithelial, partial, or mesenchymal groups.

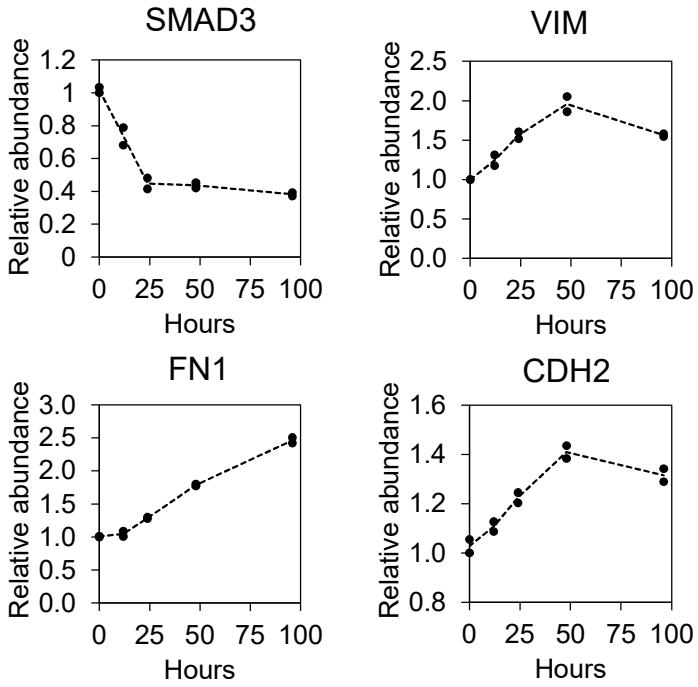

Figure 4.4. Abundance changes of example EMT protein markers from the proteomics experiment.

4.3.3 Proteins in the partial EMT state

Previously, Chang et al. performed a transcriptomics analysis of A549 cells during EMT and identified proteins that associate with super-enhancers of the partial EMT state, including ETS2, HNF4A, and JUNB.[309] The authors also demonstrated transcripts that are highly up-regulated in the partial EMT state. We extracted the partial EMT protein information from this work and compared it with the partial EMT proteins profiled from our study. Thirty-four proteins overlapped from the two experiments (Table 4.1). Clustering analysis showed that these proteins are involved in the embryonic development (N=7, P=6.8E-3) and cell surface receptor signaling pathway (N=11, P=1.7E-2). Among the 34 proteins, JUNB was also assigned to the partial group. In the study by Chang et al., the protein was reported to be the driver of the partial EMT state. This indicates that the partial EMT was achieved in this study at ~12-24 hours post-treatment. Other proteins quantified may also play important roles during the partial EMT state, such as SERPINE1 and ECE1 (Figure 4.5). Both have been reported to participate in the EMT process previously but have never been linked to the partial state.[310-311]

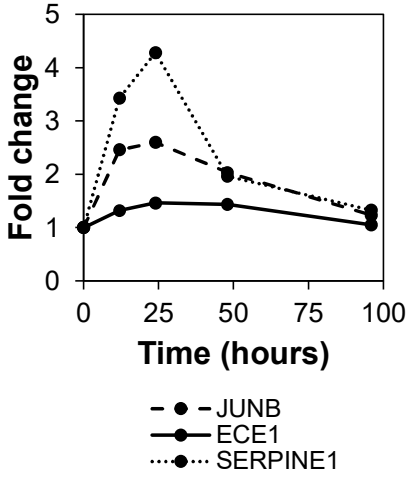

Figure 4.5. Abundance of proteins involved in the partial EMT state. JUNB was previously reported to be the driver of partial EMT.

Table 4.1. Partial EMT proteins quantified commonly from this study and from the RNA-seq experiment by Chang et al.

UniProt ID	Gene name	Protein name
Q96LD4	TRIM47	E3 ubiquitin-protein ligase TRIM47
P05121	SERPINE1	Plasminogen activator inhibitor 1
Q6QNY1	BLOC1S2	Biogenesis of lysosome-related organelles complex 1 subunit 2
Q9UHR4	BAIAP2L1	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1
P42892	ECE1	Endothelin-converting enzyme 1
P17275	JUNB	Transcription factor jun-B
P11362	FGFR1	Fibroblast growth factor receptor 1
Q86XL3	ANKLE2	Ankyrin repeat and LEM domain-containing protein 2
Q16222	UAP1	UDP-N-acetylhexosamine pyrophosphorylase
A6QL63	BTBD11	Ankyrin repeat and BTB/POZ domain-containing protein BTBD11
P53396	ACLY	ATP-citrate synthase
O14907	TAX1BP3	Tax1-binding protein 3
Q9Y6R0	NUMBL	Numb-like protein
Q9UKS6	PACSIN3	Protein kinase C and casein kinase substrate in neurons protein 3
Q9C002	NMES1	Normal mucosa of esophagus-specific gene 1 protein
Q8WUF5	PPP1R13L	RelA-associated inhibitor
Q86UU1	PHLDB1	Pleckstrin homology-like domain family B member 1
Q6ZSZ5	ARHGEF18	Rho guanine nucleotide exchange factor 18
P17812	CTPS1	CTP synthase 1
P11166	SLC2A1	Solute carrier family 2, facilitated glucose transporter member 1
O95379	TNFAIP8	Tumor necrosis factor alpha-induced protein 8
P17655	CAPN2	Calpain-2 catalytic subunit
Q92536	SLC7A6	Y+L amino acid transporter 2
Q6Y1H2	HACD2	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 2
Q01813	PFKP	ATP-dependent 6-phosphofructokinase, platelet type
P31947	SFN	14-3-3 protein sigma
Q6NYC1	JMJD6	Bifunctional arginine demethylase and lysyl-hydroxylase JMJD6
Q96EA4	SPDL1	Protein Spindly
Q8N128	FAM177A1	Protein FAM177A1
P33764	S100A3	Protein S100-A3
Q9NNX1	TUFT1	Tuftelin
P35240	NF2	Merlin
O95633	FSTL3	Follistatin-related protein 3
A6ND36	FAM83G	Protein FAM83G

4.3.4 Time-resolved analysis of surface glycoprotein changes during EMT

Once the transition in A549 cells especially the partial EMT state is shown through the time-resolved whole-cell proteomics experiments, we next investigated the changes at the surface glycoproteome level. This required a separate experiment from the whole proteomics analysis because of the low abundance of many surface glycoproteins and the specificity for targeting glycoproteins on the cell surface part since some surface glycoproteins may be synthesized and stored inside of cells. Here, we employed a method developed by Sun et al. to target surface glycoproteins using galactose oxidase to convert hydroxyl groups on Gal or GalNAc into carbonyl groups. These glycoproteins can then be enriched using hydrazide chemistry and quantified by multiplexed proteomics.[35]

Similar to the time-resolved whole-cell proteomics experiment, the cells were treated with 4 ng/mL TGF-β and collected over 0, 12, 24, 48, 72, and 96 hours. For the surface experiments, the TMT 6-plex reagents were used instead of the 10-plex reagent in the proteomics experiment to expand the time points that we can observe the cells. At each time point, the cells were collected, and surface glycoproteins were oxidized with galactose oxidase, enriched, and digested. The resulting peptides were labeled with the TMT 6-plex reagents and analyzed with LC-MS/MS.

In our trial experiment, the coverage of surface glycoproteins was found to be lower than expected. It has been reported that EMT can also up-regulate the presence of sialic acid on surface glycoproteins.[312] Therefore, we performed an optimization of the surface glycoprotein enrichment methods, including the removal of sialic acid by pretreating the cells with sialidase, the enrichment without SDS, and the enrichment at the peptide level.

The optimization showed that the pretreatment with sialidase is the only condition that improved the coverage of surface glycoproteins (Figure 4.6).

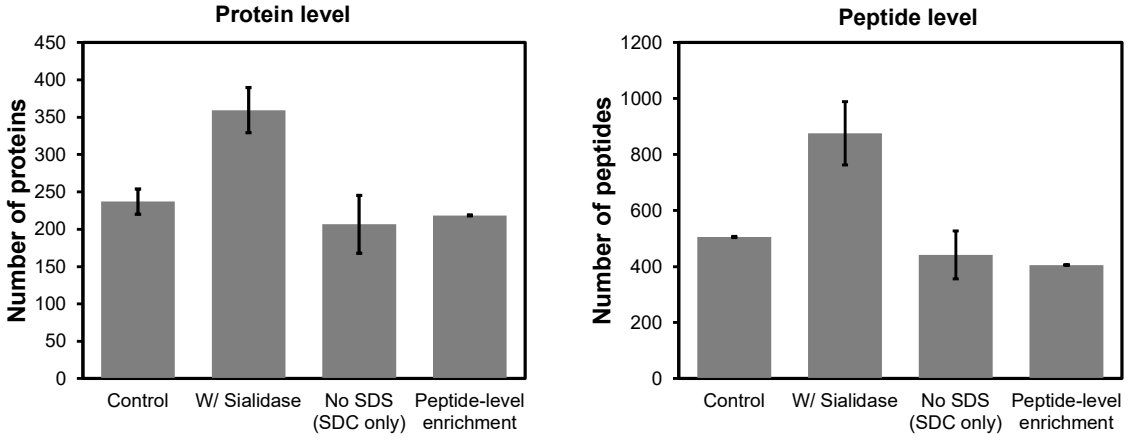

Figure 4.6. Optimization of surface glycoprotein enrichment compared with the original protocol reported by Sun et al.

We performed the experiment in biological duplicates for cells treated with TGF-β or with PBS as a control. The peptide abundance from both groups were normalized and shown in Figure 4.7. Overall, 358 surface glycoproteins were quantified from both the treated and control duplicate experiments. After normalization between the control and treated groups, 243 surface glycoproteins were quantified commonly (Figure 4.7). Among the total surface glycoproteins detected, 263 proteins are clustered to the membrane part with a very low P-value (1.7E-88). While proteomics results showed distinct proteins that are up-regulated during the partial EMT state, the current surface glycoproteomics analysis did not yield such distinct results. Therefore, the quantified glycoproteins were separated into three groups based on the abundance at the last time point ($\log_2>1$, $\log_2<1$, and others). Most changes occurred between 12 to 24 hours. Clustering analysis shows that surface glycoproteins that were up- or down-regulated are involved in cell migration and cell motility. We further investigated the abundance changes from each individual experiment

without the normalization. Interestingly, the abundance of surface glycoproteins is lower at 24 hours compared to that at 12 hours (Figure 4.8). A comparison of the results from all four experiments (duplicate treated and control) revealed that over 60% of the same glycoproteins have a lower abundance at 24 hours, and thus this might not occur randomly. The current hypothesis for the observed result is that the sialidase treatment might not be able to completely remove sialic acid on surface glycoproteins, resulting in the lower protein abundance observed after 24 hours. The future experiment is to look into the optimization of the sialic acid removal by the sialidase enzyme.

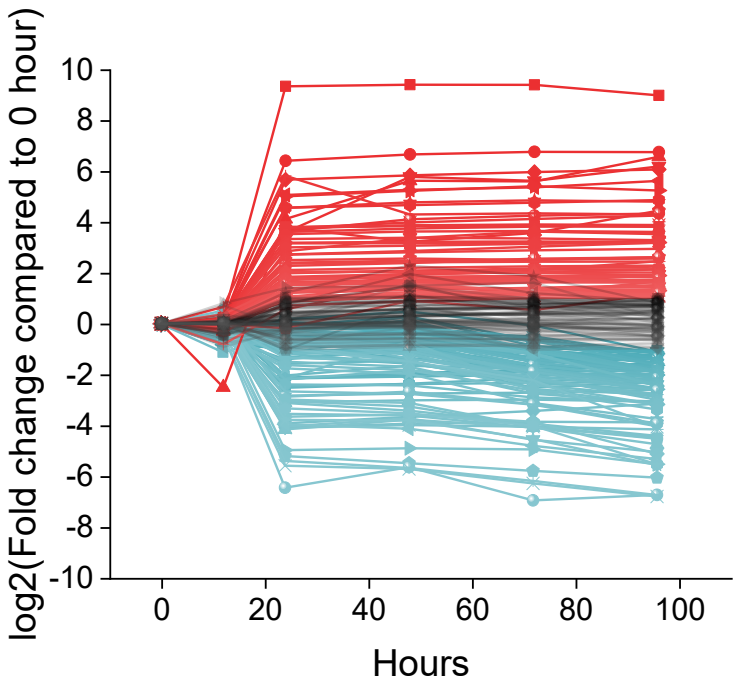

Figure 4.7. Surface glycoprotein abundance changes in A549 during EMT. The glycoprotein profiles shown has been normalized between the TGF-β-treated and PBS-treated group. The color scheme shows surface glycoproteins that were up- (red), down- (blue), or not regulated (black).

4.4 Conclusions

In this work, we studied the EMT process in A549 cells treated with TGF-β in a time-resolved manner, with a focus on the partial EMT state. Proteomics results confirmed that the transition did occur, as well as the proteins that are involved in the partial EMT state. Results are in good agreement with transcriptomics analysis and the potential drivers of the partial state. Surface glycoproteomics results showed several surface glycoproteins that are involved in the transition. With further analysis at the other protein layers, such as secretomics, glycoproteomics, and phosphoproteomics, the results will lead to a better understanding of the EMT process and their use as markers for cellular development and cancer metastasis.

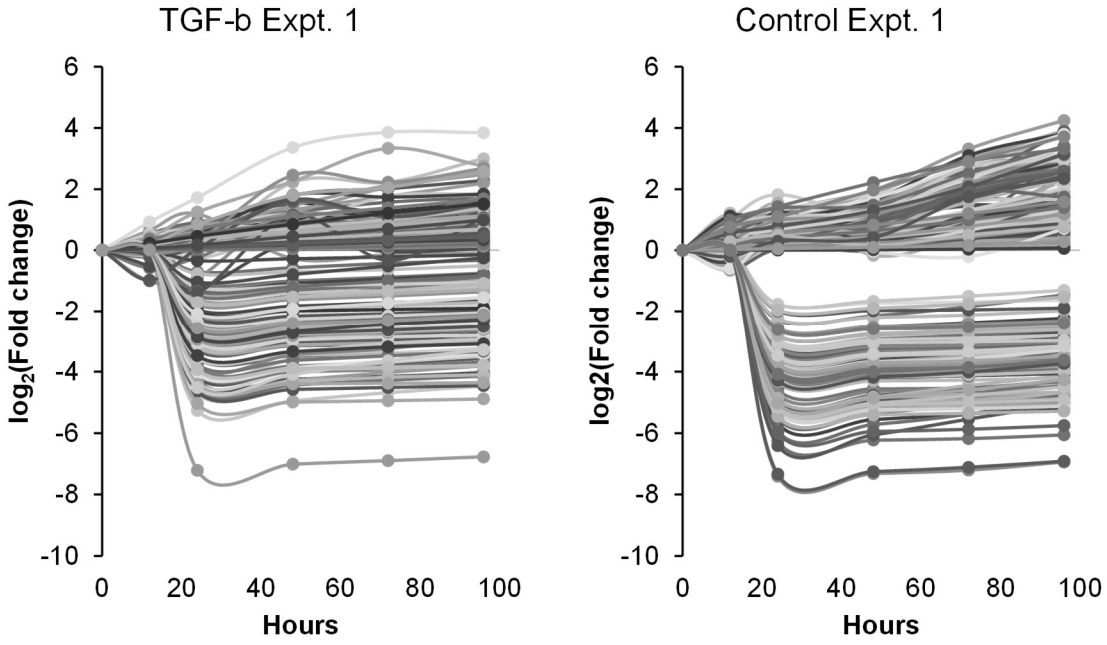

Figure 4.8. Example abundance changes of surface glycoproteins from the TGF-β-treated experiment or the control experiment without abundance normalization.

CHAPTER 5. ENHANCING COMPREHENSIVE ANALYSIS OF GLYCOPROTEINS FROM CULTURED CELLS WITHOUT SERUM STARVATION

Adapted with permission from: Suttapitugsakul, S.; Tong, M.; Sun, F.; Wu, R. Enhancing comprehensive analysis of secreted glycoproteins from cultured cells without serum starvation. Analytical Chemistry 2021, 93 (4), 2694-2705. Copyright (2021) American Chemical Society.

5.1 Introduction

Cells secrete many glycoproteins to regulate extracellular activities, including cell-cell communication and manipulation of the cell microenvironment.[56, 104, 313-314] Secreted glycoproteins, including cytokines, antibodies, growth factors, hormones, and enzymes, from easily accessible and non-invasive sources such as blood and urine contain much valuable information about cellular development and disease statuses, and thus they can serve as effective biomarkers for early disease detection.[59, 315-319] For example, secreted prostate-specific antigen (PSA) is a well-known glycoprotein used in prostate cancer diagnosis.[320] Recent studies even indicated that measuring different urinary PSA glycoforms may improve the diagnostic accuracy in the "grey zone" concentration range (4.0-10.0 ng/mL PSA).[321-322] Cells in the immune system such as monocytes and macrophages secrete glycoproteins during the infection to fight against foreign pathogens. For instance, once stimulated with lipopolysaccharides (LPS) after a bacterial infection, macrophages trigger nuclear factor kappa-B (NF-κB) to migrate into the nucleus, resulting in the secretion of cytokines such as tumor necrosis factor alpha (TNF-α), interleukin-1

(IL-1), and chemokines that attract other white blood cells such as the neutrophils to the vicinity.[313, 323-324]

Cultured cells have been extensively used as models in the fields of biology and biomedicine, and global analysis of glycoproteins secreted from these cells can aid in a better understanding of their functions and extracellular events. Despite their importance, secreted glycoproteins cannot be easily analyzed globally because many of them are present in an extremely low abundance. Furthermore, detection of secreted glycoproteins is often hindered by many other proteins from fetal bovine serum (FBS) including growth factors and proteins necessary for cell growth and proliferation used in the culture medium. Compared with serum proteins, the abundance of secreted glycoproteins is much lower.[325] While the analysis of secreted proteins can be performed using antibody-based methods such as Western blotting and ELISA, the antibody availability and specificity may be an issue. More importantly, it is almost impossible to use these methods for global analysis of secreted glycoproteins, and they cannot provide the glycosylation information.

Mass spectrometry (MS)-based proteomics has also been used for comprehensive analysis of secreted glycoproteins.[326-329] To overcome the challenges in secreted glycoprotein analysis, extensive fractionation may be performed. Yet, this method is not very effective and time-consuming. Furthermore, it does not result in the high coverage of secreted glycoproteins. Protein depletion is also commonly employed to remove some highest abundant proteins from the samples, but the reproducibility and the loss of proteins of interest are problematic.[330] Alternatively, serum-free media (SFM) have been employed to avoid the interference from high-abundance serum proteins. While this approach increases the coverage of secreted proteins detected, studies have shown that serum

starvation dramatically affects cell growth and proliferation, and thus results in protein secretion alterations.[82, 325] Taking these issues into consideration, global analysis of secreted glycoproteins and their glycosylation sites under the serum-containing conditions is much more challenging and remains to be further explored.

In this study, in order to increase the coverage of secreted glycoproteins, we integrated a selective enrichment method with a signal boosting approach for comprehensive analysis of secreted glycoproteins from cells cultured in serum-containing media. Since low-abundance secreted glycoproteins are buried among many extremely high-abundance FBS proteins in the medium, selective enrichment is critical for their global analysis. However, the enrichment alone is not sufficient because, besides non-specific binding, the detection limit of LC-MS is also unfavorable for low-abundance secreted glycoproteins. We first evaluated different glycoprotein sources to serve as the signal boosting sample. Furthermore, the boosting-to-sample ratio as another important parameter was carefully chosen to maximize the boosting effect and increase the coverage of glycoproteins. The method allowed us to quantify hundreds of secreted glycoproteins and deglycosylated peptides with the glycosylation sites from THP-1 monocytes and M0 macrophages in response to LPS and from Hep G2 treated with TGF-β. Several secreted glycoproteins were found to be regulated in these models, including those involved in the immune and inflammatory responses from THP-1 monocytes and macrophages treated with LPS, and those participated in the extracellular matrix organization and the epithelial-to-mesenchymal transition (EMT) from Hep G2 cells. Global quantification of secreted glycoproteins from cultured human cells without serum starvation provides valuable

information to the scientific community, and the developed method can be extensively applied for secreted glycoprotein analysis.

5.2 Materials and methods

5.2.1 Cell culture, treatments, and metabolic labeling of glycoproteins

Reagents were purchased from Sigma-Aldrich unless noted otherwise. For the experiment with THP-1 (ATCC), the cells were maintained in RPMI-1640 medium (Gibco) containing 10% FBS (Corning) and 1% penicillin-streptomycin (P-S) in a humidified incubator with 5% carbon dioxide at 37 °C. Once the density reached ~$8x10^5$ cells/mL, the cells were centrifuged at 500*g* for 5 minutes, washed twice with warm PBS, and resuspended in the corresponding medium before metabolic labeling. For the serum-containing medium (SCM), the original medium with 10% FBS and 1% P-S was used. For the serum-free medium (SFM), 10% FBS was not added. Depending on the experiment, the cells were pre-labeled with 100 μM N-azidoacetylgalactosamine-tetraacylated (Ac_4GalNAz; Click Chemistry Tools) in SCM or SFM in the incubator. After 12 hours, either 1 μg/mL LPS from *E. coli* O111:B4 or PBS at the same volume was added to the culture flasks and then incubated for 12 hours. For the experiments with M0 macrophages, THP-1 monocytes were differentiated as reported previously using 100 ng/mL phorbol 12-myristate 13-acetate (PMA) in the SCM for 48 hours, rested for 24 hours in the SCM without PMA, and washed with warm PBS twice before adding the corresponding medium containing Ac_4GalNAz.[208, 214] For the experiments with Hep G2 (ATCC), the cells were grown similarly in high-glucose DMEM containing 10% FBS and 1% P-S. Once the confluency reached ~50%, the cells were switched to the serum-free or serum-containing

conditions, pre-labeled with 100 μM Ac_4GalNAz for 12 hours, and treated with 10 ng/mL human transforming growth factor β1 (TGF-β, Cell Signaling Technology) for 48 hours before medium collection.

5.2.2 Medium collection, cell lysis, and click-chemistry reaction

The cells and the media were collected by centrifugation at 500*g* for 5 minutes. The cells were washed with ice-cold PBS twice and the wash solution was pooled with the media. For the experiments with the cell part, proteins were extracted using a lysis buffer containing 50 mM N-(2-hydroxyethyl)piperazine-N'-2-ethanesulfonic acid (HEPES; pH=8.8), 150 mM sodium chloride (NaCl), 10% sodium deoxycholate (SDC), 1 tablet/50 mL protease inhibitors (Roche), and 20 units/mL universal nuclease for cell lysis (Pierce) at 4 ºC with end-over-end rotation for 1 hour. The lysates were collected by centrifugation at 25,830*g* for 10 minutes. For the medium part, the same protease inhibitors were added to the final concentration of 1 tablet/50 mL before centrifugation at 5,000*g* for 10 minutes to remove any debris. The media were concentrated using an Amicon Ultra-15 centrifugal filter unit with a 3,000 Da molecular weight cutoff (Millipore) at 4,000*g* until the volume was reduced to ~5 mL. For comparison purposes, the amount of HEPES, NaCl, SDC, protease inhibitors, and universal nuclease were adjusted to the same final concentrations as those in the cell part. The final volume of the lysates or the concentrated media was adjusted with PBS to 10 mL before the click reaction. Similar steps were performed in the experiments with M0 macrophages and Hep G2 cells, except that the media were collected and the cells were washed directly from the culture flasks. Glycoproteins containing the azide group from Ac_4GalNAz labeling were tagged with 100 μM dibenzocyclooctyne (DBCO)-biotin (Click Chemistry Tools) at 37 ºC with shaking for 1 hour.

5.2.3 Protein purification and digestion, and peptide purification

After the click-chemistry reaction, the proteins were reduced with 5 mM dithiothreitol at 56 ºC for 30 minutes, alkylated with 14 mM iodoacetamide at room temperature in the dark for 30 minutes, and purified with the methanol-chloroform precipitation method. Proteins were digested with sequencing-grade modified trypsin (Promega) in the buffer containing 1.6 M urea, 50 mM HEPES pH=8.8, and 5% acetonitrile (ACN), which lasted for 16 hours with shaking at 37 ºC. The digestion was quenched with 0.4% trifluoroacetic acid (Millipore) and the pH was checked to be lower than 2. The peptides were desalted using a tC18 Sep-Pak Vac Cartridge (Waters) and dried using a vacuum concentrator.

5.2.4 Enrichment of secreted glycopeptides

The dried peptides were resuspended in PBS and glycopeptides were enriched using 150 µL NeutrAvidin Agarose Resins (Pierce) for 1 hour at 37 ºC with end-over-end rotation. The beads were subsequently washed 10 times with PBS. The enriched glycopeptides were eluted three times with 8 M guanidine hydrochloride pH=1.5 (Promega) at 56 ºC for 2 minutes each. The eluted glycopeptides were desalted using the Sep-Pak cartridge.

5.2.5 TMT labeling and high-pH reverse-phase fractionation

For the quantification experiment using the boosting method, the dried peptides were resuspended in 33 µL 200 mM HEPES pH=8.5 and 10 µL ACN. The TMT 6-plex or TMT 10-plex reagents (Thermo) were dissolved in 41 µL anhydrous ACN, and 5 µL was

added to the solution containing the peptides. The reaction was performed at room temperature with shaking for 1 hour and subsequently quenched with 8 µL of 5% hydroxylamine hydrochloride in 200 mM HEPES pH=8.5 at room temperature with shaking for 15 minutes. The peptides from all TMT channels were combined, purified, and dried using a vacuum concentrator. The mixed peptide samples were fractionated using high-pH reverse-phase HPLC with an XBridge C18 3.5 µm, 4.6x250 mm column (Waters) over a 40-minute gradient of 16-60% ACN in 10 mM ammonium formate, pH=10. The peptides were collected every 2 minutes and consolidated into 5 fractions.

5.2.6 PNGase F treatment

The dried peptides were resuspended in 40 µL 50 mM ammonium bicarbonate (pH=9) in heavy-oxygen water ($H_2{}^{18}O$) and deglycosylated using 3 units of PNGase F at 37 ºC for 3 hours with shaking. The reaction was quenched with 1% formic acid (FA) and the pH was adjusted to be ~2. The peptides were purified using the StageTip method described previously and dried before LC-MS/MS analysis.[161] For the optimization experiments, each sample was eluted into three fractions with 20%, 50%, and 80% ACN containing 1% acetic acid, respectively. For the quantification experiments with the boosting channel, each of the five fractions was eluted with 50% ACN containing 1% acetic acid. The purified peptides were dried using a vacuum concentrator.

5.2.7 LC-MS/MS analysis

The dried peptides were resuspended in a solution containing 5% ACN and 4% FA and analyzed using an on-line LC-MS/MS system. The separation was performed using a Dionex UltiMate 3000 UHPLC system (Thermo). The microcapillary column was packed

in-house (Magic C18AQ, 3 μm, 200 Å, 75 μm x 16 cm). For the optimization experiment, the peptides from the 20%, 50% and 80% fractions were separated using 112-minute gradients of 3-22%, 6-30%, and 8-35% ACN containing 0.125% FA, respectively. For the quantification experiments with the boosting channel, each of the five fractions was separated with a 112-minute gradient of 4-17% ACN with 0.125% FA. The LC is connected to an LTQ Orbitrap Elite hybrid mass spectrometer (Thermo). For each cycle, a full MS spectrum was recorded in the Orbitrap cell at the resolution of 60,000 with automatic gain control (AGC) of $1x10^6$. For the optimization experiment, the peptides were fragmented in the LTQ using a data-dependent Top20 method where a full MS scan in the Orbitrap is followed by up to 20 MS/MS in the LTQ for the most intense ions. Selected ions were excluded from further sequencing for 90 seconds. Ions with singly or unassigned charge were not sequenced. Maximum ion accumulation times were 1,000 ms for each full MS scan and 50 ms for each MS/MS scan. For the quantification experiments, peptide fragmentation was performed by higher-energy collisional dissociation (HCD) with the normalized collisional energy of 40% using a Top15 method, i.e., up to 15 MS/MS of the most abundant precursor ions were recorded in the Orbitrap cell at the resolution of 15,000.

5.2.8 Data analysis and bioinformatics analysis

The raw files were converted into an mzXML format and searched using the SEQUEST algorithm.[217] The spectra were matched against the human proteome database downloaded from UniProt (www.uniprot.org). In the optimization experiment, the peptide mass tolerance was 20 ppm and the fragment ion mass tolerance was 1.0 Da. The maximum number of missed cleavages was three and the maximum number of differential modifications per peptide was four. Differential modifications included: +15.9949 Da for

oxidation of methionine and +2.9883 Da for glycosylation on asparagine, which was deglycosylated with PNGase F in $H_2{}^{18}O$. Static modification included: +57.0215 Da for carbamidomethylation of cysteine. Similar parameters were used in the quantification experiment except that the fragment ion mass tolerance was 0.025 Da and the static modification at lysine and the peptide N-terminus (+229.1629 Da) for the TMT labeling was added. The target-decoy method was used to estimate the false discovery rates (FDRs) of peptide and protein identification.[218] Linear discriminant analysis (LDA) was applied to control the quality of peptide identification using multiple parameters such as XCorr, ppm, peptide length, and charge state. The FDR was controlled to <1% at the deglycosylated peptide level, and an additional filter was also applied to control the FDR to <1% at the glycoprotein level. The dataset was limited to only deglycosylated peptides or glycoproteins when calculating and controlling the FDRs at both levels. Note that only for the secretome quantification with THP-1 monocytes without the boosting channel, one of the five fractions did not pass the LDA and the results from four fractions were included.

The confidence of the glycosylation site localization was calculated using an algorithm similar to Ascore called ModScore.[219] A ModScore of >13 represents the site being well-localized ($P<0.05$). Reverse hits and contaminants were removed. The deglycosylated peptides were filtered so that each sequence contained the N-!P-S/T/C motif for N-linked glycosylation. For the quantification, the signal-to-noise ratio of the TMT reporter ions was used. Peptides with a zero S/N were removed. The abundance of each unique deglycosylated peptide was calculated from the sum S/N of all peptide copies detected and the one with the highest ModScore is reported. The final protein abundance change is the sum S/N ratios of all peptides for the protein. Statistical analysis was

performed using Perseus,[165] including the one-sample t-test where the significance change is defined for those with a minimum fold change of 1.5 and the P-value less than 0.05. Phobius was employed to predict the transmembrane region and the signal peptide in proteins.[166] SecretomeP was used to predict non-classical secretion.[167] Protein clustering was performed using the Database for Annotation Visualization and Integrated Discovery (DAVID) version 6.8.[220] The raw files can be accessed on ftp://massive.ucsd.edu/MSV000086461/.

5.3 Results and discussion

5.3.1 Principle of enhancing the detection of secreted glycoproteins with low abundances

Cells secrete glycoproteins into the extracellular environment to communicate with other cells and manipulate the cellular microenvironment. Many secreted glycoproteins have a very low abundance (below ng/mL) among the high-abundance background proteins in the mg/mL range. Methods to enrich secreted proteins from cell culture media were previously reported.[82, 108, 325, 331] However, they generally do not target secreted glycoproteins and/or the analysis does not reveal the glycosylation site information. While glycoprotein enrichment is imperative for their global analysis, it is not always sufficient for comprehensive analysis of secreted glycoproteins due to the high-abundance protein background in the culture media. Furthermore, even after they are enriched, the amount of many low-abundance glycoproteins may still be below the detection limit of LC-MS.

Multiplexed proteomics using the tandem mass tag (TMT) allows for the quantification of proteins from multiple samples simultaneously. This increases the reproducibility and shortens the analysis time. Through this approach, the same peptide

from different samples is labeled with each channel of the TMT reagents and then combined. The resulting peptides have the same mass-to-charge ratio and appear as a single peak in MS1. It not only increases the analysis efficiency but also improves the peptide signal in MS1, allowing for the better detection and isolation of the precursor ion for MS2. More importantly, much higher intensities of fragments in MS2 enable us to more confidently identify the peptide, while the reporter ions from the TMT tags allow for quantification of the peptides from different samples. Benefitting from this signal enhancement, this approach has been applied for post-translational modification and single-cell analyses by dedicating a separate TMT channel for peptide carriers that will result in the higher total intensity in MS1. Previously, Budnik et al. developed a method called SCoPE-MS to study single-cell proteomics,[111] and Yi et al. reported the BASIL method for phosphopeptide identification and quantification.[112]

To overcome the problems in secreted glycoprotein analysis, we envisioned that the boosting approach in combination with a glycoprotein enrichment method could be very useful to uncover low-abundance secreted glycoproteins (Figure 5.1). In previous studies, the boosting sample (channel) generally has the same composition as the quantification samples (channels), e.g., for the BASIL method, the boosting sample is generated by combining the quantification samples; in SCoPE-MS, the number of cells used in the boosting channel is much higher compared with the real single-cell samples. In this work, secreted glycoproteins are among the highly abundant FBS proteins. Using the combined media containing the highly abundant FBS proteins from the samples as the boosting channel may not work well because it further increases the abundance of FBS

proteins and will not facilitate the global analysis of secreted proteins. Therefore, a different and more effective boosting sample must be used.

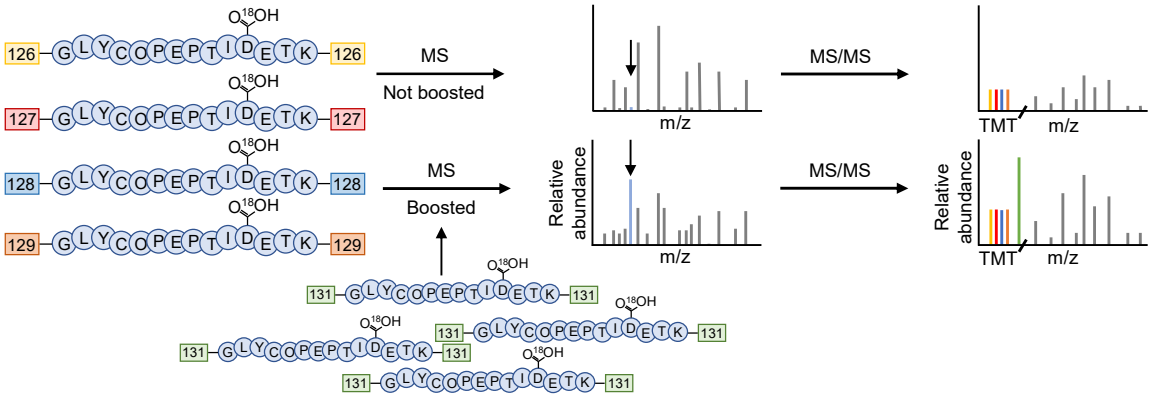

Figure 5.1. Principle of the glycopeptide signal boosting for secreted glycoprotein analysis. Enriched glycopeptides from secreted glycoproteins in the cell culture media (shown in the deglycosylated form) are labeled with the TMT reagents. When the boosting sample is added, the signal of the deglycosylated peptide in MS1 is increased, thus facilitating the precursor ion selection for MS/MS fragmentation. Higher intensities of the fragments also allow for confident identification of the glycopeptide.

Previously, we specifically analyzed cell-surface glycoproteins by integrating metabolic labeling with a sugar analog, bioorthogonal chemistry, and MS-based proteomics, and studied cell surface glycoproteins in different cell types and their dynamics.[48, 181, 215-216, 332] Briefly, cells are labeled using Ac_4GalNAz, which is incorporated into the glycan part of glycoproteins in the endoplasmic reticulum or the Golgi apparatus through the classical secretory pathway. In cells, Ac_4GalNAz is deacetylated and activated into UDP-GalNAz, which is converted to UDP-GlcNAz or remained as UDP-GalNAz. Many N-glycans can be labeled with the azido sugar. It has also been shown that UDP-GlcNAc can be converted into ManNAz for the labeling of sialic acid-containing glycans.[41, 333] Because secreted glycoproteins are normally exported through the classical secretory pathway, they can be labeled with the sugar analog and enriched from the culture

medium. It is expected that coupling the boosting approach with selective enrichment of secreted glycoproteins will be highly effective to cover low-abundance glycoproteins secreted in the medium.

5.3.2 Comparison of different secreted glycoprotein sources to most effectively boost the signals of secreted glycopeptides from SCM

While the cells from which the medium was collected present an attractive choice as a boosting sample, the abundances of secreted glycoproteins inside of the cells could be low and the inferences from many other highly abundant intracellular proteins, especially housekeeping ones, may pose a problem. To make sure that the boosting channel is appropriate and the highest coverage of secreted glycoproteins is obtained, we first compared theoretical annotated secreted glycoproteins that can be found from different sources, including the cells and the media under serum-free or serum-containing conditions.

THP-1 monocytes were first cultured in 25 mL SCM until the density reached ~$8x10^5$ cells/mL. The cells were harvested, washed twice with warm PBS to remove FBS proteins, and passaged to either SFM or SCM with 100 μM Ac_4GalNAz. After being labeled for 24 hours, the cells from both conditions were harvested and separated from the medium, resulting in four sources of glycoproteins: 1) the cell part, SFM; 2) the medium part, SFM; 3) the cell part, SCM; 4) the medium part, SCM. The cell parts were lysed and the lysates were collected. The medium parts were concentrated using a centrifugal filter. Proteins containing the azide groups from both the cell lysates and the media were then tagged with DBCO-sulfo-biotin, followed by digestion with trypsin. The resulting

glycopeptides containing the biotin group were enriched using NeutrAvidin beads, deglycosylated with PNGase F in $H_2{}^{18}O$, and analyzed with LC-MS/MS.

In the comparison experiments, different numbers of glycoproteins and deglycosylated peptides were detected from the cells and the media under serum-free or serum-containing conditions (Figure 5.2). As expected, the results from the cell parts are very similar, i.e., 989 deglycosylated peptides from 455 glycoproteins and 952 deglycosylated peptides from 443 glycoproteins in the serum-free or serum-containing media, respectively. The number of glycoproteins from the SFM is slightly lower; 698 deglycosylated peptides from 341 glycoproteins. Yet, the number of glycoproteins detected from the SCM is the lowest, with only 54 deglycosylated peptides from 41 glycoproteins found. In a separate search from this experiment where the FDR is not restricted to only glycopeptides, peptides with the highest number of hits are from bovine albumin (62 unique peptides and 330 total peptides) and fetuin-A (19 unique peptides and 219 total peptides), while the human protein detected with the highest number of peptides is progranulin (GRN) with only 4 unique peptides and 17 total peptides (all contain a glycosylation site). This demonstrated that the enrichment alone is not enough to overcome high-abundance FBS proteins left in the final sample through non-specific binding.

At first glance, the number of glycoproteins and deglycosylated peptides detected from the cell part grown in SFM is the highest and could serve as the boosting channel during the quantification. We performed several bioinformatic analyses to determine the fraction of theoretical secreted glycoproteins among these different samples. Based on the criteria used in previous studies, secreted proteins should contain a signal peptide sequence but not a transmembrane region.[82, 325] We used Phobius to predict these two components

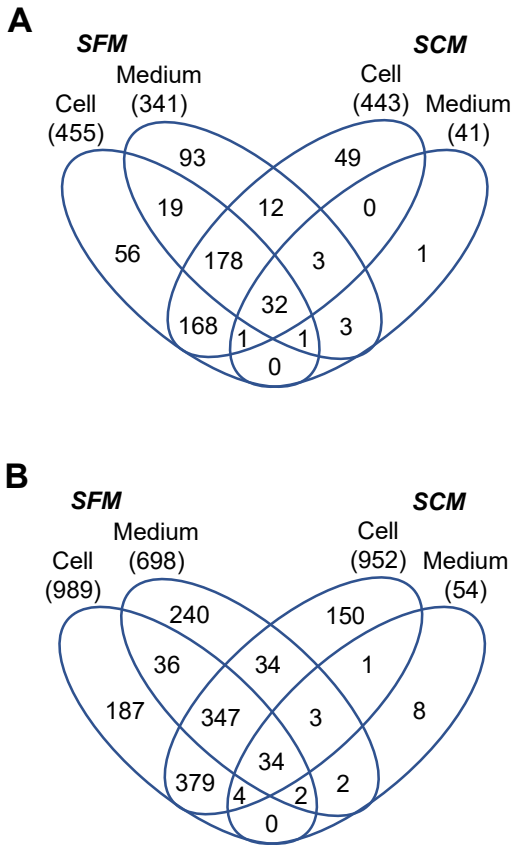

Figure 5.2. Comparison of glycoproteins and deglycosylated peptides from different boosting samples. The number of glycoproteins (A) or deglycosylated peptides (B) detected from the cells or the media from serum-free or serum-containing conditions. The number in parentheses is the number of glycoproteins or unique deglycosylated peptides identified.

and also compared with the UniProt database for proteins that contain the keyword "Signal" or "Secreted".[166] Due to its coverage, the medium part of the SCM condition has the lowest number of glycoproteins with the transmembrane region, with the signal peptide, without the transmembrane region but with the signal peptide, or containing the keywords "Signal" or "Secreted" (Figure 5.3A). Those from the cell parts in both the SFM and SCM conditions have the highest numbers of glycoproteins that matched these criteria, except for proteins that contain a signal peptide but not a transmembrane region, which is characteristic for secreted proteins and the highest in the medium part of the SFM (Figure

5.3A). On the contrary, the percentages of proteins that match these criteria revealed that, as expected, all glycoproteins (100%) enriched from the medium part of the SCM are annotated with the keywords “Signal” or “Secreted” (Figure 5.3B). Additionally, 68% have a signal peptide sequence but not the transmembrane region, 95% have a signal peptide (highest among the four conditions), and 32% have the transmembrane region (the lowest among the four). While the percentages are not as high as those from the medium part of SCM, the medium part of SFM still has 46% with a signal peptide sequence but not a transmembrane region. Those enriched from the cell parts have the highest number of proteins with the transmembrane region, the lowest for proteins containing a signal peptide sequence, or containing a signal peptide sequence but not the transmembrane region. Non-classical secretory pathways where glycoproteins are secreted through are also predicted (Table S1). Additionally, we clustered the secreted glycoproteins based on their gene-ontology (GO) terms. Glycoproteins detected from the SFM resulted in the highest number of proteins annotated in the extracellular vesicle (secreted proteins from databases are also clustered into this GO) and the lowest P-values (highest -log(P-value); as shown in Figure 5.3C). The number of proteins that belong to the integral component of membrane is also the lowest when the glycoproteins from the SCM are not considered.

Taking these together, SFM resulted in the highest number of theoretically secreted glycoproteins, and the coverage is much higher compared with that from the SCM. The SFM contains almost all glycoproteins detected from the SCM except two glycoproteins, i.e., C-C motif chemokine 24 (CCL24) and cathepsin Z (CTSZ). It also contains the lowest number of proteins with the transmembrane region, which is not characteristic of typical secreted glycoproteins. Additionally, considering that glycosylation is an important step in

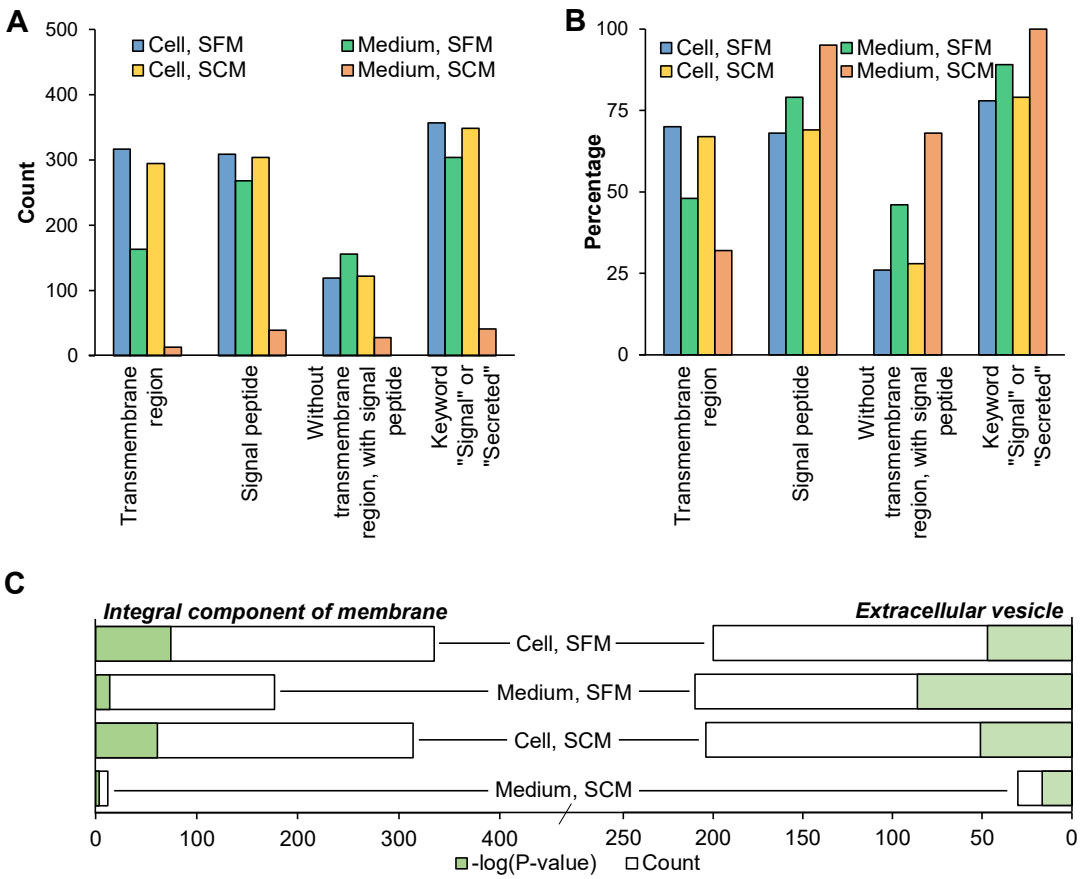

Figure 5.3. Analysis of secreted glycoproteins from different boosting samples. (A) The number of glycoproteins from Figure 2 that contain the transmembrane region, have a signal peptide sequence, does not contain the transmembrane region, but contain a signal peptide sequence, and are annotated as the keywords "signal" or "secreted" from UniProt. (B) The percentage of glycoproteins from (A) to the total number of detected glycoproteins for each sample. (C) Gene ontology (GO)-based clustering of glycoproteins related to integral component of membrane or extracellular vesicle in different samples based on cellular components.

the classical secretory pathway, almost all secreted proteins may be glycosylated and could possibly be used as the boosting sample without the enrichment. We grew the cells in the SFM, omitted the enrichment with NeutrAvidin beads, and directly performed the deglycosylation with PNGase F for glycopeptides prior to LC-MS analysis. With this approach, only 83 deglycosylated peptides from 66 glycoproteins were detected. The results are worse than expected. Therefore, we concluded that the enriched secreted

glycoproteins in the SFM best suit for boosting the signals of those from cells grown in the SCM. While previous studies showed that cells grown in the SFM may alter the abundance of secreted proteins, the glycoproteins here are merely used to boost the signal while the true quantification results will still be obtained from the cells grown under the SCM conditions.

5.3.3 Optimization of the ratio between the boosting and the quantification channels

To evaluate the boosting effect, we performed the quantification experiments without the boosting channel and with different boosting-to-sample ratios. The cells were pre-labeled with Ac_4GalNAz and treated with LPS or PBS (as a control) for 12 hours. The media were collected, and secreted glycoproteins were tagged. After the enrichment, the glycopeptides were labeled with the TMT 6-plex reagents (Figure 5.4). In the experiment without the boosting sample, channels 126-129 of the TMT reagents were used to label glycopeptides from the biological duplicate experiments of cells treated with LPS or PBS while channels 130 and 131 were left blank. With this approach, 103 unique deglycosylated peptides were detected from 71 glycoproteins (Figure 5.5A). With the boosting approach, channel 131 was dedicated for the boosting sample. Since the cells treated with LPS or PBS are expected to secrete glycoproteins differently, we combined the SFM collected from both the cells treated with LPS or PBS as a boosting sample. Channel 130 was left blank. With the boosting sample, the number of glycoproteins and deglycosylated peptides identified markedly increased. For the boosting-to-sample ratio of 2:1, i.e., the volume of the SCM for each channel is 5 mL and that of the boosting channel is 10 mL (5 mL from the SFM of cells treated with LPS and 5 mL from cells treated with PBS), 179 unique deglycosylated peptides were quantified from 107 secreted glycoproteins (Figure 5.5A).

With the ratio of 10:1 where 5 mL of the SCM for each channel and 50 mL of the SFM for the boosting channel were used (25 mL from the SFM of cells treated with LPS and 25 mL of cells treated with PBS), the number further increased to 308 deglycosylated peptides from 178 glycoproteins (Figure 5.5A). Compared with the results from the identification in the comparison experiment, the secreted glycoproteins quantified at the 10-to-1 ratio cover 89% of the secreted glycoproteins (36 of 41) detected from the SCM. Nonetheless, the coverage is lower than the identification experiment in the SFM (599 deglycosylated peptides and 341 glycoproteins). One major reason is that the ion trap used for MS2 in the identification experiment is much more sensitive than the Orbitrap cell employed for MS2 in this quantification experiment because of the detection of the TMT reporter ions.

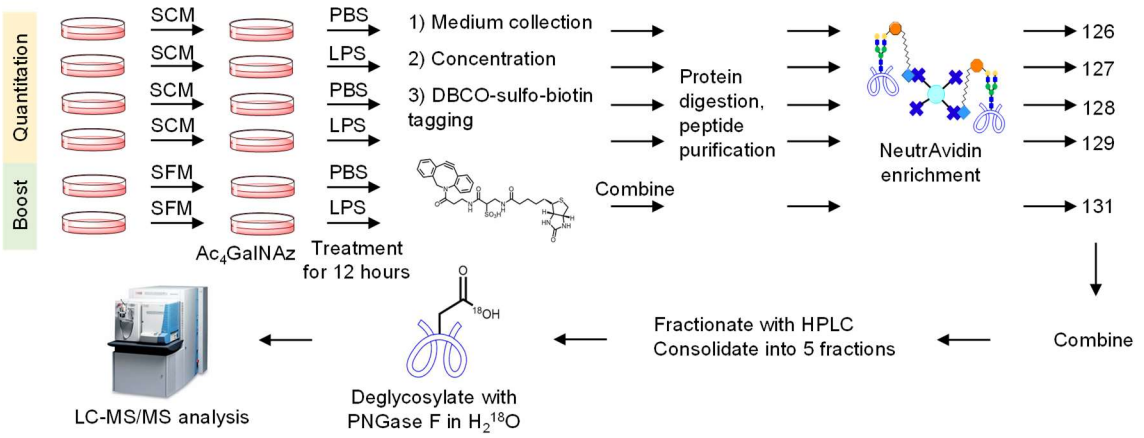

Figure 5.4. Experimental procedure for quantitative analysis of secreted glycoproteins using a boosting channel. The cells are pre-labeled with Ac4GalNAz for glycoprotein enrichment in either SCM (quantification channels) or SFM (boosting channel). The media are collected, and the glycoproteins are tagged with DBCO-sulfo-biotin. The enriched glycopeptides are labeled with the TMT 6-plex reagents, deglycosylated with PNGase F, and analyzed with LC-MS/MS.

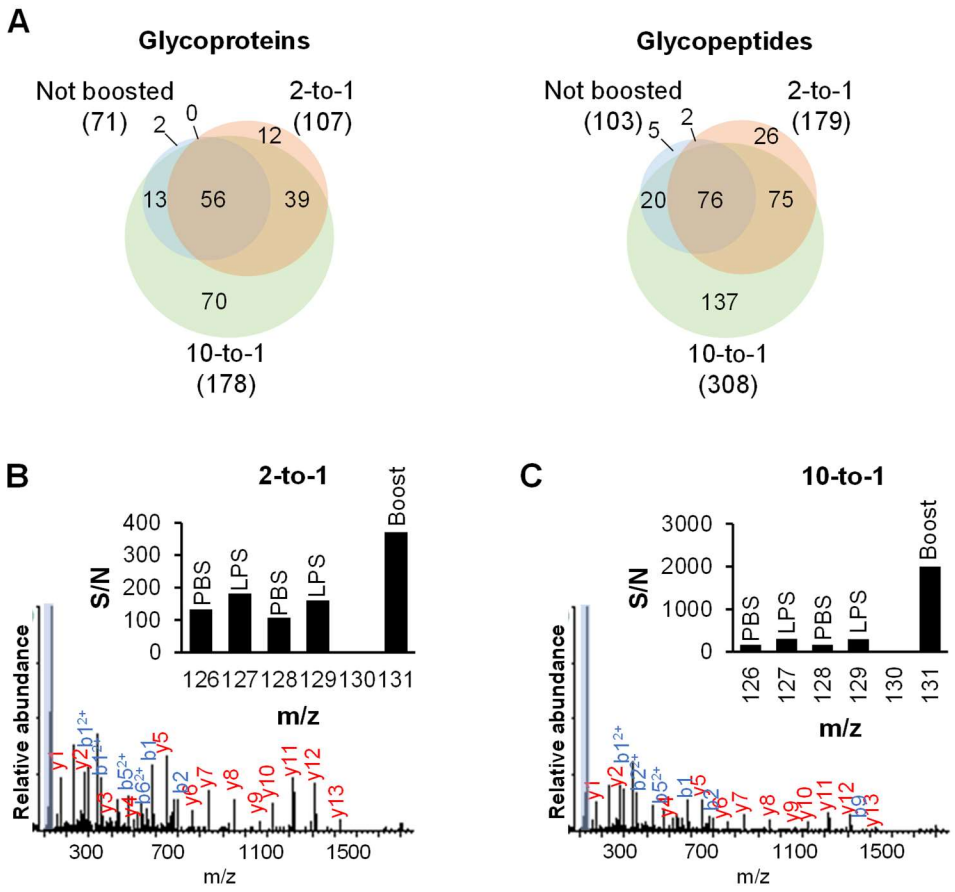

Figure 5.5. Optimizing the boosting-to-sample ratio. (A) The number of secreted glycoproteins or deglycosylated peptides from the experiments without the boosting channel, boosted at the 2-to-1 ratio, or boosted at the 10-to-1 ratio. (B) and (C) Example tandem mass spectra of the deglycosylated peptide KLPPGLLAN#FTLLR from LRG1 from the experiments boosted at the 2-to-1 ratio (B) or boosted at the 10-to-1 ratio (C). The TMT reporter ion region is highlighted and shown in the inset.

In one example, the deglycosylated peptide KLPPGLLAN#FTLLR, where # represents the glycosylation site, was quantified in the boosting experiments with the 2:1 and 10:1 boosting ratios (Figure 5.5B and C). The peptide was confidently identified with the XCorr values of 4.86, and 4.93, respectively. The glycosylation site is well-localized at N186 with the ModScore of 1,000 in all experiments. The peptide is from leucine-rich alpha-2-glycoprotein (LRG1). As the boosting-to-sample ratio increases, the signal at the channel of 131 also increases. Note that the abundance of this glycoprotein is high enough

so that it can be detected even without the boosting channel. It might also be due to the LPS treatment that increased the glycoprotein abundance in the secretome. For other glycopeptides that were not detected from the quantification experiment without the boosting channel, their abundances may be too low to be selected for further fragmentation, or not enough fragments are produced for glycopeptide identification in MS2. The addition of the boosting channel increases the signals in MS1 and MS2 and improves the glycopeptide coverage.

5.3.4 Quantification of secreted glycoproteins from THP-1 monocytes in response to LPS

In the experiment with THP-1 monocytes, we quantified 309 deglycosylated peptides from 179 glycoproteins (Figure 5.6C). The reproducibility from two biological duplicate experiments is reasonably high at both the deglycosylated peptide and glycoprotein levels (Figure 5.6A and B). GO analysis demonstrates that over one hundred glycoproteins are related to the extracellular vesicle (P=8.5E-50), 170 are annotated with the keyword “Secreted” or “Signal”, while 82 are predicted to not have the transmembrane region but contain a signal peptide sequence. At the peptide level, the ratios of 66 unique deglycosylated peptides from secreted glycoproteins are ≥1.5 from the cells treated with LPS compared with the control sample (Figure 5.7A). Clustering of 27 glycoproteins that were up-regulated demonstrated that they are involved in biological processes such as positive regulation of programmed cell death, extracellular matrix organization, inflammatory response, and response to cytokine, which are the expected responses (Figure 5.6D). Several deglycosylated peptides with the glycosylated sites of N118 and N258 from tumor necrosis factor-inducible gene 6 protein (TNFAIP6) were up-regulated by ~16 times. TNFAIP6 has been associated with several inflammatory diseases and was expressed in

response to TNF-α, one of the known cytokines secreted from cells treated with LPS.[334] Another study showed that TNFAIP6 inhibited pro-inflammatory proteins while increasing the anti-inflammatory ones. Among these, TNFAIP6 inhibited toll-like receptor 4 (TLR4), the direct receptor of LPS, from associating with myeloid differentiation primary response protein MyD88 (MYD88) to suppress NF-κB activation.[335] This suggests that after the inflammation induced by LPS, cells may secrete more TNFAIP6 to counter this inflammation. Surprisingly, only one glycoprotein containing the N170 site from angiotensinogen (AGT) was down-regulated. AGT normally has roles in blood pressure regulation and body fluid and electrolyte homeostasis, but a recent study showed its involvement in the inflammatory response.[336] If the abundance of a secreted glycoprotein is already very low and the even lower amount of the glycoprotein is secreted from cells with the LPS treatment, the boosting samples may not be able to effectively improve its detection in the secretome. This potentially results in the under-estimated number of down-regulated glycoproteins and glycopeptides with very low abundances from the LPS-treated cells. Some exemplary regulated glycosylation sites are shown in Table 3.1.

Five cytokines were also quantified despite their abundances at the ng/mL level, including interleukin-12 subunit beta (IL12B), macrophage colony-stimulating factor 1 (CSF1), growth/differentiation factor 11 (GDF11), bone morphogenetic protein 1 (BMP1), and progranulin (GRN). Among these, the deglycosylated peptide with the site of N303 from IL12B was up-regulated by 7.5 times, and this has been reported previously.[337-338] We also found that the two deglycosylated peptides with the site of N183 or N267 from intercellular adhesion molecule 1 (ICAM1) were up-regulated by 2.1 and 2.5 times, respectively. ICAM1 is a known cell-surface glycoprotein involved in leukocyte adhesion

and is often up-regulated under inflammatory conditions.[339] ICAM1 can be shed and was found in a study that the LPS challenge increased its concentration in the culture medium.[340] This may explain the detection of some glycoproteins with the transmembrane region and a signal peptide in the current experiment, despite that they are generally considered as being localized at the cell surface but not in the secretome.

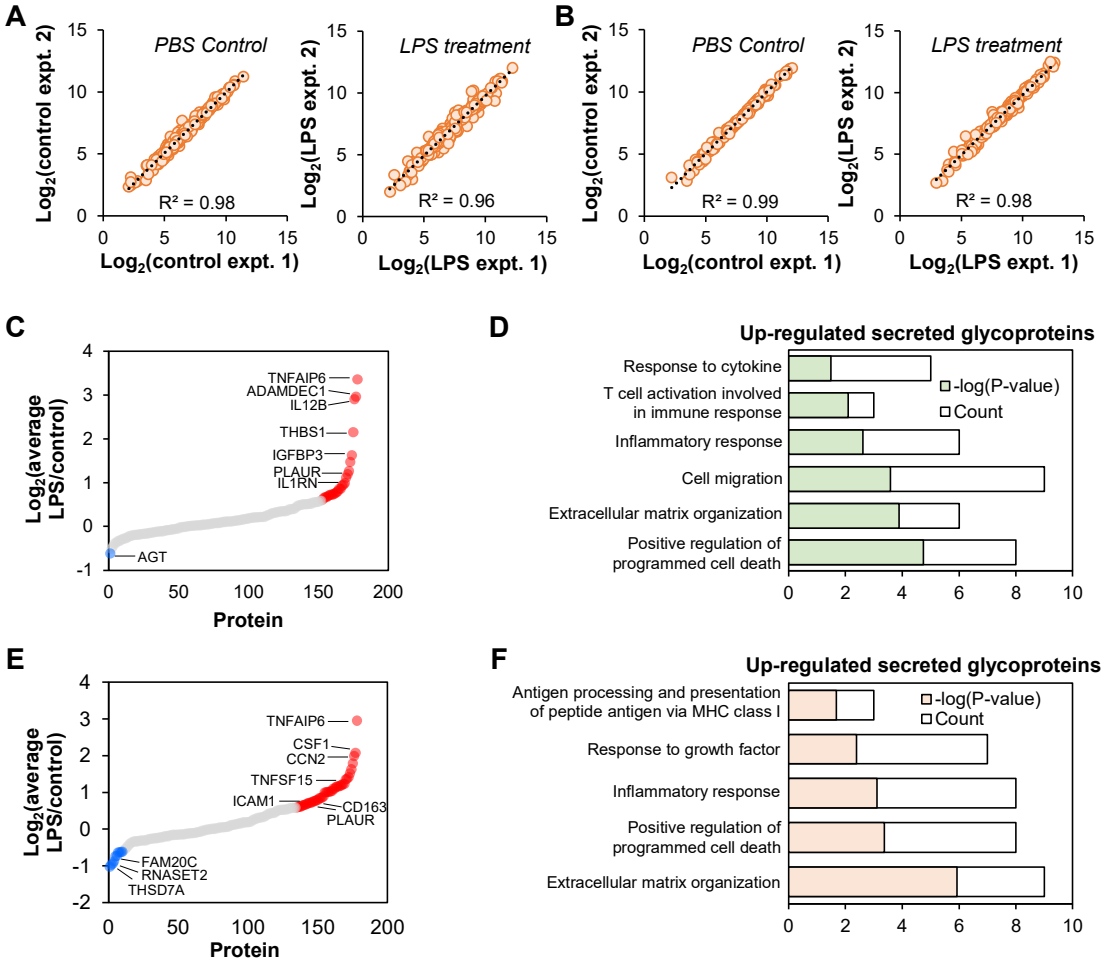

Figure 5.6. Quantification of secreted glycoproteins from THP-1 monocytes and M0 macrophages treated with LPS. (A-B) Correlation of deglycosylated peptides (A) and glycoproteins (B) quantified from THP-1 monocytes. (C) The abundance changes of secreted glycoproteins from THP-1 monocytes stimulated with LPS. The cutoff for up-or down-regulation is the fold change of 1.5. (D) GO clustering of up-regulated secreted glycoproteins from THP-1 monocytes after the LPS treatment based on biological process. (E) The abundance changes of secreted glycoproteins from M0 macrophages stimulated with LPS. (F) GO clustering of up-regulated secreted glycoproteins from M0 macrophages after the LPS treatment based on biological process.

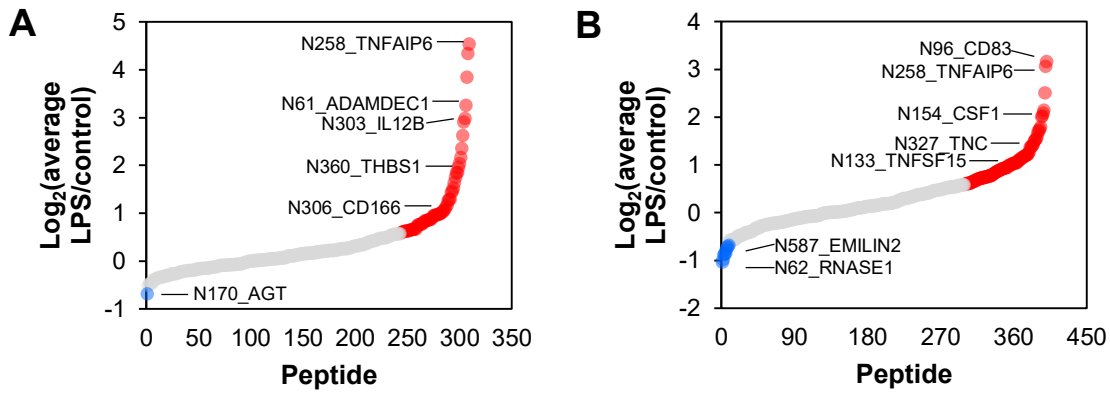

Figure 5.7. Quantification deglycosylated peptides from THP-1 monocytes and M0 macrophages treated with LPS. (A and B) Abundance changes at the deglycosylated peptide level from THP-1 monocytes (A) or M0 macrophages (B).

5.3.5 Analysis of glycoproteins secreted from THP-1 macrophages treated with LPS

THP-1 cells were further differentiated into M0 macrophages using PMA similar to the previously reported protocols.[208, 214, 341] The cells were subsequently labeled with $Ac_4GalNAz$ and treated with LPS as the above experiment with THP-1 monocytes. In total, 400 deglycosylated peptides were quantified from 178 glycoproteins. The results from both replicates indicate the relatively high R^2 values (Figure 5.8B). More than one hundred glycoproteins are related to the extracellular vesicle (P=2.9E-37), and 170 are annotated with the keyword "Secreted" or "Signal". Moreover, 81 are predicted to not have the transmembrane region but contain a signal peptide. The results at the protein level are shown in Figure 5.6E. Forty-four glycoproteins were up-regulated while ten were down-regulated. Clustering of up-regulated secreted glycoproteins revealed that they are involved in extracellular matrix organization (P=1.2E-6), inflammation response (P=7.8E-4), and response to growth factor (P=4.1E-3) (Figure 5.6F). Among the quantified deglycosylated peptides, 101 were up-regulated in the cells treated with LPS while 9 were down-regulated

compared with the control group (Figure S2B). The deglycosylated peptide with the highest fold change is the one containing the N96 site from CD83 antigen (CD83), a well-known surface glycoprotein, which was up-regulated by 9.2-fold in cells treated with LPS. In dendritic cells, CD83 was reported to be pre-formed inside the cells and transported to the surface once the cells are activated.[235] The soluble form of CD83 has been well-documented and may interact with the TLR4/MD-2 complex (receptor for LPS) to reduce the inflammation.[342] The peptide containing the glycosylation site was also quantified from the experiment with THP-1 monocytes, but the secretion was not regulated (fold change=1.0), suggesting that the genetically identical cells responded differently to LPS. Five cytokines were quantified, including CSF1, tumor necrosis factor ligand superfamily member 15 (TNFSF15), GRN, BMP1, and CCL24. Among these, the deglycosylated peptide containing the N154 site from CSF1 was up-regulated by over 4 times, in contrast to the experiment with THP-1 monocytes where the peptide was not regulated. Yet, some deglycosylated peptides, such as those with the N258 site of TNFAIP6, were also up-regulated from cells treated with LPS, which is similar to the experiment with THP-1 monocytes. Some examples of deglycosylated peptides are displayed in Table 5.1.

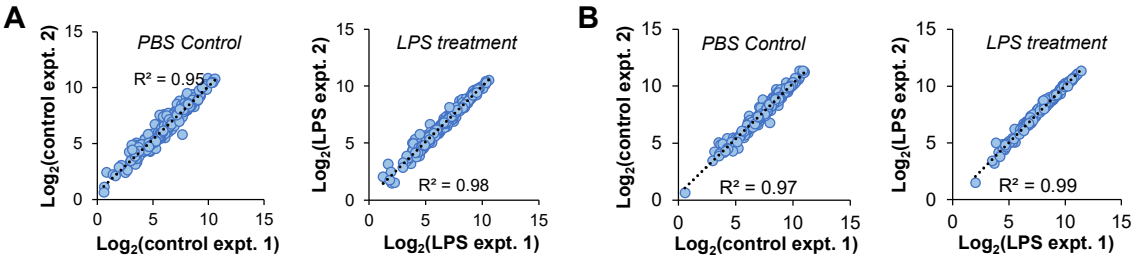

Figure 5.8. Quantification of glycoproteins and deglycosylated peptides from THP-1 M0 macrophages treated with LPS. Correlation of deglycosylated peptides (A) and glycoproteins (B) quantified from THP-1 M0 macrophages.

5.3.6 Analysis of secreted glycoproteins from Hep G2 cells treated with TGF-β

To further demonstrate the effectiveness of the current method, we applied it to study secreted glycoproteins from Hep G2 cells treated with TGF-β. TGF-β has been reported to induce the epithelial-to-mesenchymal transition (EMT) in Hep G2 cells, which is a process involved in cancer metastasis where epithelial cells become more invasive and mobile.[332, 343] The cells were pre-labeled with 100 μM Ac_4GalNAz for 12 hours and then treated with 10 ng/mL TGF-β for 48 hours before medium collection. We expanded the number of samples analyzed by employing the TMT 10-plex reagents and performing the experiment in biological triplicates (Figure 5.9A). We quantified 531 unique deglycosylated peptides from 236 glycoproteins (Figure 5.9B). The glycoproteins related to the extracellular vesicle are highly enriched with a very low P-value (P=8.8E-73), and 231 proteins are annotated with the keyword "Secreted" or "Signal". Furthermore, 147 are predicted to not have the transmembrane region but contain a signal peptide. We found 21 secreted glycoproteins being up-regulated and 6 being down-regulated in response to TGF-β. The proteins that were up-regulated are involved in processes such as extracellular matrix organization (Figure 5.9C). Latent-transforming growth factor beta-binding protein 2 (LTBP2) was up-regulated by over 6 times (Figure 5.9B). The protein was reported to be up-regulated at both the mRNA and protein levels after TGF-β stimulation,[344] corroborating the results from this experiment. Other glycoproteins that have also been reported to be up-regulated after TGF-β stimulation are also shown in the figure, including metalloproteinase inhibitor 1 (TIMP1) and lysyl oxidase homolog 4 (LOXL4).[345-346] Several examples of quantified deglycosylated peptides are listed in Table 5.1.

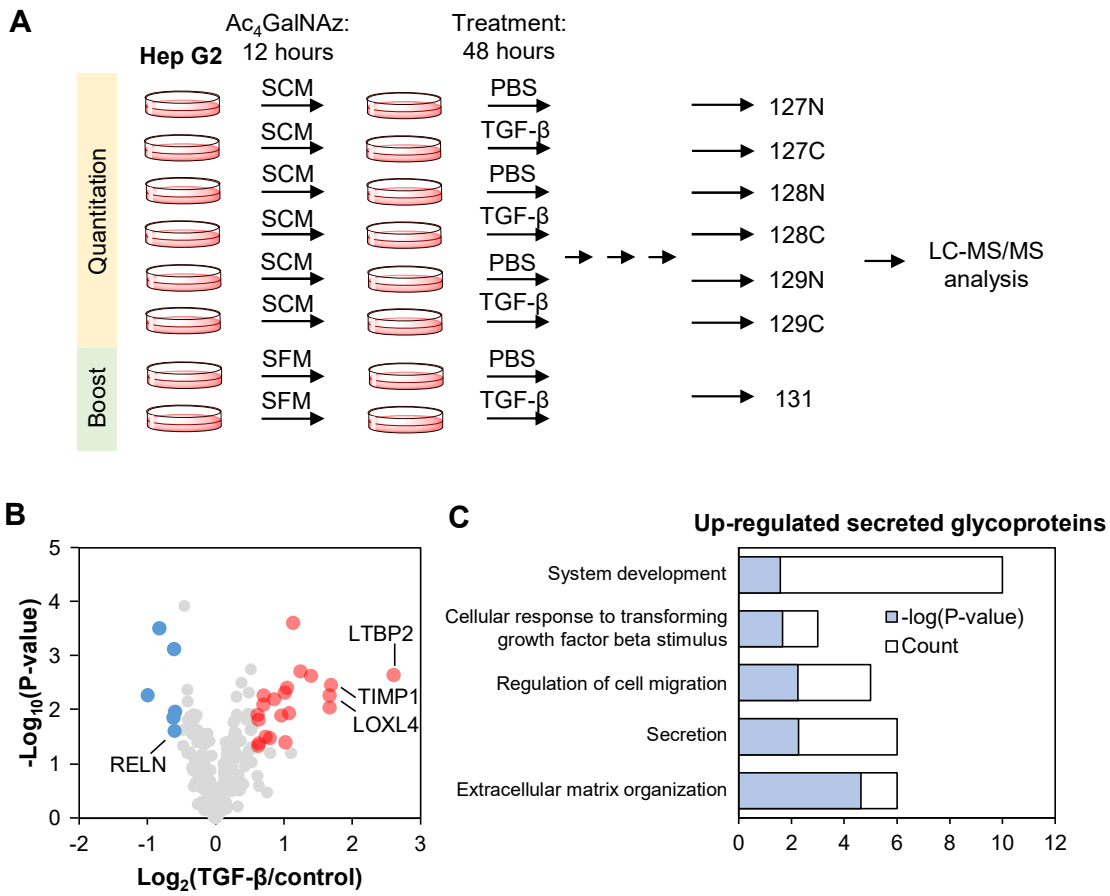

Figure 5.9. Quantification of secreted glycoproteins from Hep G2 cells treated with TGF-β. (A) Experimental scheme using the TMT 10-plex platform. (B) Glycoprotein abundance changes after the TGF-β treatment from the triplicate experiments. A t test was used to determine the significance (shown with P<0.05 and the ratio cutoff of 1.5). (C) GO clustering of secreted glycoproteins that are up-regulated after the TGF-β treatment based on biological process.

5.4 Conclusions

In this work, we coupled the selective enrichment of glycopeptides with a signal boosting approach to enhance the coverage of secreted glycoproteins with low abundances among many extremely high-abundance FBS proteins. This method was applied to globally quantify glycoproteins secreted from THP-1 monocytes and macrophages in response to LPS, and from Hep G2 cells treated with TGF-β. While the interference from highly

abundant FBS proteins on analysis of secreted proteins can be eliminated by using serum-free media, the serum starvation alters protein secretion even after a short period of time, and thus the quantification of secreted proteins is questionable. Here, we aimed to quantify secreted glycoproteins when cells are grown under the normal conditions, i.e., the serum-containing media. The enrichment of secreted glycopeptides alone is not sufficient for their comprehensive analysis because some high-abundance peptides are still present in the enriched samples due to non-specific binding. Moreover, LC-MS has its detection limit, which is biased against low-abundance glycoproteins. To overcome these issues, we combined the selective enrichment of glycopeptides from secreted glycoproteins with a boosting approach for global analysis of secreted glycoproteins by LC-MS. The boosting sample was carefully chosen by comparing the coverage of theoretical secreted glycoproteins from the cells or the media under serum-free or serum-containing conditions. The results demonstrated that the serum-free medium as a glycoprotein source can best serve as the boosting sample due to the higher coverage and the selectivity of secreted glycoproteins. Nonetheless, the quantitative information of secreted glycoproteins was obtained solely from the cells grown under the serum-containing conditions. We first applied this integrated method to globally quantify secreted glycoproteins in THP-1 monocytes and macrophages in response to LPS. Almost 400 deglycosylated peptides from ~200 secreted glycoproteins were quantified from THP-1 monocytes and macrophages treated with LPS. We further expanded the number of samples compared by using the TMT 10-plex reagents to determine secreted glycoproteins from Hep G2 cells treated with TGF-β in the EMT context, in which we quantified over 200 secreted glycoproteins related to the extracellular matrix organization. Here, the glycosylation site information was also

obtained, which provides the solid experimental evidence, minimizing the false-positive rates. Considering the importance of secreted glycoproteins, this method can be extensively applied in the biological and biomedical research fields.

Table 5.1. Example deglycosylated peptides from secreted glycoproteins quantified from THP-1 monocytes and macrophages in response to LPS or from Hep G2 cells treated with TGF-β.

Cell	UniProt ID	Gene	Deglycosylated peptide	Site	Mod Score	Ratio	Annotation
THP-1 Monocytes + LPS	P98066	TNFAIP6	K.N#TSTTSTGNK.N	258	150.77	23.10	Tumor necrosis factor-inducible gene 6 protein
	O15204	ADAMDEC1	R.EIKNN#QTEK.H	61	17.01	9.58	ADAM DEC1
	P29460	IL12B	R.KN#ASISVR.A	303	1000	7.51	Interleukin-12 subunit beta
	P07996	THBS1	K.KVSCPIMPCSN#ATVPDGECCPR.C	360	1000	4.11	Thrombospondin-1
	P05362	ICAM1	R.LNPTVTYGN#DSFSAK.A	267	127.53	2.46	Intercellular adhesion molecule 1
THP-1 M0 Macrophages + LPS	Q01151	CD83	R.N#TTSCNSGTYR.C	96	100.21	8.94	CD83 antigen
	P98066	TNFAIP6	K.N#TSTTSTGNK.N	258	150.77	8.34	Tumor necrosis factor-inducible gene 6 protein
	P09603	CSF1	K.NVFN#ETK.N	154	73.31	4.23	Macrophage colony-stimulating factor 1
	P17936	IGFBP3	K.VDYESQSTDTQN#FSSESKR.E	199	1000	4.06	Insulin-like growth factor-binding protein 3
	Q9BXX0	EMILIN2	K.SLN#DTMHRK.F	587	1000	0.62	EMILIN-2
	Q8IXL6	FAM20C	R.MVN#MTK.E	335	1000	0.60	Extracellular protein kinase FAM20C
	O14657	TOR1B	R.EERPLN#ASALK.L	64	1000	1.04	Torsin-1B

Table 5.1. Continued

Hep G2 + TGF-β	Q14767	LTBP2	R.DECWCPAN#STGK.F	421	1000	10.60	Latent-transforming growth factor beta-binding protein 2
	P19883	FST	R.CVCAPDCSN#ITWK.G	124	1000	3.65	Follistatin
	P15018	LIF	K.LN#ATADILR.G	138	1000	2.12	Leukemia inhibitory factor
	P01031	C5	R.AN#ISHK.D	741	1000	0.56	Complement C5

CHAPTER 6. DIFFERENTIAL REGULATION OF THE PROTEOME AND PHOSPHOPROTEOME OF NEUROBLASTOMA TUMORS IN RESPONSE TO CHEMOTHERAPEUTIC TREATMENT BY TOPO/CTX

6.1 Introduction

Neuroblastoma is a type of extracranial solid tumor found in children especially in early childhood.[347-348] According to the American Cancer Society, the occurrence of neuroblastoma is about 6% of all cancers in children.[349] Neuroblastoma tumors typically have highly heterogeneous clinical characteristics.[350-351] Primary tumors can form in different locations such as the adrenal glands, abdomen, chest, orbital of the eye, and bones, and the symptoms are linked to the location of these primary tumor sites, including bone pain, fever, lumping, swelling, and weight loss.[348] In the tumor microenvironment, other cells such as myeloid cells and fibroblasts may be recruited to enhance the chemoresistance and immune evasion of the tumor. The immune response may be mediated by antibodies, natural killer cells, and T cells.[352-354] Metastasis to the lymph node, bone marrow, and bone are detected in about 50% of patients at diagnosis depending on the age and biological characteristics of the tumors.[348, 355] Even though it is not certain what causes genetic alterations in these cells, several genes have been identified, with most malignant neuroblastoma tumors having the amplification of the *MYCN* oncogene while the rare familial ones having the mutation of the *ALK* gene.[356-357] In addition, other environmental factors, such as maternal use of drugs, hair dyes, and pesticides in residential area, may be associated with neuroblastoma.[358-360]

Based on the criteria established by the International Neuroblastoma Risk Group (INRG), patients can be classified into very low-, low-, intermediate-, and high-risk groups.[361] While some neuroblastoma may disappear without any treatment, the survival rates can vary from over 95% for low-risk patients to 40-50% for high-risk ones.[362] Relapse occurs in about 50-60% of high-risk patients.[347] A number of compounds have been used to target proteins associated with neuroblastoma, such as aurora kinase inhibitors that inhibit cell-cycle regulator proteins aurora kinase A and aurora kinase B.[363-364] Combinations of drugs have also been used, including topotecan (TOPO) and cyclophosphamide (CTX), which is one of the most effective approach to treat relapse patients.[348, 365] TOPO intercalates between DNA strands and causes replication-mediated DNA damage, i.e., as DNA topoisomerase I (TOP1) unwinds DNA with the intercalated TOPO during replication, it cannot religate the strand. Consequently, the TOP1 cleavage complex (TOP1cc) is formed and collapses with the replication fork, resulting in DNA damage and eventually apoptosis. The drug is very selective, and its interaction with DNA is only stable in the complex with TOP1.[366] CTX is a DNA alkylating reagent. This prodrug is converted by hepatic cytochrome P-450 into the active 4-hydroxycyclophosphamide that eventually crosslinks two guanine bases together, and thus DNA replication is interfered. Normal cells are minimally affected by CTX since detoxification can be achieved by the highly abundant aldehyde dehydrogenase.[367] CTX was originally used in chemotherapy, and has recently been repurposed as a potent immune modulator in cancer immunotherapy by targeting suppressive regulatory immune cells within the tumor microenvironment. This enhances effector and memory cytotoxic T cells, which further increases the body immune response against the tumors.[368-369]

With recent technological advances, multiple omics approaches have been employed to study the characteristics and treatments of neuroblastoma, including genomics, metabolomics, transcriptomics, and proteomics, with the goal of obtaining a personalized treatment.[363] Next-generation sequencing allows for a faster and deeper analysis of the genome, which has obtained meaningful results such as the newly discovered missense mutation in the *ALK* gene and the mutation in the bromodomain PHD finger transcription factor from whole exome sequencing.[370-371] Imperiale et al. described the difference in metabolic contents, such as a higher level of creatine and glutamine, from neuroblastoma patients compared with normal adrenal medulla using nuclear magnetic resonance spectroscopy-based metabolomics.[372] Egler et al. employed SELDI-TOF MS to analyze proteins from plasma samples of high-risk neuroblastoma patients and identified seven biomarker candidates, which need further verification.[373] Proteomic analyses in other models, including mice and cell culture, have also been performed.[374-375] However, proteomic and phosphoproteomic analyses of neuroblastoma tumors treated with TOPO and CTX have yet to be reported.

In this work, we systematically quantified protein expression and phosphorylation changes in neuroblastoma tumors from the IMR5 cell line-derived xenograft (CDX) and COG424X patient-derived xenograft (PDX) models generated in mice after the treatment with TOPO/CTX. Multiplexed proteomics using the TMT10plex reagent, extensive peptide fractionation, and phosphopeptide enrichment by TiO_2 beads were employed to uncover the cellular response to the treatment. Among the two untreated tumors, several proteins and phosphoproteins were differently expressed such as those involved in mitotic cell cycle, programmed cell death, and DNA integrity checkpoint regulation. The response

to the treatment centered around the regulation of the cell-cycle process, which is one of the well-known cellular responses to DNA damage. For the CDX model, the response occurred at the proteome level while that in the PDX model occurred at the phosphoproteome level, indicating the sensitivity of each tumor to the treatment. Several protein targets, phosphorylation events, and kinases that may possibly be used in neuroblastoma treatment have also been identified. Among those, Ca^{2+}/calmodulin-dependent protein kinase II (CAMK2) may be responsible for the lower sensitivity of COG424X as determined by phosphorylation motif analysis, corresponding very well with the kinase activation by autophosphorylation detected in this tumor. The results will lead to a better understanding and aid in a better treatment of neuroblastoma with TOPO/CTX.

6.2 Materials and methods

6.2.1 Generation of xenograft models and TOPO/CTX treatment

All reagents were from Sigma-Aldrich unless noted otherwise. *In vivo* studies were performed using SCID mice (Jackson) between the ages of 4 and 6 weeks old. To establish subcutaneous tumors, Matrigel (Corning) was frozen at -80 °C in 200 µl aliquots and thawed on ice at 4°C for 3 hours prior to cell harvest. IMR5, a human derived neuroblastoma cell line, (courtesy of COG Cell Line Repository and Michael D. Hogarty, MD, Laboratory, Children's Hospital of Philadelphia), was grown in RPMI-1640 (Life Technologies) supplemented with 10% fetal bovine serum, 2 mM L-Glutamine, and 100 U/mL of penicillin. Tissue culture was performed at 37 °C in a humidified atmosphere of 5% CO_2. At confluence, cells were harvested and suspended in 200 µl of Matrigel matrix at a concentration of 10^6 cells/mL and injected subcutaneously into the flank. To develop

COG424X tumors, chunks were viably frozen in filtered 90% FBS (Gemini) and 10% dimethyl sulfoxide at -80 °C and stored in long term liquid nitrogen storage. To thaw, these chunks were placed in a hot water bath at 37°C for 5 minutes. Tumor chunks were mechanically dissociated and resuspended in Matrigel at a concentration of 10^6 cells/mL. When established tumors reached approximately 150 mm^3, the mice began randomized treatment. Tumor growth was assessed 2-3 times per week by caliper measurement by a single technician with volume calculated using the formula (length x width x height) x 0.52 mm^3. For treatment, mice received intraperitoneal injections of either vehicle control (normal saline, daily for 5 days) or a combination of cyclophosphamide (20 mg/kg) with topotecan (0.05 mg/kg). Animals were euthanized 1 day after treatment ended and the tumors were excised. All animal work was performed under a protocol approved by the Children's Hospital of Philadelphia and Emory Institutional Animal Care and Use Committee.

6.2.2 Protein extraction and peptide purification

The tumors were roughly dissociated into individual cells using the blunt end of a 10-mL syringe plunger in ice-cold PBS buffer, collected by centrifugation at 500*g*, 4 ºC for 5 minutes, and washed with ice-cold PBS buffer twice to remove contaminants from blood. The cells were lysed in a Dounce tissue grinder (Wheaton) with 40 strokes on ice using the lysis buffer containing 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH=7.4), 1% sodium deoxycholate, 150 mM sodium chloride, 1 pill/10 mL cOmplete mini protease inhibitor cocktail (Roche), 20 units/mL benzonase, 50 mM sodium fluoride, 50 mM β-glycerophosphate disodium salt hydrate, 1 mM sodium orthovanadate, and 1 mM phenylmethanesulfonyl fluoride. The lysates were collected after centrifugation

at 25,830*g*, 4 ºC for 10 minutes to remove the cell debris, and the protein concentration was measured with BCA assay (Thermo). The proteins were reduced with 5 mM dithiothreitol at 56 ºC for 25 minutes, alkylated with 14 mM iodoacetamide at room temperature in the dark for 30 minutes, and purified with the methanol/chloroform precipitation method.[160] The protein pellets were resuspended in a digestion buffer containing 50 mM HEPES (pH=8.8), 1.6 M urea, and 5% acetonitrile (ACN), and digested with sequencing-grade modified trypsin (Promega) with the enzyme to protein ratio of 1:100 at 37 ºC with shaking for 16 hours. The digestion was quenched by acidification with trifluoroacetic acid (TFA, EMD Millipore) to pH<2. The supernatants were collected after centrifugation at 5,000*g* for 10 minutes. The peptides were desalted with a Sep-Pak tC18 vac cartridge (Waters), split equally into two parts for duplicate analysis, and dried in a vacuum concentrator.

6.2.3 Peptide labeling with TMT10plex

About 1 mg peptides from each sample, based on the protein concentration measured with the BCA assay, were resuspended in a buffer containing 200 mM HEPES (pH=8.5) with 40% ACN. Eight out of the ten TMT10plex (Thermo) channels were resuspended in 41 µL anhydrous acetonitrile and used to label the peptide samples as followed: 127N for control IMR5 #1; 127C for control IMR5 #2; 128N for treated IMR5 #1; 128C for treated IMR5 #2; 129N for control COG424X #1; 129C for control COG424X #2; 130N for treated COG424X #1; 130C for treated COG424X #2. The reaction was performed at room temperature with shaking for 1 hour and subsequently quenched with 8 µL of 5% hydroxylamine hydrochloride in 200 mM HEPES (pH=8.5) at room temperature with shaking for 15 minutes.

6.2.4 Peptide fractionation for proteomics analysis

About 50 μg of the labeled peptides from each sample were mixed in a 1:1:1:1:1:1:1:1 ratio according to a test run, dried using a vacuum concentrator, and desalted with the Sep-Pak cartridge. After being dried in the vacuum concentrator, the peptides were fractionated with high-pH chromatography using an Agilent 1260 Infinity HPLC and an XBridge C18 (3.5 μm, 4.6 x 250 mm) column (Waters). Buffer A was 10 mM ammonium formate (pH=10) in water and buffer B was 10 mM ammonium formate (pH=10) in 90% ACN. The peptides were separated by a 40-minute gradient of 5% to 70% buffer B and collected every one minute into 40 fractions, consolidated into 25 fractions, and dried in the vacuum concentrator. Eventually, the peptides were purified with StageTip before LC-MS/MS analysis.[161]

6.2.5 Peptide fractionation and enrichment for phosphoproteomics analysis

Similar to the proteome analysis, the rest of the peptides after TMT labeling were mixed equally according to the test run, fractionated with high-pH HPLC and collected during a 40-minute gradient of 5-55% buffer B, consolidated into 25 fractions, desalted with the Sep-Pak cartridge, and dried in the vacuum concentrator. Phosphopeptides from each fraction were enriched with ~1 mg 5 μm Titansphere TiO_2 beads (GL Sciences). The beads were washed twice with a binding buffer containing 30% ACN, 1% TFA, and 70 mM glutamic acid. The phosphopeptides were enriched with the beads in the binding buffer at 37 ºC with shaking for 1 hour. The supernatant was collected and re-enriched similarly with another 1 mg of the beads. The beads were washed twice with a wash buffer containing 30% ACN and 1% TFA. The enriched phosphopeptides were eluted three times

with 40% ACN and 15% ammonium hydroxide for 10 minutes at 37 ºC with shaking. The eluates were pooled, dried, and purified with StageTip before LC-MS/MS analysis.

6.2.6 LC-MS/MS analysis

The peptides were dissolved in a solution containing 5% ACN and 4% formic acid and loaded into a microcapillary column packed in-house (Magic C18AQ, 3 μm, 200 Å, 75 μm x 16 cm). The peptides were separated with reverse-phase liquid chromatography on a Dionex UltiMate 3000 UHPLC system (Thermo) using 112-minute gradients of 3-14%, 4-17%, 8-24%, and 15-40% of ACN with 0.125% formic acid for fractions 1-5, 6-12, 13-23, and 24-25, respectively. The UHPLC was connected on-line to an LTQ Orbitrap Elite Hybrid Mass Spectrometer (Thermo). For each cycle, a full MS spectrum was recorded in the Orbitrap at the resolution of 60,000 with $1x10^6$ automatic gain control (AGC). Peptide fragmentation was performed by higher-energy collisional dissociation (HCD) with the normalized collisional energy of 40% using a Top15 method, i.e., up to 15 MS/MS were performed for the 15 most abundant peaks with an isolation width of 1.2 m/z. Spectra of the fragment ions were recorded in the Orbitrap cell at the resolution of 30,000 and $2x10^5$ AGC.[376] Selected ions were excluded from further sequencing for 90 seconds. Ions with a single or unassigned charge states were not sequenced. Maximum ion accumulation times were 1,000 ms or 50 ms for each full MS scan or each MS/MS scan, respectively.

6.2.7 Data processing and database searching

The raw files were converted into an mzXML format and searched using the SEQUEST algorithm (version 28).[217] The spectra were matched against the human

proteome downloaded from UniProt (www.uniprot.org). The peptide mass tolerance was 20 ppm, and the fragment ion mass tolerance was 0.025 Da. The maximum number of missed cleavages was 3 and the maximum number of differential modifications was 4. The following amino acid modifications were used; static modifications of +57.0215 Da for the carbamidomethylation of cysteine and +229.1629 Da for the TMT labeling at lysine and the N-terminus, and differential modification of +15.9949 Da for the oxidation of methionine. For the phosphoproteome experiment, an additional differential modification of +79.9663 Da for the phosphorylation of serine, threonine, and tyrosine were included. The target-decoy method, where each protein sequence is listed in both forward and reverse orders, was employed to estimate the false discovery rates (FDRs) of peptide and protein identifications.[218] Linear discriminant analysis was applied to control the quality of peptide identification using multiple parameters such as XCorr, ppm, peptide length, and charge state. The FDR for peptide identifications was controlled to <1%, and an additional filter was also performed to control the FDR to <1% at the protein level. For the phosphoproteomic analysis, the dataset was limited to only phosphopeptides and phosphoproteins when calculating the FDRs at both levels.

6.2.8 Bioinformatics analysis

Data were processed with Microsoft Excel and Perseus.[165] Contaminants and falsely identified peptides were removed. The signal-to-noise (S/N) ratios for each of the eight TMT channels were recorded and corrected according to the manufacturer's instruction for the quantification experiment. The TMT peak match tolerance was 0.003 Da and the nearest m/z peak in that tolerance window was chosen for each reporter ion. Peptides with S/N less than 5 were removed. The signals from all channels were normalized

assuming equal loading of all eight channels so that the summed S/N of all peptides from each channel is equal to the others in the proteome experiment.[377] The TMT protein intensity was eventually calculated from the median TMT S/N of all peptides, and the final protein intensity was from the average of the two replicates. The ratios for statistical analyses of the treated and control groups were from the same N and C channels of the 127 and 128, or 129 and 130 of the TMT reagents, i.e., the ratios of the treated to the control of the IMR5 tumors were calculated from 128N/127N and 128C/127C while those for the COG424X tumors were from 130N/129N and 130C/129C. Gene-ontology (GO) clustering of proteins were performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID).[378-379] For the volcano plots, the statistical significance of the regulation of protein abundance was calculated using a one-sample t-test (S_0=0). Proteins were considered being up- or down-regulated when the abundance changed by at least 1.5-fold ($\log_2(1.5)$=0.58) from the control group and the P-value is less than 0.05. ANOVA was performed using S_0=0 and the false discovery rate was set to 0.05 using permutation-based FDR.

For the regulation of phosphopeptides, the criteria were set similar to the proteome experiment. The median intensity is reported for each unique phosphopeptide and used when determining the regulation of phosphorylation event. A probabilistic-based algorithm called Ascore was used to determine the confidence of the phosphorylation site localization.[219] An Ascore of 13 showing 95% confidence for the site localization is reported. When analyzing the phosphorylation motifs, the dataset was first filtered so that only peptides with a single phosphorylation site and an Ascore of >13 were kept. Motif

analysis was performed using MEME suite based on motif-X algorithm.[380-381] The minimum number of occurrences was 20 and the cutoff P-value was 0.000001.

6.3 Results and discussion

6.3.1 Identification and quantification of proteins and phosphoproteins by MS-based multiplexed proteomics

In this study, we aimed to explore protein expression and cellular signaling changes in two neuroblastoma xenograft models after TOPO/CTX treatment. CDX and PDX models were generated in mice from IMR5 neuroblastoma cell line and COG424X neuroblastoma tumor from a patient, respectively. The models were treated with a five-day regimen of the drugs. The control and treated tumors were harvested, and the proteome and phosphoproteome changes were analyzed using MS-based proteomics combining with TMT10plex isobaric tagging. Extensive fractionation with high-pH chromatography was employed to increase the proteome coverage. In addition, titanium dioxide (TiO_2)-based enrichment was also performed in the phosphoproteome experiment to enrich all phosphopeptides from the samples (Figure 6.1).

In the proteome experiment, we identified 113,120 total and 60,855 unique peptides, corresponding to 6,640 proteins with <1% FDR at both the protein and peptide levels. The identification covered a broad range of proteins in several cellular compartments and biological processes according to a GO clustering analysis. While in the phosphoproteome experiment, 12,984 unique phosphopeptides from 2,465 phosphoproteins were identified (Table S2). The majority of the identified phosphoproteins were from the nucleus, and those involved in cellular macromolecule metabolic process,

gene expression, and cell cycle process were highly enriched according to GO clustering analysis. Combining the results from both the proteome and phosphoproteome experiments, we identified a total of 7,275 proteins, with 1,830 common proteins from both experiments (Figure 6.2A).

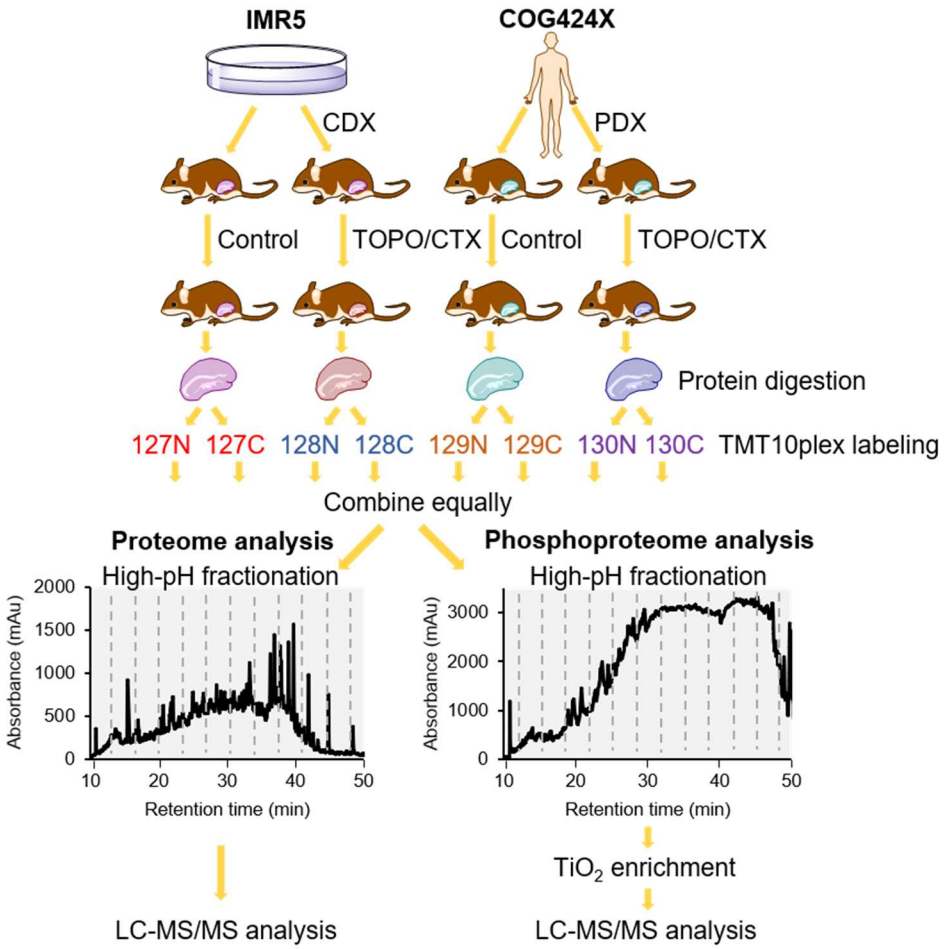

Figure 6.1. Experimental procedure. CDX and PDX models were generated in mice from IMR5 cell line and COG424X tumor, respectively. The tumors were either treated with a five-day regimen of TOPO/CTX or used as a control. The proteins were extracted from the tumors and digested into peptides, which were subsequently labeled with TMT10plex reagents, fractionated with high-pH, reverse-phase chromatography, and analyzed with LC-MS/MS. Phosphopeptides were also enriched with TiO2 beads.

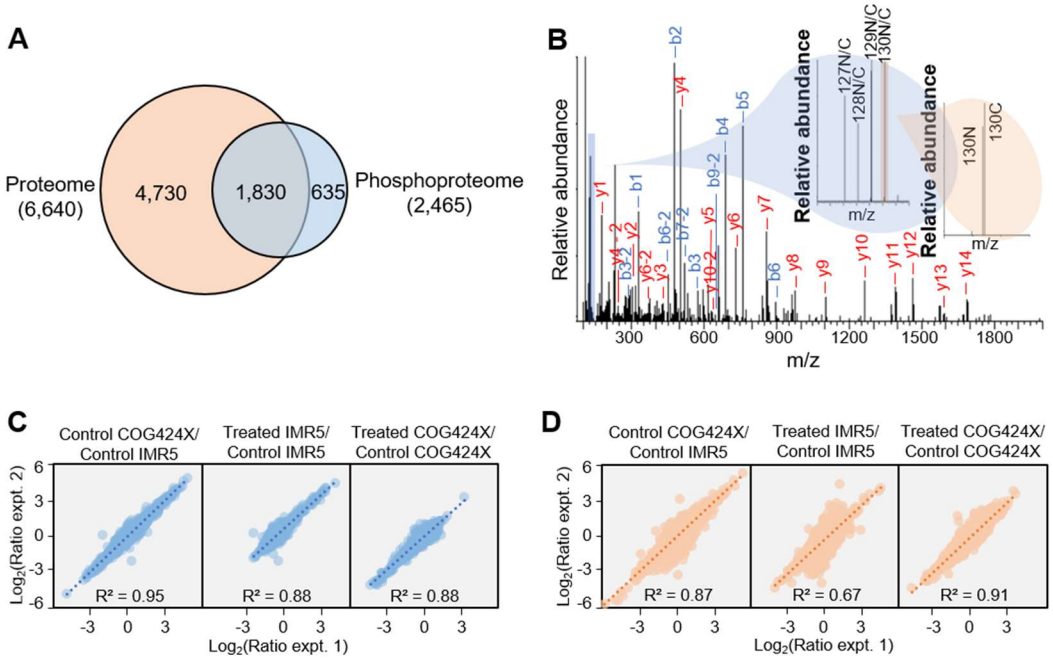

Figure 6.2. Proteins and peptides identified from the experiments. (A) Overlap of proteins identified from the proteomic and phosphoproteomic experiments. (B) An example tandem mass spectrum of a phosphopeptide. The insets show the region where TMT reporter ions are used for the quantification. (C) and (D) Reproducibility of changes in protein abundance (C) or phosphopeptide abundance (D) from replicate experiment.

Since the isobaric TMT reagents were used, the same peptides from different samples coeluted and fragmented together, and the peptide identification and quantification can be obtained from the same tandem mass spectrum across multiple samples and treatments. A powerful Orbitrap mass spectrometer with HCD fragmentation was employed for the identification and quantification of proteins. The resolution and AGC used in MS/MS analysis were carefully chosen so that the neutron-encoded reporter ions can be resolved, and the coalesce artifact was reduced.[376] An example tandem mass spectrum of phosphopeptide TFPLAHS#PQAECEDQLDAQER (# represents the phosphorylation site) is shown (Figure 6.2B). The peptide is from protein KAT8 regulatory NSL complex subunit 1 (KANSL1), which is involved in acetylation of nucleosomal

histone H4. We confidently identified this phosphopeptide with XCorr of 5.42 as well as Ascore of 33.36, showing the well localization of the phosphorylation site on S1045. The insets of the figure show clusters of the TMT reporter ions with the neutron encoded N and C peaks that can be resolved at the minimum resolution of 30,000 in the Orbitrap cell. These reporter ions in the m/z=127-130 region are used for peptide quantification. After filtering out some peptides based on the S/N ratio (see the method section), 6,602 proteins were quantified across the eight samples in the proteome experiment with relatively high reproducibility, while a total of 12,213 unique phosphopeptides from 2,458 proteins were quantified in the phosphoproteome experiment. The correlation of abundance changes of proteins or phosphopeptides among different sample groups is relatively high, with the lower correlation in the phosphoproteome experiment that might be due to the low abundance of phosphopeptides affecting the quantification accuracy (Figure 6.2C and D).

6.3.2 Protein and phosphoprotein expression profiles of IMR5 and COG424X tumors

Even though both tumors are considered as neuroblastomas, the protein expression is expected to be different because of the intrinsic heterogeneity and the different tumor sources. We first compared the protein expression in normal state of both tumors and found that a total of 1,539 proteins, corresponding to 23% of the quantified proteins, were significantly differently expressed (Figure 6.3A). In the IMR5 tumor, 666 proteins were expressed with higher abundance, and they participate in processes such as membrane organization, exocytosis, respiratory electron transport chain, and substance transport processes. COG424X tumor expressed more of 873 proteins, which are involved in cytoskeletal organization, phosphate-compound metabolic process, and regulation of transferase activity. Proteins involved in mitotic cell cycle and programmed cell death, two

characteristics that allow for cancer progression, were highly up-regulated in COG424X. Interestingly, the major molecular function of these differently expressed proteins is RNA binding, indicating that the difference in protein expression of both tumors might result from regulation through non-covalent RNA bindings. Nonetheless, multiple proteins engaging in the same biological processes, such as nervous system development and cellular homeostasis, were differently expressed in the tumors. The highest protein expression in IMR5 compared to COG424X is from 2-hydroxyacylsphingosine 1-beta-galactosyltransferase (UGT8), which catalyzes the synthesis of galactocerebrosides and is expressed 29 times higher in IMR5. The protein was recently found to be elevated in urine of patients with basal-like breast cancer, and the knockdown of its gene resulted in suppression of tumorigenicity and metastasis.[382] For COG424X, histone H1.3 (HIST1H1D) is the highest expressed protein compared to that in IMR5 by 28 times. This protein was previously proposed as a biomarker for pancreatic ductal adenocarcinoma.[383] In spite of that, the use of these proteins as general biomarkers for neuroblastoma would be problematic since their expression is distinct for each tumor type.

Phosphoproteomic analysis also revealed different extents of phosphorylation between the two tumor types (Figure 6.3B); 736 phosphopeptides from 290 phosphoproteins have higher abundance in IMR5 tumor, while 1,482 phosphopeptides from 750 phosphoproteins have higher abundance in COG424X tumor. The majority of GO terms enriched from these phosphopeptides based on their phosphoproteins are found in both tumors. Nonetheless, phosphoproteins involved in processes such as protein-DNA complex organization and regulation of cell projection organization are enriched only in a specific tumor. Among the extremes, a phosphopeptide from monocarboxylate transporter

1 (SLC16A1) with pS498 is expressed over 50 times higher in IMR5 than COG424X while its protein abundance is relatively similar. This protein was previously found to be the direct Wnt target in colon cancer, and its high expression affected the tumor sensitivity to drugs that target cancer metabolism due to the Warburg effect.[384] The higher phosphorylation level may indicate the activation or inhibition of transport of molecules such as lactate that is a byproduct of the Warburg metabolism in the IMR5 tumor.

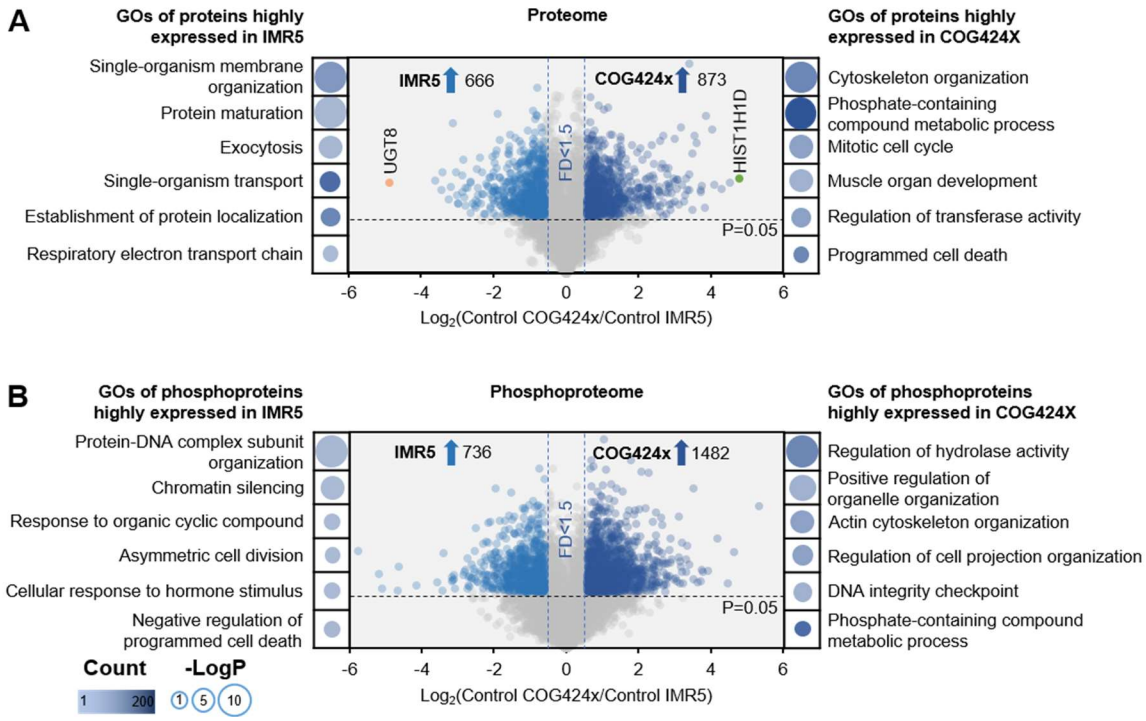

Figure 6.3. Volcano plots show proteins and phosphopeptides that were differently expressed in the control IMR5 (A) or COG424X (B) tumors. The cutoff for significant regulation was P<0.05 and protein ratio change of 1.5. There enriched GO terms based on biological functions are also shown. FD means fold difference.

Proteins that have been previously reported as markers for neuroblastoma were also quantified, including microtubule-associated protein 2 (MAP2),[385] with both the canonical form and isoform 4 detected and having similar expression in both tumors. The high level of multidrug resistance-associated protein 1 (MRP) was found to indicate the poor

outcome, and its expression is also similar in both models.[386] Meanwhile, the high expressions of CD44 antigen (CD44) and high affinity nerve growth factor receptor (NTRK1) could indicate the good clinical outcome.[387] The abundance of CD44 is slightly higher in COG424X while NTRK1 was not detected, which may due to several mutations of its gene in neuroblastoma.[388] At the phosphoproteome level, phosphopeptides from proteins that participate in the negative regulation of programmed cell death were enriched in IMR5, which could result in the tumor's greater ability to evade tumor suppression processes. In a study from Xie et al., three protein signatures that could be used to predict the outcome of neuroblastoma were investigated, including RNA-binding protein Musashi homolog 1 (MSI1), DNA-binding protein inhibitor ID-1 (ID1), and proliferating cell nuclear antigen (PCNA).[389] While ID1 was not detected in our study, MSI1 was quantified with similar abundance in both tumors, indicating the similar prognosis. However, COG424X expressed over 2-fold of PCNA compared to IMR5, which could show that COG424X is more proliferative among the tumors even though proteins relevant to programmed cell death are up-regulated, suggesting that there may be an alternative pathway or proteins responsible for the progression of COG424X.

6.3.3 Direct effects of TOPO/CTX treatment on proteins

The previous section showed that the untreated tumors differentially expressed proteins and phosphoproteins. Accordingly, each tumor might employ a unique mechanism to respond to the treatment. TOPO affects TOP1 through the formation of DNA-TOPO-TOP1 complexes, i.e., TOP1ccs, which inhibits DNA replication and eventually leads to TOP1 degradation through the 26S proteasome-mediated signaling pathway as a method for cells to restore to the normal conditions, and is not always found in all cancers.[390-391] In

the control tumors, the abundance of TOP1 in IMR5 is slightly higher than that in COG424X. After the treatment, the abundance of the protein decreased significantly by 1.66 and 1.47 times in IMR5 and COG424X, respectively (Figure 6.4A, D, and E). The degradation of TOP1 could lead to an up-regulation of DNA topoisomerase 2-alpha (TOP2A), which, in contrast to TOP1, cleaves both strands of the DNA simultaneously to relieve the strain due to the absence of TOP1.[392] We observed a similar change, i.e., the abundance of TOP2A increased significantly after the treatment by 1.73 and 1.15 times in IMR5 and COG424X, respectively (Figure 6.4A, 4D, and 4E). For CTX, the drug crosslinks two guanine bases together and, consequently, DNA replication is inhibited. Unlike the mechanism of TOPO, the inhibition by CTX does not directly target a specific protein. Nonetheless, the abundance of some proteins in the process were affected. CTX is a prodrug that must be activated by hepatic cytochrome P450, which is primarily expressed in the liver and was not detected in this experiment.[393] Retinal dehydrogenase 1 (ALDH1A1) is highly abundant in normal cells and has a role in detoxification of the drug.[394] The increased expression of this protein is known to be involved in tumor resistance to CTX.[395] The protein was detected and its abundance increased significantly in the IMR5 tumor, but only slightly in the COG424X tumor, potentially a cellular response to CTX treatment (Figure 6.4C). While individual target proteins for each drug were investigated in this section, it should be noted that the treatment using a combination of both TOPO and CTX may also have combinatorial, off-target, and secondary effects. Yet, the results in this part demonstrated that the treatment affected their targets as expected.

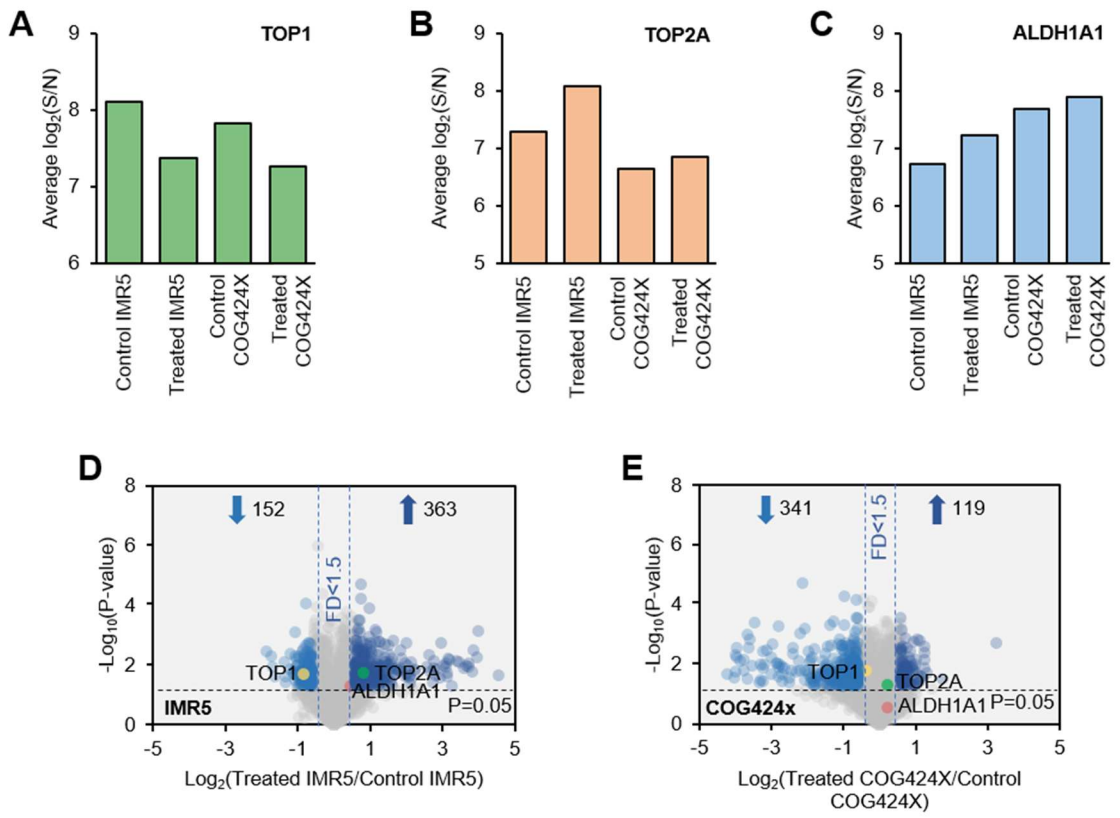

Figure 6.4. (A), (B), and (C) Protein abundance changes under the TOPO/CTX treatment including DNA topoisomerase 1 (TOP1) (A), DNA topoisomerase 2-alpha (TOP2A) (B), and retinal dehydrogenase 1 (ALDH1A1) (C). P-values were calculated by unpaired t-tests. The error bars represent one standard deviation. (D) and (E) Volcano plots show proteins that were up- and down-regulated in IMR5 (D) or COG424X (E). The cutoff for significant regulation was P<0.05 and protein ratio change of 1.5. FD means fold difference.

6.3.4 Protein expression pleiotropically affected by TOPO/CTX treatment

Besides the target proteins above, we systematically quantified the abundance changes of over 6,000 proteins in both tumors after TOPO/CTX treatment. To get an overview of the protein expression changes, an unsupervised hierarchical clustering analysis was performed based on the protein expression profile of each tumors (Figure 6.5). The control and treated IMR5 tumors clustered together, as did the control and treated COG424X. The protein expression varies across the tumors, and the treatment further

altered the expression of these proteins to the level that still resembles each original tumor. An ANOVA test revealed that 68% proteins are differently expressed in at least one tumor. The proteins have also been grouped based on their expression patterns across the samples and the enriched GO terms based on biological process are displayed.

We then compared proteins that were affected by the treatment in each tumor type, shown as volcano plots in Figure 6.4D and Figure 6.4E. Proteins that participate in the nervous system development processes were generally not affected by the treatment in both tumors. In the IMR5 tumor, 363 proteins were up-regulated while 152 were down-regulated after the treatment, corresponding to 5.5% and 2.3% of the total quantified proteins, respectively (Figure 6.4D). Proteins involved in the cell cycle and cytoskeletal processes were up-regulated, while those participating in nucleic metabolic processes and gene expression were down-regulated. In the COG424X tumor, 119 proteins were up-regulated while 341 were down-regulated corresponding to 1.8% and 5.2% of the total quantified proteins, respectively (Figure 6.4E). Proteins involved in the cell cycle process were still up-regulated even though, compared to the IMR5 tumors, the number of these proteins decreased by over 50%. Moreover, proteins involved in cytoskeleton processes were mostly down-regulated in contrast to the results from IMR5, and those that regulate the nucleic metabolic processes were not enriched among the down-regulated proteins.

The abundance changes of these proteins can be explained as the cellular response to DNA breakage caused by the treatment and the efforts of these cancerous cells to revert to their normal proliferative state. The response, including nucleic acid regulation, gene expression, and cytoskeletal processes, center around the regulation of the cell-cycle process, which is expected when cells undergo a treatment with DNA damaging ability.

The formation of DNA-TOPO-TOP1 complexes and the alkylation of DNA strands by CTX initially halt DNA replication or RNA transcription. Cells may try to counter the effects by degrading TOP1 and regulating other proteins involved in the degradation mechanisms. Eventually when the DNA is damaged, cells repair it by the regulation of proteins in the cell cycle so that DNA strands can be repaired through several checkpoints or DNA-repair mechanisms. One of the most important proteins in the cell-cycle process is cyclin-dependent kinase 1 (CDK1), which coordinates with cyclins to regulate the cell-cycle transition. CDK1 was up-regulated by 2.4 times in IMR5 and 1.7 times in COG424X. The up-regulation of cytoskeletal process-related proteins and cyclin-A2 (CCNA2) suggests that the majority of IMR5 cells after the treatment could be in the S/G2/M phase, which requires structural proteins to segregate the two daughter cells apart and the presence of phase-specific cyclin, and is supported by the enriched GO term for G2/M phase transition. It may also indicate that the cells were halted in a specific phase that caused the accumulation of these proteins.[396] Meanwhile, the down-regulation of cytoskeletal proteins in the COG424X tumor may result from a different cellular response or alternative pathways/mechanisms and will be discussed further in the phosphoproteome analysis section. Proteins involved in programmed cell death were also up-regulated in both tumors, which is the expected outcome of the treatment. Interestingly, those involved in the negative regulation of programmed cell death were also up-regulated only in the IMR5 tumor, which may show its ability to counter the effects of the treatment.

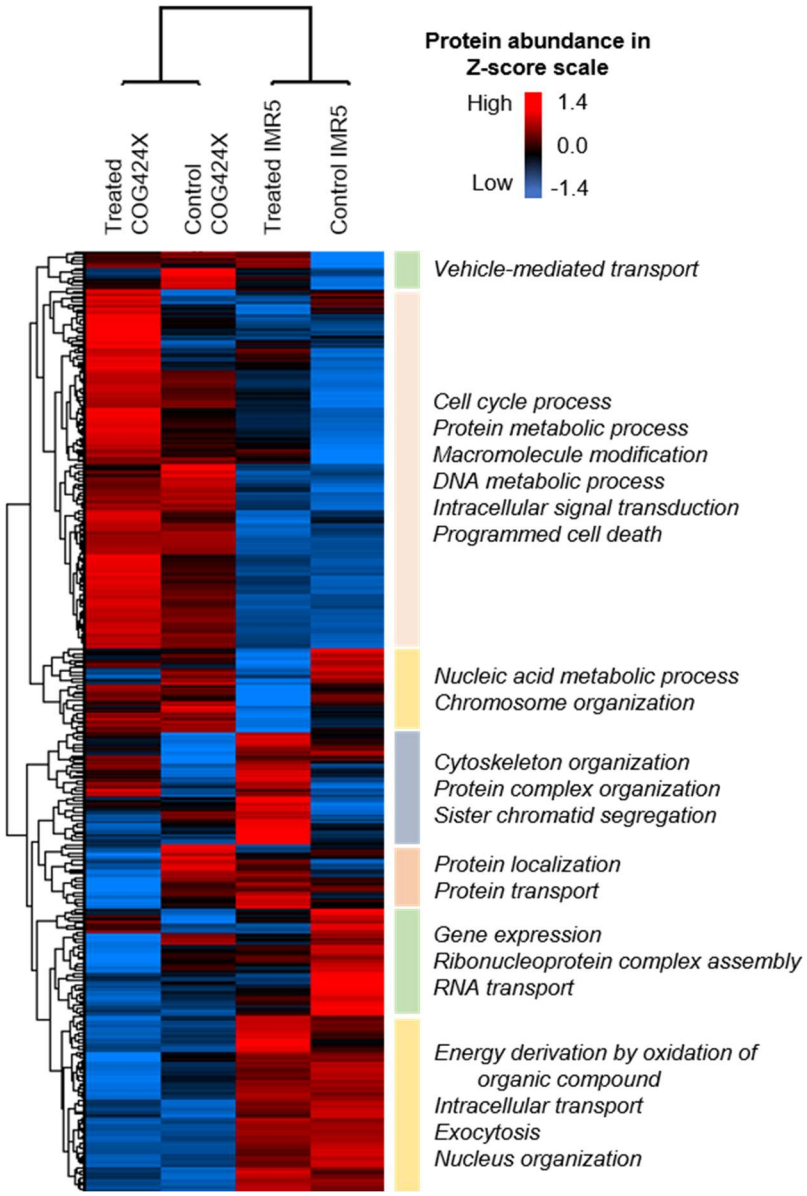

Figure 6.5. Hierarchical clustering and heat map of protein abundances from the proteome experiment. TMT signals were $\log_2$- and Z-score-transformed. Differential expression of proteins from each part of the heat map were clustered and the GO enriched terms based on biological process are shown in the plot.

6.3.5 Phosphorylation changes in response to DNA strand breakage

Protein phosphorylation plays extremely important roles in the regulation of cell signaling, and the large-scale analysis of these proteins can reveal the signaling pathway

changes in the tumor cells after the drug treatment. Because of the TOPO/CTX treatment, we first looked into phosphorylation events that are indicative of DNA strand breakage. One of the hallmarks for this event is the phosphorylation on S139 of histone H2AX (H2AFX), a variant of histone H2A, to produce γH2AX by serine-protein kinase ATM (ATM).[397] We quantified H2AFX protein as well as several phosphorylation sites, including S140 (the protein is cleaved at the N-terminus after the translation and therefore previous reports referred to this site as S139). This phosphorylation event triggers other cellular activities that lead to DNA repair and arrest of cell cycle progression, where two major DNA repair pathways may occur; homologous recombination (HR) and non-homologous end joining (NHEJ) that are active mainly during the S or G2 and G1 phases, respectively.[398-399] However, phosphorylation on S140 in IMR5 after the treatment stayed relatively the same at 1.04 times compared to the control tumor, while that in COG424X decreased by over five times (localized site with Ascore>13). The protein abundance in IMR5 stayed relatively the same, in contrast to COG424X in which its abundance decreased by over 5 times, indicating that there may be a mechanism that interfered with the phosphorylation on this site. Nonetheless, ATM can also phosphorylate histone-lysine N-methyltransferase NSD2 (NSD2) at S102 in response to DNA damage, and the event in COG424X increased by over 3.6 times.[400] Another interesting phosphorylation event is at S51 of DNA ligase I (LIG1), which is critical for the protein to repair DNA.[401] Phosphorylation on this site increased by 2.17 and 1.75 times in IMR5 and COG424X, respectively.

6.3.6 Phosphorylation pleiotropically affected by TOPO/CTX treatment

Other phosphorylation events were also affected since cells need to recover to the cancerous state after the treatment by employing various mechanisms through the regulation of protein phosphorylation. These events can be observed through the changes in phosphopeptide abundance. For IMR5, 256 phosphopeptides from 193 proteins and 351 phosphopeptides from 219 proteins were up- and down-regulated, respectively (Figure 6.6A), corresponding to a total of 4.7% of the quantified unique phosphopeptides. In contrast, 17.4% of all quantified unique phosphopeptides were regulated in the COG424X tumor, including 1,228 phosphopeptides from 700 phosphoproteins and 1,033 phosphopeptides from 453 phosphoproteins were up- and down-regulated, respectively (Figure 6B). Clustering analysis of proteins from the regulated phosphopeptides revealed some similarly enriched biological processes as in the proteome experiment, e.g., cell-cycle processes were enriched from up-regulated phosphopeptides in both tumors. Surprisingly, the processes enriched from the regulated phosphopeptides of COG424X are now reflective of the expected changes from the treatment and comparable to those enriched from the proteome experiment of IMR5. For example, phosphoproteins involved in cytoskeleton organization were greatly up-regulated and more enriched than those from the proteome and phosphoproteome experiments of IMR5 (N=120 proteins, P=1.8E-24 for phosphopeptides of COG424X; N=48 proteins, P=5.4E-16 for phosphopeptides of IMR5; N=74 proteins, P=1.3E-19 for the up-regulated proteins of IMR5), unlike at the proteome level of COG424X where cytoskeleton organization proteins were down-regulated (N=33, P=1.1E-2). The number of proteins participating in the cell-cycle process is also higher at the phosphoproteome level compared with the proteome level, and more of them were enriched in the COG424X tumor than in the IMR5 tumor (N=37 proteins, P=3.4E-7 for

IMR5 and N=123 proteins, P=1.0E-19 for COG424X). For gene expression and nucleic acid metabolism, phosphorylation of proteins involved in these processes was found to be down-regulated in both tumors while only proteins from IMR5 were down-regulated at the proteome level (N=154 proteins, P=1.3E-40 for IMR5 and N=266 proteins, P=5.3E-44 for COG424X). Among those, several phosphorylation events are still reflective of the different regulation of the cell cycle process among the two tumors. For example, CDK1 that was up-regulated at the proteome level may phosphorylate other proteins to regulate the cell cycle process. One of those is the serine/threonine-protein kinase PLK1 (PLK1), which has increased abundance in both tumors. Phosphorylation on site T214 was also increased by over two times in COG424X, which is essential and required for PLK1 mitotic function.[402] CDK1 can also phosphorylate retinoblastoma-associated protein (RB1), and several peptides containing these phosphorylation sites were up-regulated in both tumors.

Benefiting from the large-scale proteomic and phosphoproteomic analyses, we extracted proteins and phosphorylation events of the phosphoinositide 3-kinase (PI3K) signaling pathway from the KEGG database.[403] The pathway is often dysregulated in cancers since the downstream effects include cell survival, proliferation, or cell cycle regulation, and several cancer treatments have targeted proteins this pathway.[404] While protein and phosphoprotein regulations are observed in both tumors, the majority of the changes happened in COG424X (Figure 6.7A and B). Proteins such as AKT serine/threonine kinase 3 (AKT3, displayed as AKT in the figure), GTPase KRas (KRAS, displayed as Ras in the figure), phosphoenolpyruvate carboxykinase, cytosolic [GTP] (PCK1, displayed as PEPCK in the figure), and 5'-AMP-activated protein kinase catalytic subunit alpha-2 (PRKAA2, displayed in AMPK in the figure) increased their abundance

after the treatment. Phosphorylation events such as that on S118 of Bcl2-associated agonist of cell death (BAD) increased in COG424X, which prevents apoptosis, may explain the lower sensitivity of the tumor and could also be an interesting drug target for some neuroblastoma treatment.[405]

Combining the proteomic and phosphoproteomic results, IMR5 is potentially more sensitive to the treatment than COG424X considering that changes in phosphorylation is a faster and immediate response, and these phosphorylation or de-phosphorylation events eventually lead to changes in the protein expression. For IMR5, the responses involving the regulation through the cell-cycle process appeared at the proteome level, in contrast with the changes in COG424X where most of the responses are at the phosphoproteome level. As noted previously, the xenografts were generated from a cell line or a patient, and some characteristics of the cell culture model could allow for a faster or more effective delivery of the drugs even when the treatment was done *in vivo* for both models. Nonetheless, the response could be due to the heterogeneity of the tumors or the different uptake and metabolism pathways.

6.3.7 Motif analysis of up- and down-regulated phosphorylation sites

Kinases normally target the S, T, or Y residues with specific docking motifs surrounding the phosphorylation site. These motifs can be classified into acidic, basic, proline-directed, tyrosine, and other classes based on the decision tree classification described previously.[406] Site-specific analysis of the regulated peptides with a single, localized phosphorylation site showed several enriched motifs. The majority of the sites are proline-directed (SP, Figure 6.6C), which is typically the motif for CDKs, and this

enrichment of this motif could be due to the increased regulation of the cell-cycle process after the treatment. Notably, the abundance of phosphopeptides with the SP motif from IMR5 showed a relatively normal distribution and centered around the zero value (no regulation) while the that from COG424X shifted slightly toward the positive x-axis (Figure 6.6D). The SP motif was the only one enriched in IMR5 while many others were enriched in COG424X. For example, phosphorylation on the R..S.E (“.” refers to any amino acid residue, Figure 6.6C) motif that was up-regulated in COG424X could result from Ca^{2+}/calmodulin-dependent protein kinase II (CAMK2). The kinase is responsible for Ca^{2+} homeostasis and was recently found to regulate cancer progression, cell cycle, apoptosis, and therapy response.[407] Subunits beta, delta, and gamma were quantified in the proteome experiment, and phosphorylation on site T287 increased in all subunits, indicating the kinase activation.[408] The distribution of phosphopeptides with the R..S.E motif shifted more toward the positive x-axis in COG424 compared with IMR5, showing the different regulation of this kinase in the two tumors (Figure 6.6E). This could affect the tumor growth and proliferation, and may be responsible for the lower sensitivity to the treatment of COG424X. Other motifs, such as the S.E motif that could be targeted by CAMK2 or Golgi casein kinase (G-CK), were enriched in both up- and down-regulated phosphopeptides of COG424X (Figure 6.6C). The overall results showed specific kinases/phosphatases that differently controlled the phosphorylation in the tumors and may be responsible for the treatment outcome. These enzymes present a special and interesting type of target for cancer therapy since their inhibition would pleiotropically affect other phosphoprotein targets.

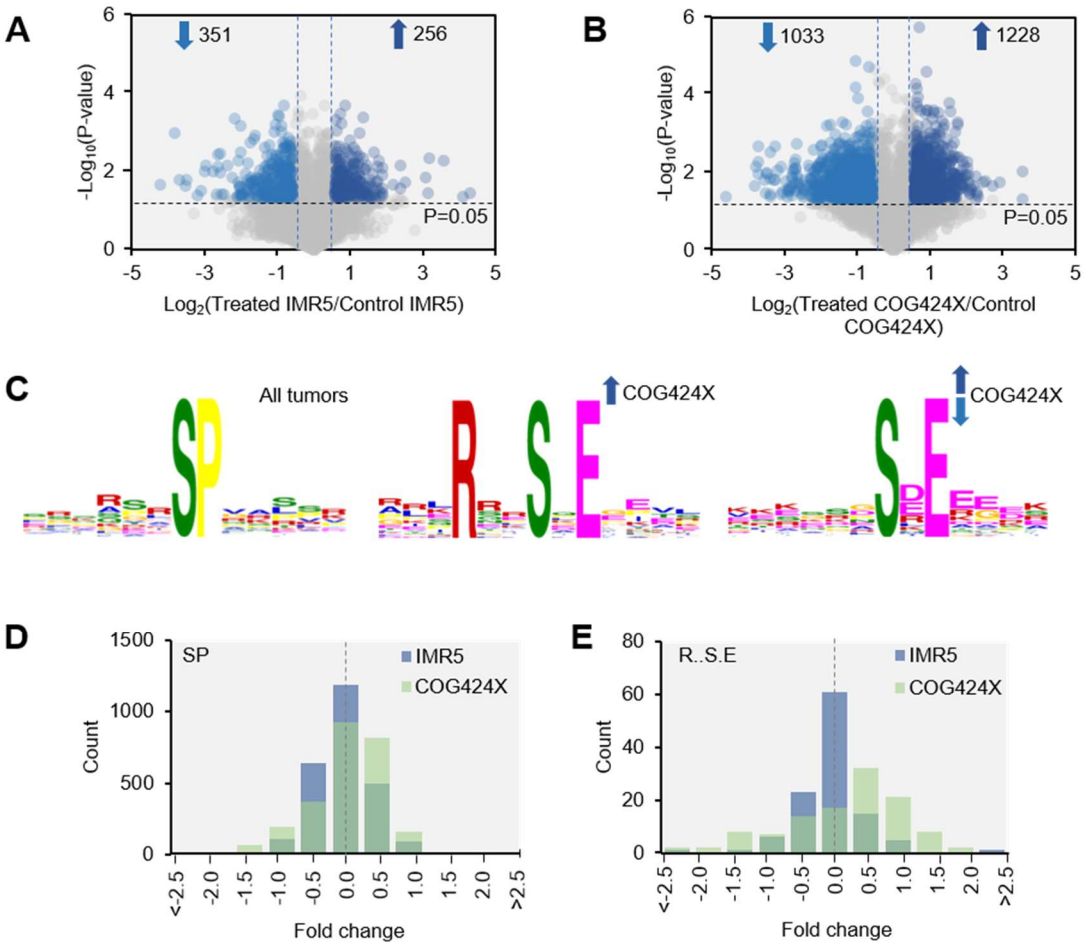

Figure 6.6. (A) and (B) Volcano plots show phosphopeptides that were up- and down-regulated in IMR5 (A) or COG424X (B) tumors. FD means fold difference. (C) Some enriched motifs of up- and down-regulated phosphorylation sites from both tumors. (D) and (E) The abundance distribution of phosphopeptides with motif SP (D) or R..S.E (E).

6.3.8 Common effects of TOPO/CTX on neuroblastoma tumors

Since neuroblastoma tumors generally have heterogeneous characteristics, the identification of proteins and phosphoproteins that are similarly up- or down-regulated in different tumor types could indicate the intrinsic response of neuroblastoma to TOPO/CTX treatment. In the proteome experiment, fifteen proteins were similarly up-regulated, including those that regulate the cell-cycle process such as aurora kinase A (AURKA),

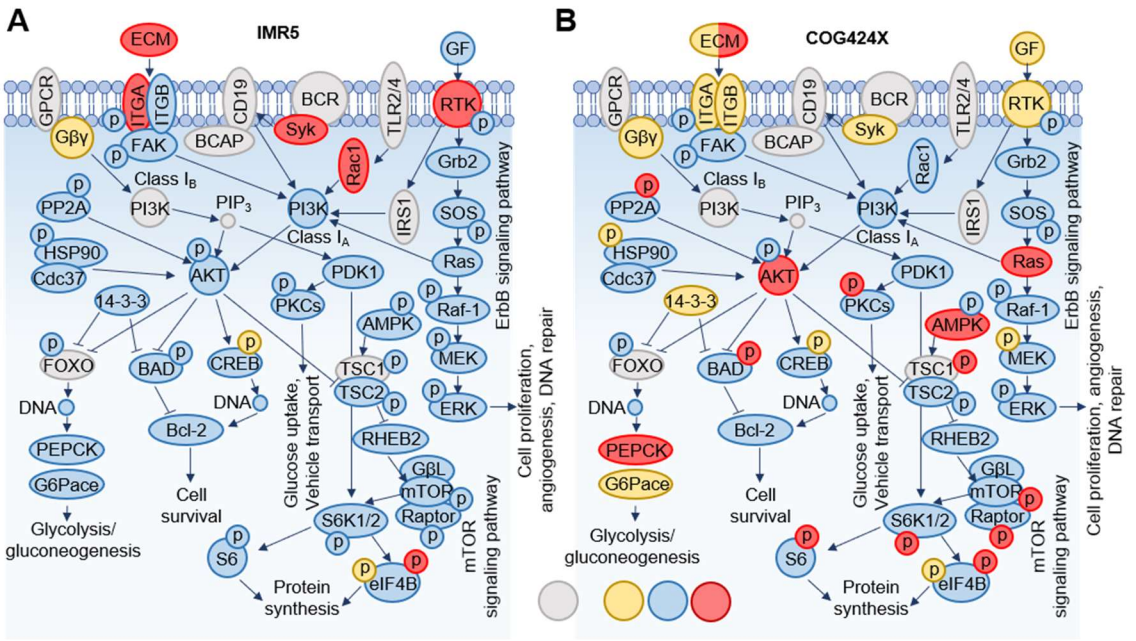

Figure 6.7. The PI3K signaling pathway extracted from KEGG database. Changes in protein abundance and phosphorylation in IMR5 (A) and COG424X (B) are mapped on the pathway and color-coded based on the regulation. Some are shown as a family of proteins and the color-coded up- and down-regulation may not reflect all proteins in that family. For ECM, one protein was up-regulated while the other was down-regulated.

CDK1, and cell division cycle protein 20 homolog (CDC20) (Table 6.1). Eleven proteins were down-regulated in both tumors, which are related to chromatin organization or silencing, such as histone H3.2 (HIST2H3A), high mobility group protein HMG-I/HMG-Y (HMGA1), and high mobility group nucleosome-binding domain-containing protein 3 (HMGN3) (Table 6.1). Some of these proteins are well-known drug targets, such as AURKA that was mentioned in the introduction section, or centromere protein F (CENPF) that has been proposed as a drug target in breast cancer cells. Similar to the proteome experiment, multiple phosphorylation sites were also regulated in both tumors. Considering phosphopeptides with only a single localized site, 16 and 25 phosphorylation sites were up- and down-regulated in both tumors, respectively. Most of them are still involved in the regulation of the cell cycle and gene expression processes. For example, phosphorylation

on S674 of kinesin-like protein KIF18A (KIF18A) that affects the movement of chromosomes in the congression state as cells enter anaphase in the cell cycle was found to increase in both tumors, possibly by CDK1 phosphorylation.[409] Some of these phosphorylation events may be critical for tumor progression and proliferation. Other proteins and phosphoproteins on the list may also serve as potential drug targets for neuroblastoma treatment. Further validation is required to confirm their utilization.

6.4 Conclusions

Treatment by TOPO and CTX has been used in several types of cancers as a single drug treatment or a combination of both. While the toxicity of the drugs to normal cells is unavoidable, their effectiveness in cancer therapy is very distinct. CTX is considered as one of the most potent immunosuppressors available and is currently used in cancer immunotherapy. Nonetheless, the effects of these two drugs on proteins and phosphorylation events have not been systematically studied. In this work, using MS-based proteomics, we globally analyzed proteins and phosphorylation in two neuroblastoma tumor models, i.e., the cell culture-derived xenograft (CDX) and patient-derived xenograft (PDX) models generated in mice and treated with a five-day regimen of TOPO/CTX. Quantification was performed by combining the method with isobaric tagging using the neutron-encoded TMT10plex reagent. Proteomic results demonstrate that there are dramatic differences in the protein expression of the two tumor models, and both also respond to the treatment differently at both the proteome and phosphoproteome levels, which may result from the difference in cells originating from *in vitro* or *in vivo* sources, or the generally heterogeneous nature of neuroblastoma tumors. Depending on the tumor, the cells regulate proteins involved in the cell-cycle process and a response to DNA damage

caused by the treatment, as well as those involved in gene expression and cytoskeleton organization processes. Most changes occurred in at the proteome level for the CDX tumor while those in the PDX tumor occurred at the phosphoproteome level, showing the sensitivity of different tumors to the treatment. Motif analysis showed that some kinases, such as CAMK2, are more active in a specific tumor and may be responsible for the sensitivity difference among the tumors. Nonetheless, we present several proteins and phosphorylation events that were affected similarly in both tumors, as well as kinases that regulates the response at the phosphoproteome level, which may be served as possible drug targets for neuroblastoma treatment and need further validation and verification. The comprehensive analysis of proteins and phosphoproteins provides insights into the response mechanisms of neuroblastoma to chemotherapy and eventually lead to a better treatment.

Table 6.1. Proteins that were up- or down-regulated in both tumors.

UniProt ID	Gene symbol	Annotation	Treated/control	
			IMR5	COG424X
Q92522	H1FX	Histone H1x	0.40	0.48
Q71DI3	HIST2H3A	Histone H3.2	0.40	0.51
P02786	TFRC	Transferrin receptor protein 1	0.56	0.56
Q9P0B6	CCDC167	Coiled-coil domain-containing protein 167	0.57	0.56
P17096	HMGA1	High mobility group protein HMG-I/HMG-Y	0.64	0.57
Q02487	DSC2	Desmocollin-2	0.30	0.58
Q5IJ48	CRB2	Protein crumbs homolog 2	0.64	0.65

Table 6.1. Continued

Q9Y2U8	LEMD3	Inner nuclear membrane protein Man1	0.66	0.66
O95989	NUDT3	Diphosphoinositol polyphosphate phosphohydrolase 1	1.55	1.51
P07996	THBS1	Thrombospondin-1	1.52	1.52
P49454	CENPF	Centromere protein F	2.00	1.53
Q9UJ70	NAGK	N-acetyl-D-glucosamine kinase	1.75	1.57
Q99661	KIF2C	Kinesin-like protein KIF2C	1.96	1.66
Q9H492	MAP1LC3A	Microtubule-associated proteins 1A/1B light chain 3A	1.68	1.68
P68371	TUBB4B	Tubulin beta-4B chain	1.60	1.73
P06493	CDK1	Cyclin-dependent kinase 1	2.38	1.74
O14965	AURKA	Aurora kinase A	2.64	1.78
P07358	C8B	Complement component C8 beta chain	1.53	2.14
J3KRF5	CLTC	Clathrin heavy chain 1	9.38	2.19
Q9P2X3	IMPACT	Protein IMPACT	2.16	2.36

CHAPTER 7. EVALUATION AND OPTIMIZATION OF REDUCTION AND ALKYLATION METHODS TO MAXIMIZE PEPTIDE IDENTIFICATION WITH MS-BASED PROTEOMICS

Adapted with permission of the Royal Society of Chemistry: Suttapitugsakul, S.; Xiao, H.; Smeekens, J.; Wu, R. 2017. Evaluation and optimization of reduction and alkylation methods to maximize peptide identification with MS-based proteomics. Molecular BioSystems 13:2574-2582.

7.1 Introduction

Mass spectrometry (MS)-based proteomics is currently very powerful for protein identification and quantification.[410-414] There are three major approaches, i.e., top-down, middle-down, and bottom-up proteomics.[415-420] For bottom-up proteomics, proteins are digested enzymatically and/or chemically into peptides. The resulting peptides are separated by high-performance liquid chromatography (HPLC), and the eluted peptides are subsequently subjected to tandem mass spectrometric (MS/MS) analysis. Despite the fact that some protein structure information is compromised, bottom-up proteomics currently is still the most popular method because it has been proven to be extremely powerful to identify and quantify proteins,[11, 421-427] including the study of protein post-translational modifications (PTMs),[10, 149, 154, 156, 227, 428-432] and investigation of protein interactions with other proteins and small molecules.[433-436]

Disulfide bonds between sulfhydryl groups of cysteine side chains often regulate protein folding and final functional structures.[437-439] A typical workflow for bottom-up proteomics includes the reduction of disulfide bonds and the alkylation of sulfhydryl groups.[416, 440-441] Without the reduction and alkylation, peptides involved in the disulfide

bonds would be difficult to identify during database searching.[442] Currently, the reduction and alkylation are routine steps for bottom-up proteomics.

Several reducing and alkylating reagents have been commonly used in this field. The following compounds have frequently been reported to serve as reducing reagents: dithiothreitol (DTT), 2-mercaptoethanol (2-ME), tris(2-carboxyethyl)phosphine (TCEP) and tris(3-hydroxypropyl)phosphine (THPP). Sulfur-containing reagents reduce disulfide bonds through the thiol-disulfide exchange (Eq. 1),[443] while those containing phosphorous form phosphine oxide after disulfide reduction (Eq. 2).[444]

$$\text{HS-R-SH} + \text{R'SSR'} \rightarrow \text{HS-R-SSR'} + \text{R'SH} \quad (1)$$

$$R_3P + \text{R'SSR'} + H_2O \rightarrow R_3P{=}O + 2\text{R'SH} \quad (2)$$

Regarding the alkylation, reagents typically alkylate nucleophiles through bimolecular nucleophilic substitution (S_2N) for haloalkanes or Michael addition for maleimide.[445] The following compounds have normally been employed for alkylation: iodoacetamide (IAA), acrylamide (AA), N-ethylmaleimide (N-EM), and 4-vinylpyridine (4-VP).[440, 446-449]

Although the reduction and alkylation are critical to bottom-up proteomics, this process may pose some problems, including the incompleteness of the reactions and side reactions. Besides the sulfhydryl group of cysteine, the alkylation may also happen at other chemical groups, including the amino group of the peptide N-terminus, and those on the side chains of lysine, aspartic acid, glutamic acid, methionine, and histidine.[450-451] Recently, Muller and Winter have reported the systematic evaluation of protein reduction

and alkylation, which has also included side reactions on the side chains of tyrosine, serine, and threonine.[452] Therefore, it is critical to push the desired reaction toward completion and minimize the side reactions. Compared to that paper,[452] the current work has several differences: (1) The experimental designs are different. Here we compared only one parameter with all other parameters fixed for each individual experiment, while in that paper, different combinations of reducing and alkylating reagents were compared. (2) The reducing and alkylating reagents compared are different. For the alkylating reagents, we compared IAA, AA, N-EM, and 4-VP while they evaluated IAA, AA, iodoacetic acid (IAC), and chloroacetamide (CAA). For the reducing reagents, besides DTT, TCEP, and 2-ME in that paper, we also included THPP. (3) Here we systematically investigated the effect of the alkylation reaction conditions (concentration, temperature, and reaction time) on the cysteine alkylation completion and side reactions on other amino acids and the peptide N- and C-termini. (4) We further identified cysteine-containing peptides in yeast cells using the best reducing and alkylating reagents and optimal reaction conditions after the enrichment with Thiopropyl-Sepharose 6B resins.

In this work, we systematically evaluated the reduction and alkylation using the most commonly used reagents, investigated side reactions during the alkylation step, and further optimized the experimental conditions to maximize peptide identifications. The reduction of the disulfide bonds with commonly used DTT, 2-ME, TCEP, and THPP was systematically compared. After the reduction, the alkylation of sulfhydryl groups with the popular reagents of iodoacetamide, acrylamide, N-EM, and 4-VP was also evaluated, and the completion of cysteine alkylation and undesired side reactions were systematically investigated. Alkylating conditions were further optimized for iodoacetamide, and the

optimal conditions were then used for the analyses of the enriched cysteine-containing peptides from yeast.

7.2 Materials and methods

7.2.1 Yeast cell culture, lysis, and digestion

BY4742 MAT alpha yeast (*Saccharomyces cerevisiae*) was cultured in Difco™ YPD broth (BD) overnight at 31 ºC with shaking until the optical density was ~1.0 as measured by UV-Vis spectrometry at 600 nm. Cells were harvested by centrifugation at 800*g*, 4 ºC for 5 minutes, and washed with water. Lysis buffer containing 0.5% sodium deoxycholate (SDC, Sigma-Aldrich), 8 M urea (Sigma-Aldrich), 1 pill/10 mL cOmplete ULTRA Tablets protease inhibitor cocktail (Roche), and 75 mM sodium chloride (Sigma-Aldrich) in 50 mM, pH=8.2 N-(2-hydroxyethyl)piperazine-N'-2-ethanesulfonic acid (HEPES, Biobuffer) buffer, along with 0.5 mm zirconia/silica beads (BioSpec), were added to the XXTuff 2 mL microvials (BioSpec). The cells were lysed using a Mini BeadBeater (BioSpec) at the maximum speed for three cycles, 30 seconds each with a resting period of 2 minutes on ice between cycles. The lysate was then collected.

Chloroform/methanol precipitation of proteins was performed by adding methanol (EMD Millipore), chloroform (EMD Millipore), and water to the lysate in the ratio of 4:1:3, respectively. The protein pellet was collected and air-dried. Digestion buffer containing 5% acetonitrile (ACN, Sigma-Aldrich), 1.6 M urea, and 50 mM, pH=8.2 HEPES was added to the dried pellet so that the concentration of proteins is ~1 mg/ mL. Protein digestion was performed with lysyl endopeptidase (Lys-C, Wako) at 31 ºC overnight and subsequently with sequencing grade modified trypsin (Promega) at 37 ºC for 4 hours (enzyme:substrate

ratio of 1:100 for both enzymes). Trifluoroacetic acid (TFA, Sigma-Aldrich) was added to a final pH of ~2 to quench the digestion. Peptides were desalted using a Sep-Pak Vac tC18 cartridge (Waters). Desalted peptides were aliquoted, dried using a vacuum concentrator, and frozen at -80 ^{o}C until used.

7.2.2 Comparison of reducing reagents

Dried peptides were reconstituted in 50 mM, pH=8.2 HEPES buffer to the concentration of 1 μg/ μL. Four reducing reagents, DTT (Sigma-Aldrich), 2-ME (Sigma-Aldrich), TCEP (Calbiochem), and THPP (Santa Cruz Biotechnology), were compared at the same concentration of 5 mM. The reduction took place at 56 ^{o}C for 25 minutes. Then the mixtures were cooled down to room temperature, and iodoacetamide was added to the final concentration of 14 mM to alkylate the peptides. The alkylation was performed in the dark at room temperature for 30 minutes, followed by addition of the same amount of each reducing reagent to quench the alkylation. The mixtures were left in the dark for another 15 minutes. After the comparison, DTT was chosen as the reducing reagent. Triplicate runs were performed in each experiment. For parallel experiments, except the reducing reagents, the amount of starting materials and any other steps were kept the same.

7.2.3 Comparison of alkylating reagents

Similar to the reduction experiment in the previous section, dried peptides were dissolved in 50 mM, pH=8.2 HEPES buffer to the final concentration of 1 μg/ μL. After the reduction with DTT, the alkylation with iodoacetamide, acrylamide, N-EM, or 4-VP (all from Sigma-Aldrich) was performed and compared at the same concentration of 14 mM in the dark at room temperature for 30 minutes. Then DTT was added to the final

concentration of 5 mM to quench the alkylation. Based on the current experimental results, iodoacetamide was selected as the alkylating reagent for further experiments. The following parameters were optimized: concentration (1, 2, 4, 8, 14, and 20 mM), temperature (room temperature, 40, 70, and 85 ^{o}C), and reaction time (10, 20, and 30 minutes). Similarly, triplicate runs were performed in each experiment.

7.2.4 Enrichment of peptides with cysteine from yeast whole-cell lysate

The protocol for the enrichment of peptides containing cysteine using Thiopropyl-Sepharose 6B resin was modified from the method described by Guo et al.[453] In summary, 300 µg of dried peptides from a yeast whole-cell lysate in 100 µL 50 mM, pH=8.2 HEPES buffer was reduced with 5 mM DTT for 25 minutes at 56 ^{o}C. The mixture was cooled down and incubated with rehydrated resin. Unbound or non-specifically bound peptides were washed according to the original protocol. Enriched peptides were eluted by incubation with DTT and alkylated with 14 mM iodoacetamide at room temperature in the dark for 30 minutes. The samples were quenched with 5 mM DTT for another 15 minutes in the dark at room temperature prior to drying in the vacuum concentrator.

7.2.5 LC-MS/MS analysis

Eluted peptides were purified, and dried peptides were subsequently dissolved in a solution with 5% ACN and 4% formic acid (FA) prior to LC-MS/MS analysis. Peptides were separated by a Dionex UltiMate 3000 UHPLC system (Thermo Scientific), which is coupled to an LTQ Orbitrap Elite Hybrid Mass Spectrometer (Thermo Scientific) with Xcalibur software (version 3.0.63). A total of ~1 µg of peptides was loaded onto a C18-packed microcapillary column (Magic C18AQ, 3 µm, 200 Å, 100 µm x 16 cm, Michrom

Bioresources) by a Dionex WPS-3000TPL RS autosampler (Thermostatted Pulled Loop Rapid Separation Nano/Capillary Autosampler). Peptides were separated by reversed-phase chromatography using an UltiMate 3000 binary pump with a 90-minute gradient of 4-30% ACN containing 0.125% FA. MS/MS analysis was performed with a data-dependent Top20 method. For each cycle, a full MS scan (resolution: 60,000) in the Orbitrap with 1 million automatic gain control (AGC) target was followed by up to 20 MS/MS in the LTQ for the most intense ions.[454-455] Selected ions were excluded from further sequencing for 90 seconds. Ions with single or unassigned charge were not sequenced. Maximum ion accumulation times were 1,000 ms for each full MS scan and 50 ms for MS/MS scans.

7.2.6 Database search

Raw files from MS/MS analysis were converted to mzXML files and searched by SEQUEST algorithm (version 28).[217] The following parameters were used: 10 ppm precursor mass tolerance, 0.5 Da fragment ion mass tolerance, fully tryptic digestion, up to two missed cleavages, methionine oxidation (+15.9949 Da), alkylation at cysteine, histidine, lysine, aspartic acid, glutamic acid, tyrosine, the peptide N- and C-termini (+57.0215 Da for iodoacetamide, +71.0371 Da for acrylamide, +126.0555 Da for N-EM, and +105.0579 Da for 4-VP). The target-decoy method was employed to estimate the false discovery rate (FDR).[218, 456] Linear discriminant analysis (LDA) was used to filter and control peptide identifications by parameters such as XCorr, ΔC_n, and precursor mass error.[457] The minimum-length peptides contain at least six amino acid residues. Peptide spectral matches were filtered to less than 1% FDR.

7.3 Results and discussion

7.3.1 Comparison of reducing reagents

In this study, Baker's yeast (*Saccharomyces cerevisiae*) was chosen as a model, and the peptides used were from its whole-cell lysate. The typical bottom-up approach includes the reduction of disulfide bonds, followed by the alkylation of sulfhydryl groups prior to chromatographic separation and MS analysis. For the comparison of the reduction, the amount of the starting peptides and all other steps were the same except using different reducing reagents, and the same is for the alkylation comparison (Figure 7.1A and Figure 7.1B).

In the first experiment, four commonly used reducing reagents were compared, i.e., DTT, 2-ME, TCEP, and THPP. Each of them was used to reduce peptides at the same concentration of 5 mM at 56 ^{o}C for 25 minutes. A control group without the reduction and alkylation was included (Figure 7.1). Each experiment was run in triplicate, and the number of identified peptides and proteins are the average ones with the standard deviation shown in Figure 7.2. There were no obvious differences among four reducing reagents, which clearly demonstrated that all these reducing reagents are similarly effective to reduce disulfide bonds. Interestingly, without the reduction and alkylation, we obtained very similar results. Theoretically, the reduction and alkylation increase the identification of cysteine-containing peptides, albeit only a small portion of peptides contain cysteine. In addition, some peptides may be lost due to the side reactions of the alkylation, as discussed below.

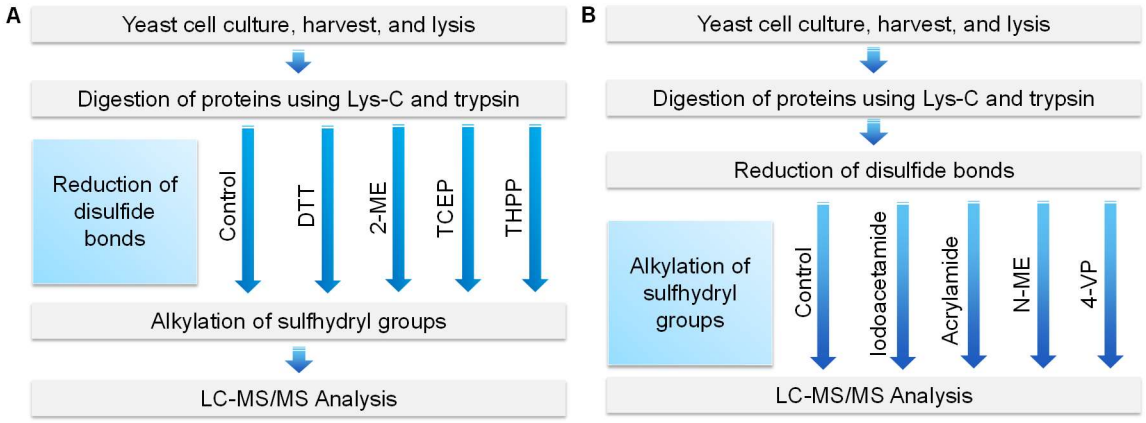

Figure 7.1. Experimental procedure to compare different reducing and alkylating reagents. (A) Comparison of the reduction with DTT, 2-ME, TCEP, or THPP. (B) Comparison of the alkylation with iodoacetamide, acrylamide, N-EM, or 4-VP. The reduction and alkylation were performed at the peptide level.

Each reducing reagent has its advantages and disadvantages. TCEP is not stable in phosphate buffer under physiological pH,[444] but it is more stable than DTT when there is no metal chelator. However, a metal chelator, such as ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA), can increase the stability of DTT but decrease TCEP stability.[458] Both DTT and 2-ME have shown to decrease their stability at higher pH and temperature.[459] THPP is more stable than both TCEP and DTT at room temperature and pH=8.[460] The current results demonstrated that all reducing reagents had similar performance. DTT was chosen for further experiments because of its popularity. It should be noted that the number of proteins or peptides may be more fairly compared within the same experiment, as there are some variations among different batch of samples, including variations from the LC column and MS conditions.

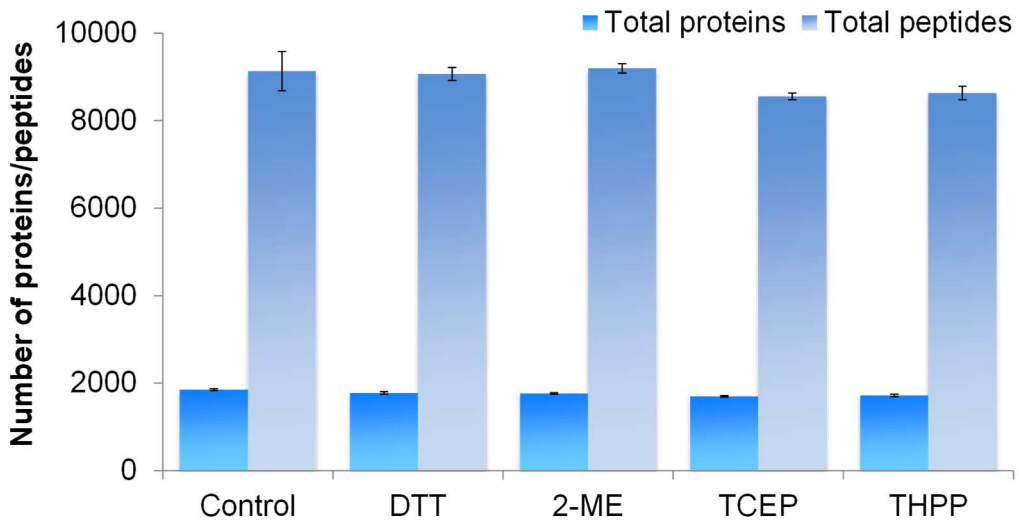

Figure 7.2. Comparison of the reducing reagents. The number of proteins and total peptides identified from each reducing reagent were compared.

7.3.2 Side reactions during the alkylation reaction

After the reduction, the alkylation follows to stabilize free sulfhydryl groups. Several alkylating reagents have been commonly used, including iodoacetamide, acrylamide, N-EM, and 4-VP. Figure 7.3A shows the desired alkylation at the side chain of cysteine by iodoacetamide. Three other possible side reactions at the peptide N-terminus and the side chains of lysine and aspartic acid are displayed in Figure 7.3B. Side reactions may occur during the alkylation reactions with different reagents. There have been reports regarding the alkylation at the side chains of other amino acids besides the desired alkylation of cysteine, including aspartic acid, glutamic acid, lysine, histidine, tyrosine, serine, and threonine.[450-452, 461] Even di- and tri-alkylation on the side chains of some amino acid residues of peptides may further occur,[462] but it is less common. In this study, we mainly focused on mono-alkylation at cysteine, histidine, lysine, aspartic acid, glutamic acid, tyrosine, and the peptide N- and C-termini.

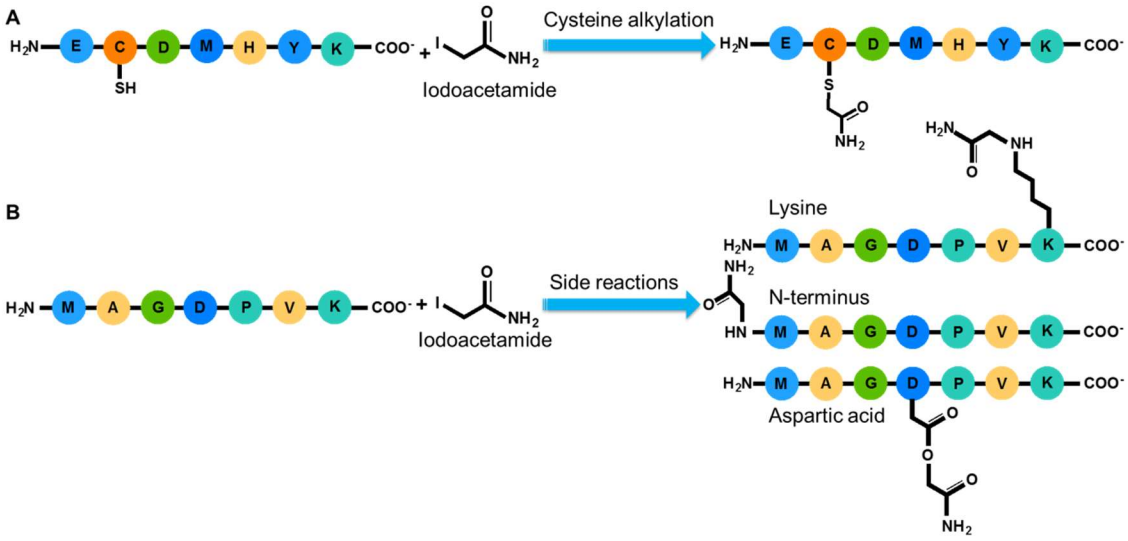

Figure 7.3. Possible alkylation reactions with iodoacetamide. Ideal alkylation is on the sulfhydryl group of cysteine (A). Side reactions may occur, and the alkylation reactions at the peptide N-terminus and the side chains of lysine and aspartic acid are shown here as examples (B).

Examples of MS/MS spectra of peptides with side reactions, i.e., carbamidomethylated N-terminus and lysine, are in Figure 7.4A and Figure 7.4B, respectively. Peptide G]LVSDPAGSDALNVLK ("]" denotes the alkylation of the amino group of the peptide N-terminus) is highly confidently identified with XCorr of 4.30 and a mass accuracy of -0.48 ppm (Figure 7.4A). This peptide is from protein YML126C, which is a hydroxymethylglutaryl-CoA synthase protein. The peptide does not contain any cysteine, and the peptide N-terminus was undesirably alkylated.

Another example is peptide LQETNPEEVPK#FEK ("#" refers to the alkylation at the side chains of other amino acids except cysteine) (Figure 7.4B). Similarly, this peptide from protein YKL056C, a translationally-controlled tumor protein homolog, was also confidently identified with XCorr of 4.43 and a mass accuracy of -0.21 ppm. There is no cysteine in the identified peptide, and undesired alkylation occurred at the side chain of lysine.

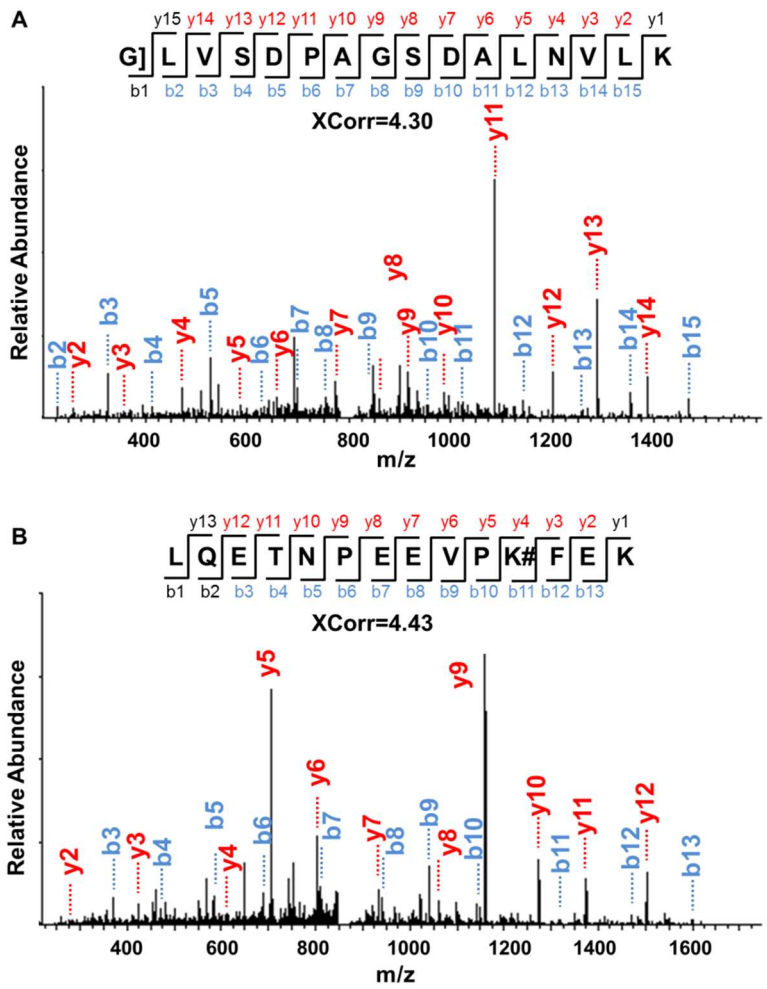

Figure 7.4. Examples of MS/MS spectra from peptides with (A) the alkylated N-terminus, as shown by the "]" sign and (B) alkylated lysine, as shown by the "#" sign.

7.3.3 Comparison of alkylation with different reagents

Compared to the reduction, the alkylation of free sulfhydryl groups of cysteine is more complex. Besides the potential incompletion of the reaction, side reactions from reactive alkylating reagents can also occur. Here we compared four commonly used alkylating reagents: iodoacetamide, acrylamide, N-EM, and 4-VP. Peptide samples were reduced with 5 mM DTT for 25 minutes at 56 ºC before alkylated with each reagent at the

same concentration of 14 mM for 30 minutes in the dark at room temperature. A control group without the reduction or alkylation was included as well.

The number of proteins detected was similar, i.e., 1,700-1,800 proteins in each experiment, except for the results with N-EM (1,447 ± 153 proteins) (Figure 7.5A). Similarly, the number of total peptides was about 8,200 except for the samples treated with N-EM (6,672 ± 589 peptides) (Figure 7.5B), as further discussed below.

7.3.4 Side reactions of the alkylation from different reagents

With reactive alkylating reagents, side reactions frequently happen, especially in complex biological samples. These four alkylating reagents resulted in different degrees of alkylation at other functional groups, especially the amino group at the peptide N-terminus and the side chain of lysine. Among the four alkylating reagents, N-EM also resulted in the greatest number of peptides with the alkylated N-termini (791 ± 73), which is even greater than the alkylation on cysteine. There are 133 ± 9 peptides with the alkylated N-termini from acrylamide, 92 ± 8 peptides from iodoacetamide, and 73 ± 8 peptides from 4-VP (Figure 7.5B).

For the side reaction on the side chain of lysine, the alkylation with N-EM resulted in the highest number of peptides with alkylated lysine among four alkylating reagents. These results are highly consistent with previous data, i.e., the least number of total peptides was identified using N-EM as the alkylating reagent because more peptides were lost due to the side reactions than those rescued from the cysteine alkylation. Compared to 4-VP, N-EM was more effective to alkylate cysteine, but the side reactions were much severer.

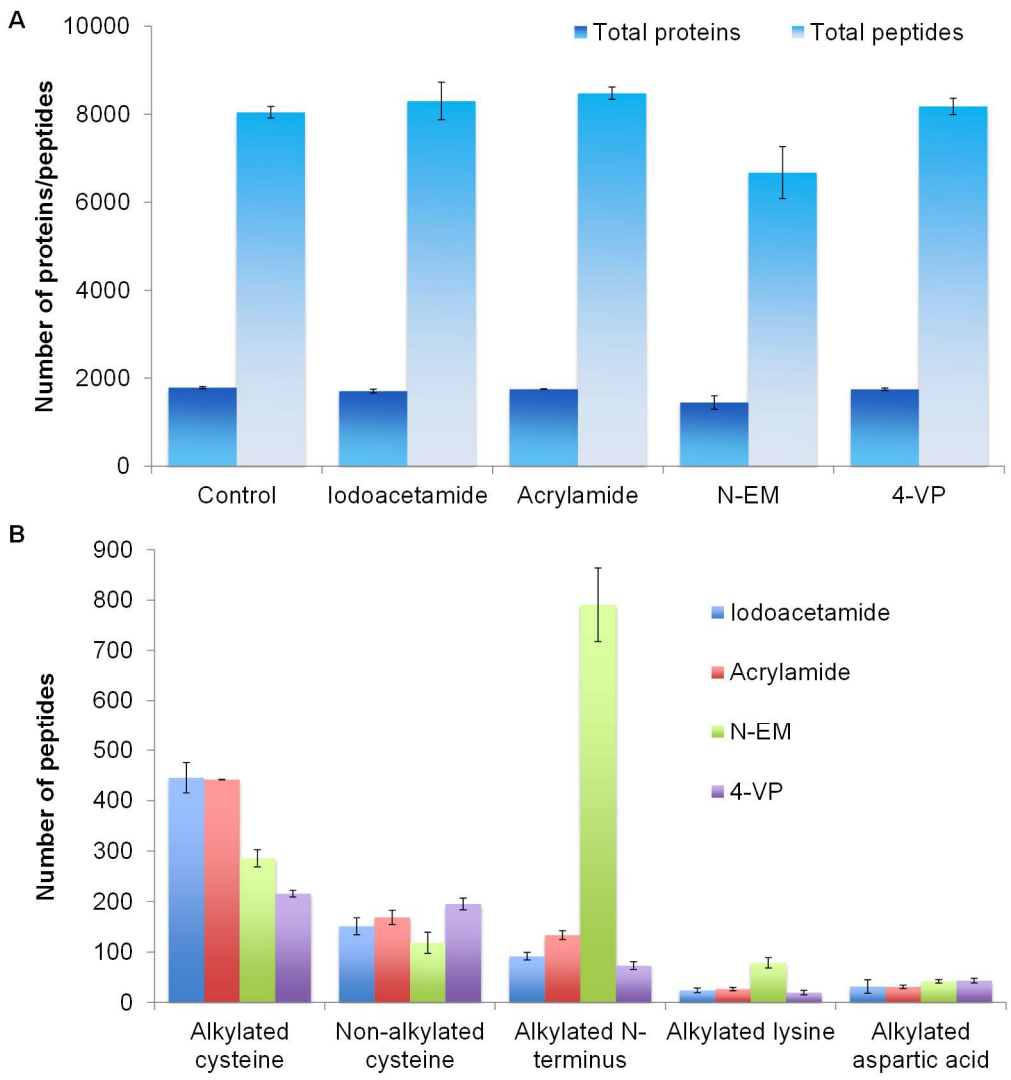

Figure 7.5. Comparison of the alkylating reagents. (A) Effects of different alkylating reagents on the identification of proteins and peptides. (B) Comparison of identified peptides with alkylated cysteine, free cysteine (due to incomplete reaction), or the side reactions on the peptide N-terminus and the side chains of lysine and aspartic acid.

Side reactions at other amino acid residues were also studied, as shown in Figure 7.6. Based on the current results, the extent of side reactions was the N-terminus > glutamic acid > the C-terminus ≈ lysine > aspartic acid > tyrosine > histidine. As mentioned previously, the alkylation using iodoacetamide could occur through bimolecular nucleophilic substitution, where the nucleophilic sulfhydryl group attacks the C2 of

iodoacetamide with iodine leaving the molecule at the same time.[463] The electron-rich groups of the side chains of other amino acids also carry out the nucleophilic substitution reactions, as shown in this work.

Among the parallel experiments, 4-VP resulted in the least side reactions, but the completion rate of the cysteine alkylation was the lowest. These results suggest that 4-VP is relatively less reactive, and it is not a good choice as the alkylating reagent. N-EM has the highest level of side reactions, especially with the amino group at the peptide N-terminus and the side chain of lysine. After evaluating all these results with different alkylating reagents, iodoacetamide is still the best choice because it provided the highest completion rate of the cysteine alkylation and relatively fewer side reactions. Therefore, iodoacetamide was chosen for further experiments. Acrylamide could also be used as an alternative since similar results were obtained.

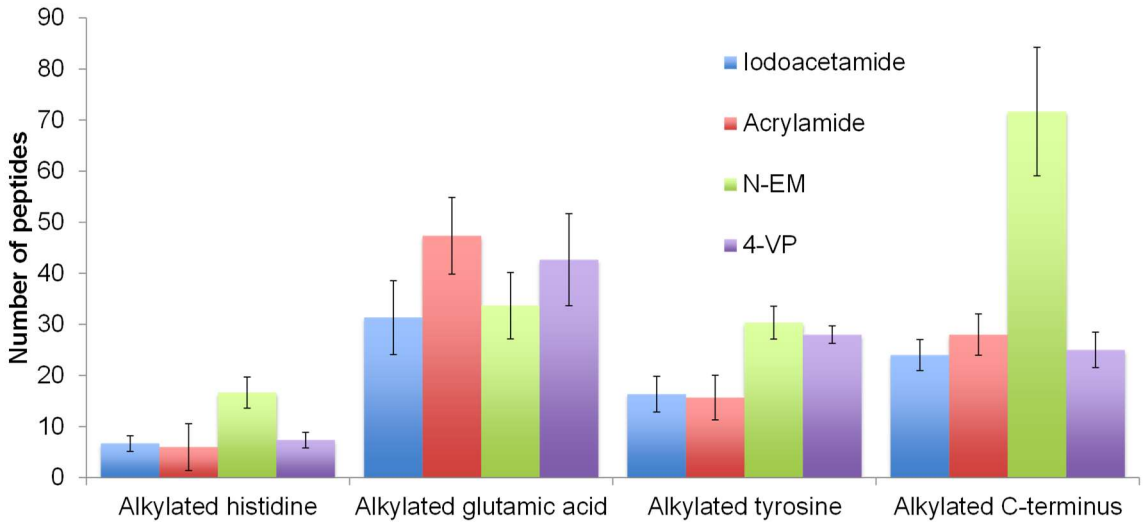

Figure 7.6. Comparison of alkylating reagents: other side reactions on the side chains of histidine, glutamic acid, tyrosine, and the peptide C-termini.

7.3.5 Optimization of the alkylation with iodoacetamide

As shown above, even for the best alkylating reagent, the reaction completion and side reactions still pose problems. In order to push the reaction towards completion and to minimize side reactions, we further optimized several reaction conditions for iodoacetamide, including concentration, temperature, and reaction time. Peptides were first reduced with 5 mM DTT in 50 mM, pH=8.2 HEPES buffer, and then the samples were alkylated with iodoacetamide under different conditions.

The iodoacetamide concentrations of 1, 2, 4, 8, 14, and 20 mM were examined. Overall, the number of proteins and peptides identified were similar in all cases (Figure 7.7A). However, more peptides with alkylated cysteine were identified as the concentration increased. The number of peptides with alkylated cysteine was the highest at the concentration of 14 mM (446 $\pm$ 13 peptides) and started to level off at 20 mM, while at the lowest iodoacetamide concentration tested here, 1 mM, 217 $\pm$ 10 peptides were detected (Figure 7.7B). However, at 14 mM iodoacetamide, 144 $\pm$ 11 peptides were still not alkylated. Alkylation at the N-terminus is still the major side reaction, and the number of peptides with side reactions slightly increased as the concentration went up (Figure 7.7B and Figure 7.8A).

The temperature effect on the alkylation was also investigated, and the following temperatures were tested: room temperature, 40 ^{o}C, 70 ^{o}C, and 85 ^{o}C. As the temperature became higher, the number of detected proteins and peptides slightly decreased. The highest number of proteins identified were at room temperature (1,631 $\pm$ 33 proteins) and 40^{o}C (1,655 $\pm$ 9) and the number of peptides were 7,982 $\pm$ 183 and 8,132 $\pm$ 72,

respectively. At 85°C, only 1,157 ± 32 proteins and 6,178 ± 315 peptides were identified (Figure 7.7C). Temperature affects the alkylation at the peptide N-terminus the most, as shown in Figure 7.7D. The number of peptides with the alkylated N-terminus dramatically increased from 87 ± 4 at room temperature to 1,065 ± 55 at 85°C (Figure 7.7D). Alkylation at the side chains of lysine, glutamic acid, aspartic acid, and histidine also increased slightly with the temperature (Figure 7.8B). This resulted in lower number of peptides with alkylated cysteine as the temperature elevated.

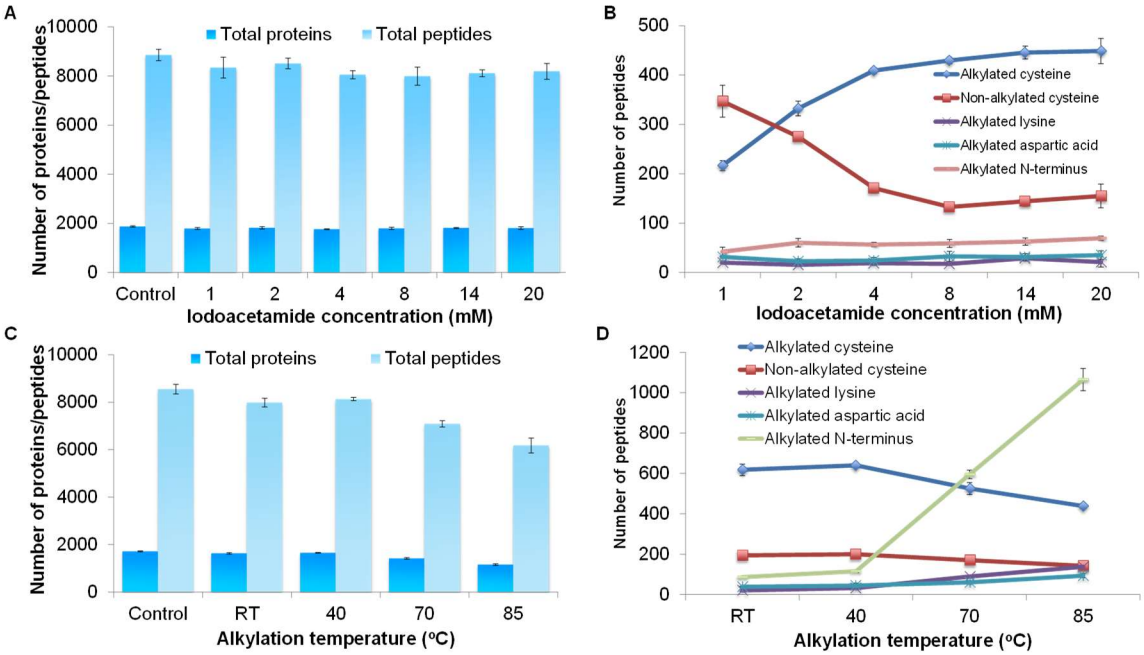

Figure 7.7. Optimization of alkylation conditions. (A) Effects of the iodoacetamide concentration on the identification of proteins and peptides. (B) The number of identified peptides with alkylated cysteine, free cysteine (due to incomplete reaction), or the side reactions on the peptide N-terminus and the side chains of lysine and aspartic acid as a function of the iodoacetamide concentration. (C) Effects of the alkylation temperature on the identification of proteins and peptides. (D) The number of identified peptides with alkylated cysteine, free cysteine, or the side reactions on the peptide N-terminus and the side chains of lysine and aspartic acid at different alkylation temperatures.

The effect of the alkylation reaction time on the identification of peptides and proteins was tested as well. Peptide samples were incubated with iodoacetamide for 10, 20,

and 30 minutes, followed by quenching with DTT as mentioned above. Alkylation time did not markedly affect the degrees of the cysteine alkylation with iodoacetamide because similar number of peptides with alkylated cysteine were identified. Similarly, degrees of side reactions barely changed for the different reaction durations tested here (Figure 7.9).

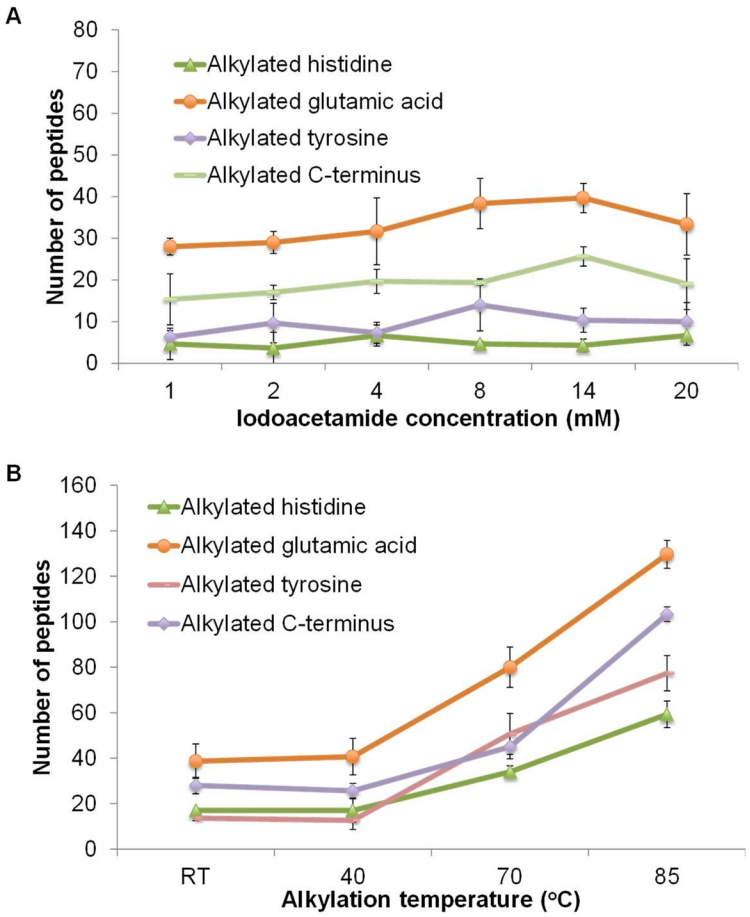

Figure 7.8. Other side reactions from the alkylation with iodoacetamide at different concentrations (A) and temperatures (B).

7.3.6 Enrichment of peptides with cysteine

Next, we sought to identify peptides with cysteine from the yeast whole-cell lysate. To maximize the number of cysteine-containing peptides, enrichment of peptides containing cysteine from the lysate is critical. In addition, we need to effectively reduce

disulfide bonds and alkylate sulfhydryl groups to boost the identification of these peptides. The enrichment was performed by incubating peptides with Sepharose resin containing sulfhydryl groups.

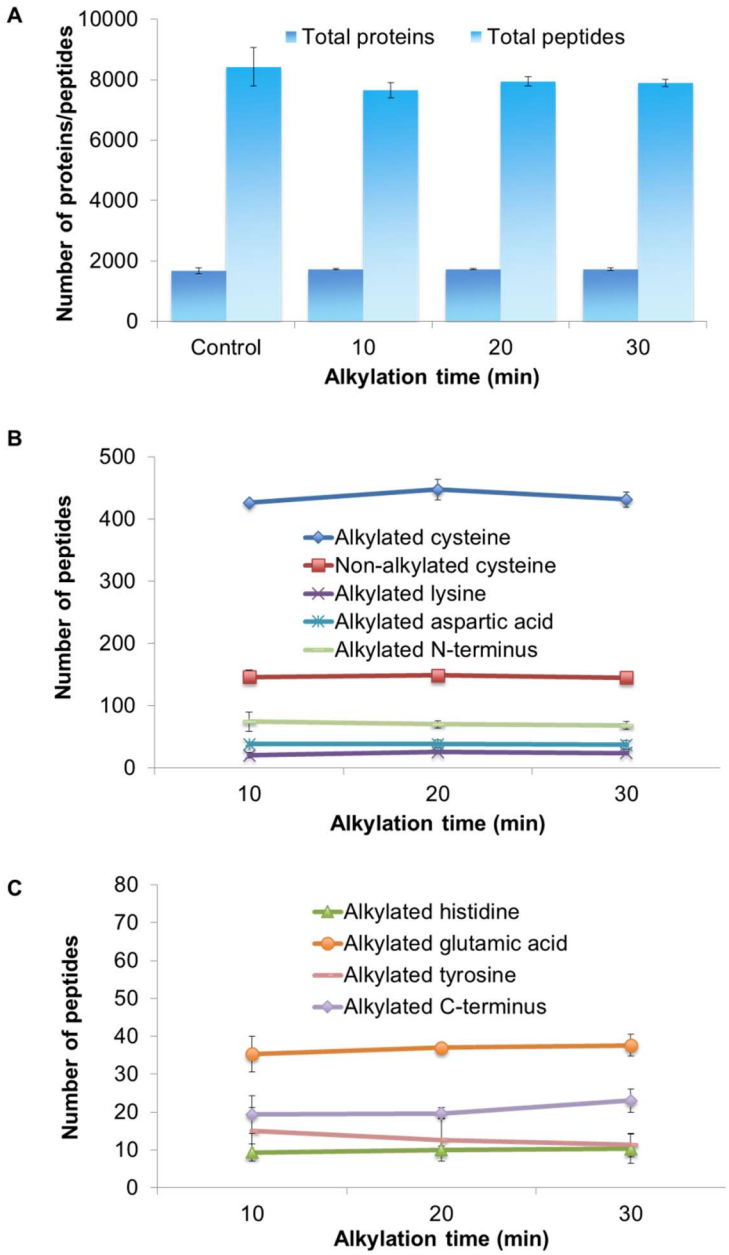

Figure 7.9. Optimization of alkylation conditions. (A) Effects of alkylation time on the identification of proteins and peptides. (B) The number of identified peptides with alkylated cysteine, free cysteine (due to incomplete reaction), or the side reactions on the N-terminus and the side chains of lysine and aspartic acid from different alkylation times. (C) Other side reactions from the alkylation with iodoacetamide at different reaction times.

Figure 7.10 shows the enrichment and alkylation of peptides containing cysteine. A commercially available Thiopropyl-Sepharose 6B was employed for the enrichment. The resin contains 2-thiopyridyl disulfide groups attached to Sepharose through an ester linkage. The lyophilized resin was first rehydrated in water and subsequently incubated with peptides that were treated with DTT to reduce disulfide bonds. Several washing steps were performed to remove non-specifically bound peptides. Enriched peptides were eluted with 20 mM DTT solution in 25 mM NH_4HCO_3 buffer before alkylating with 14 mM iodoacetamide. During the incubation, active sulfhydryl groups on the resin should form disulfide bonds with cysteine on peptides. This resulted in effective and selective alkylation of peptides containing cysteine. A control group without the enrichment nor the reduction/alkylation and a group without the enrichment, but with the reduction/alkylation were also compared.

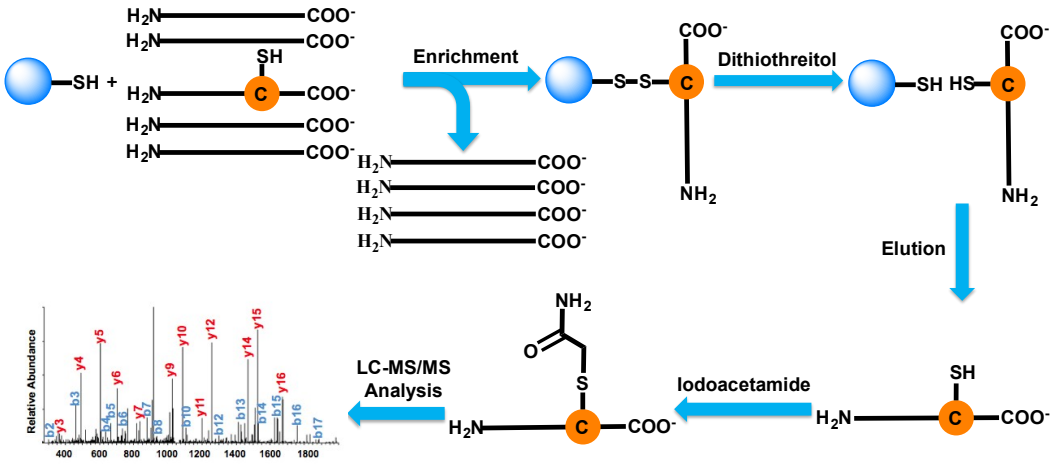

Figure 7.10. Enrichment of peptides containing cysteine using Thiopropyl Sepharose 6B resin. Peptides from the yeast whole-cell lysate are incubated with the resin. Peptides containing cysteine are bound to the resin through disulfide bonds while unbound peptides are removed. Enriched peptides are cleaved with DTT, and then alkylated with iodoacetamide prior to LC-MS/MS analysis.

The overlap of cysteine-containing peptides identified from the three experiments is in Figure 7.11, and only peptides with unique sequence were compared. Without the reduction and alkylation, 288 peptides, corresponding to 216 proteins, containing cysteine were identified (Figure 7.11 and Figure 7.12). With the reduction and alkylation, but not the enrichment, 831 peptides with alkylated and free cysteine, corresponding to 504 proteins, were detected. However, 2,730 peptides with alkylated and free cysteine and 1,398 proteins were identified with the enrichment (Figure 7.11). Cysteine-containing peptides identified from the enrichment were about nine times of those without the enrichment and reduction/alkylation. The cysteine-containing peptides identified with the enrichment covers 81.9% of those detected from the experiment without the enrichment, but with the reduction/alkylation. The enrichment dramatically enhanced the identification of peptides and proteins containing cysteine.

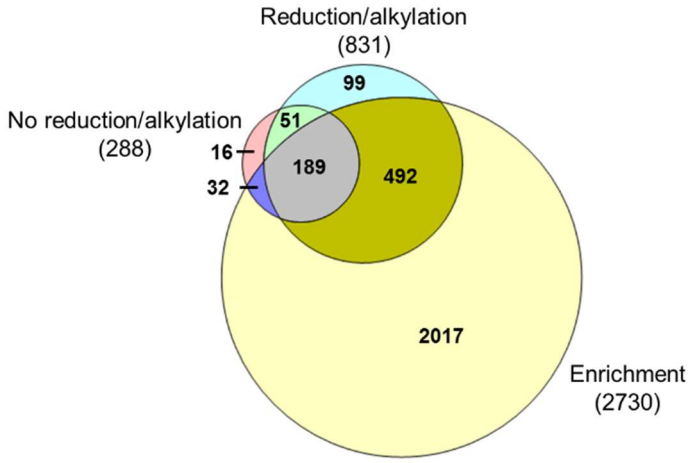

Figure 7.11. Comparison of peptides containing cysteine identified in yeast cells with or without the enrichment. Overlap of unique peptides containing cysteine (both alkylated and non-alkylated) identified from the experiments without enrichment nor reduction/alkylation (No reduction/alkylation), the experiments with only reduction/alkylation (Reduction/alkylation), and the experiments with both the enrichment and reduction/alkylation (Enrichment).

Proteins with cysteine from the enrichment experiment were subjected to clustering analysis using The Database for Annotation, Visualization and Integrated Discovery 6.8 (DAVID).[464] Based on the biological process, proteins with the organic acid metabolic process were highly enriched with a very low *P*-value of 2.7×10^{-32} (Figure 7.12A). Other enriched processes include the organic acid biosynthesis process ($P=9.4 \times 10^{-17}$), nuclear transport ($P=7.2 \times 10^{-8}$), carbohydrate biosynthesis process ($P=8.6 \times 10^{-5}$), and cellular response to oxidative stress ($P=1.3 \times 10^{-5}$), which correspond excellently with the well-known function of cysteine that plays a critical role in oxidation-reduction processes.

7.4 Conclusions

Bottom-up proteomics has proven to be very powerful for protein analysis. The reduction and alkylation are routine steps during the sample preparation prior to MS analysis. In this work, we systematically evaluated the commonly used reducing and alkylating reagents. Four most commonly used reducing reagents were examined here and they had very comparable performances while the results from four alkylating reagents are notably different. In addition to the completion rate differences among alkylation reagents, side reactions also varied at several side chains of the amino acid residues, including histidine, lysine, aspartic acid, glutamic acid, tyrosine, and the peptide N- and C-termini. The extents of the alkylation reaction with iodoacetamide on cysteine and undesired side reactions were also affected by its concentration and the alkylation temperature. Based on the current results, either of the tested reducing reagents (DTT, 2-ME, TCEP, or THPP) can be used to perform the reduction reaction. Regarding the alkylation, iodoacetamide provided us the best results considering the reaction completion rate and side reactions. The optimal alkylation reaction conditions included 14 mM iodoacetamide, room

temperature, and the reaction time of 30 minutes. The similar results were obtained using acrylamide, which may be used as an alternative alkylation reagent. The optimal alkylation conditions were employed to maximize the coverage of cysteine-containing peptides identified from the yeast whole-cell lysate. The enrichment markedly enhanced the identification of cysteine-containing peptides. The current results provide valuable information for choosing right reagents and optimal experimental conditions to maximize the identification of peptides, especially cysteine-containing peptides.

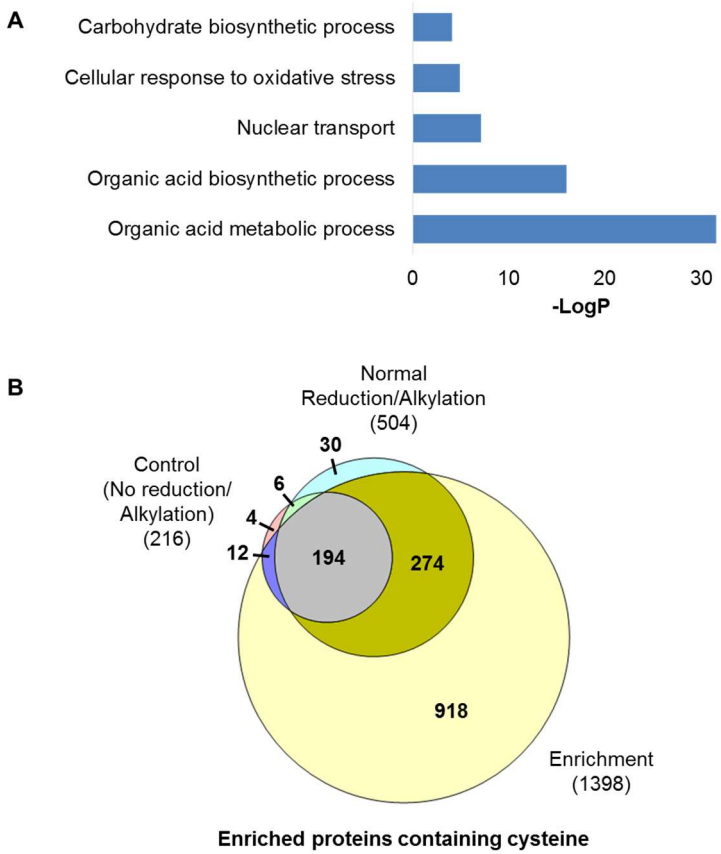

Figure 7.12. Comparison results of the enrichment at protein level. (A) Clustering of proteins with cysteine from the enrichment experiment according to the biological process. (B) Overlap of identified proteins containing cysteine from the experiments without the enrichment nor reduction/alkylation (No reduction/alkylation), the experiments with only the reduction/alkylation (Reduction/alkylation), and the experiments with both the enrichment and reduction/alkylation (Enrichment).

CHAPTER 8. CONCLUSIONS AND FUTURE DIRECTIONS

8.1 Conclusions

8.1.1 Surface glycoproteomic analysis reveals that both unique and differential expression of surface glycoproteins determine the cell type

Proteins on the cell surface are frequently glycosylated and they are essential for cells. Surface glycoproteins regulate nearly every extracellular event, but compared to proteins, comprehensive and site-specific analysis of surface glycoproteins is much more challenging and dramatically understudied. Here, combining metabolic labeling, click-chemistry and enzymatic reactions, and mass spectrometry-based proteomics, we characterized surface glycoproteins from eight popular types of human cells. This integrative and effective method allowed for the identification of 2,172 N-glycosylation sites and 1,047 surface glycoproteins. The distribution and occurrence of N-glycosylation sites were systematically investigated, and protein secondary structures were found to have dramatic influence on glycosylation sites. As expected, most sites are located on disordered regions. For the sites with the motif N-!P-C, about one third of them are located on helix structures while those with the motif N-!P-S/T prefer strand structures. There is almost no correlation between the number of glycosylation sites and protein length, but the number of sites corresponds well with the frequencies of the motif. Quantification results reveal that besides cell-specific glycoproteins, the uniqueness of each cell type further arises from differential expression of surface glycoproteins. The current research indicates that multiple surface glycoproteins including their abundances need to be considered for cell classification rather than a single cluster of differentiation (CD) protein normally used in

conventional methods. These results provide valuable information to the glycoscience and biomedical communities and aid in the discovery of surface glycoproteins as disease biomarkers and drug targets.

8.1.2 Temporal analysis of surface glycoproteins reveals distinct responses of monocytes and macrophages to infection

Glycoproteins on the surface of immune cells play extremely important roles in response to pathogens, including sensing their presence, initiating signal transduction, receiving signals from other immune cells, and facilitating with cell migration to the infection site. Nonetheless, a systematic and time-resolved investigation of surface glycoproteins in immune response processes remains to be explored. Combining selective enrichment of surface glycoproteins with multiplexed proteomics, we globally and site-specifically quantitated the dynamics of surface glycoproteins on THP-1 monocytes and macrophages by using lipopolysaccharides (LPS) as an immune modulator and during the monocyte-to-macrophage differentiation. The time-resolved analysis reveals transient changes and differential remodeling of surface glycoproteins on both cell types, as well as the priming of cells during the differentiation that contributed to the different responses to LPS in monocytes and macrophages. Besides the well-documented glycoproteins, we identified novel surface glycoproteins participating in the immune response such as APMAP, TSPAN3, and IGSF8. Potential upstream regulators and downstream effects of the regulated glycoproteins were also determined. Additionally, site-specific glycosylation changes were identified in response to the infection, for instance, glycosylation at the site N229, but not N195 and N223 of CSF2RA after the LPS challenge. The study provides unprecedented and valuable information about surface glycoprotein changes in response to

the infection, facilitating a better understanding of glycoprotein functions and cellular immune responses.

8.1.3 Quantification of the proteome and surface glycoproteome dynamics during the partial epithelial-to-mesenchymal transition (EMT) state

Epithelial-to-mesenchymal transition (EMT) is integral for biological processes including embryonic development, wound healing, and cancer metastasis. While several inducers have been reported, the most notorious one is transforming growth factor beta (TGF-β). TGF-β initiates the transition through surface TGF-β type I and type II receptors that subsequently activate the SMAD pathway. This results in the regulation of transcription factors such as SNAIL and ZEB, and the transformation of epithelial to mesenchymal cells. Recent studies have shown that the partial EMT state is metastable and can be observed. In fact, the full mesenchymal state might not occur *in vivo* due to the requirement that cells must be able to propagate again. In this work, we study the EMT process in A549 cells treated with TGF-β in a time-resolved manner. The proteomics analysis revealed proteins that are involved in the epithelial, mesenchymal, as well as partial states. Comparing with transcriptomics results, several novel partial EMT drivers were discovered, including JUNB that was found to be the regulator of the partial EMT state. Surface glycoproteomics results showed several surface glycoproteins that are involved in the transition, but the roles of sialic acid-modified surface glycoproteins still needs to be further investigated. The results will lead to a better understanding of the EMT process and their use as markers for cellular development and cancer metastasis.

8.1.4 Enhancing comprehensive analysis of secreted glycoproteins from cultured cells without serum starvation

Glycoproteins secreted by cells play essential roles in the regulation of extracellular activities and cellular communication. Secreted glycoproteins are often reflective of cellular status, and thus glycoproteins from easily accessible bodily fluids can serve as excellent biomarkers for disease detection. Cultured cells have been extensively employed as models in the research fields of biology and biomedicine, and global analysis of glycoproteins secreted from these cells provides insights into cellular activities and glycoprotein functions. However, comprehensive identification and quantification of secreted glycoproteins is a daunting task because of their low abundances compared with the high-abundance serum proteins required for cell growth and proliferation. Several studies employed serum-free media to analyze secreted proteins, but it has been shown that serum starvation, even for a short period of time, can alter protein secretion. To overcome these issues, we developed a method to globally characterize secreted glycoproteins and their N-glycosylation sites from cultured cells by combining selective enrichment of secreted glycoproteins with a boosting approach. The results demonstrated the importance of the boosting sample selection and the boosting-to-sample ratio to improve the coverage of secreted glycoproteins. The method was applied to globally quantify secreted glycoproteins from THP-1 monocytes and macrophages in response to lipopolysaccharides (LPS) and from Hep G2 cells treated with TGF-β without serum starvation. We found differentially secreted glycoproteins in these model systems that showed the response to the immune activation or the epithelial-to-mesenchymal transition. Benefiting from the selective enrichment and the signal enhancement of low-abundance secreted glycoproteins,

this method can be extensively applied to study secreted glycoproteins without serum starvation, which will provide a better understanding of protein secretion and cellular activity.

8.1.5 Differential regulation of the proteome and phosphoproteome of neuroblastoma tumors in response to chemotherapeutic treatment by TOPO/CTX

Neuroblastoma is a pediatric solid tumor that develops from immature nerve cells. High-risk patients are often vulnerable to relapse with a low survival rate. One commonly used treatment for relapsed patients is a combination of topotecan and cyclophosphamide (TOPO/CTX), which causes DNA damage by interfering with the activity of topoisomerase I and alkylating DNA strands, respectively. Using multiplexed proteomics through TMT10plex isobaric tagging, we systematically quantified the abundance changes of proteins and protein phosphorylation in the cell culture-derived xenograft (CDX) and patient-derived xenograft (PDX) models generated in mice treated with a five-day regimen of TOPO/CTX. Both tumors have distinct proteome and phosphoproteome expression profiles considering the inherent heterogeneous nature of neuroblastoma. The response suggests different sensitivity of the tumors to the treatment based on their proteome and phosphoproteome changes through the regulation of the cell-cycle process, which occurred mostly at the proteome level of the CDX models and at the phosphoproteome level of the PDX model. Phosphorylation motif analysis revealed that Ca^{2+}/calmodulin-dependent protein kinase II (CAMK2) is activated in the PDX model, which may be responsible for the different behaviors and the sensitivity of the tumors to the treatment. Remarkably, several proteins and phosphorylation events similarly affected in both tumors were identified, and some of which have already been studied as drug targets in neuroblastoma

such as aurora kinase A. The results will lead to a better understanding of the response of neuroblastoma to chemotherapy and eventually a better treatment in patients.

8.1.6 Evaluation and optimization of reduction and alkylation methods to maximize peptide identification with MS-based proteomics

Mass spectrometry (MS) has become an increasingly important technique to analyze proteins. In popular bottom-up MS-based proteomics, reduction and alkylation are routine steps to facilitate peptide identification. However, the reaction incompletion and side reactions may occur, which will compromise the experimental results. In this work, we systematically evaluated the reduction step with the commonly used reagents, i.e., dithiothreitol, 2-mercaptoethanol, tris(2-carboxyethyl)phosphine, or tris(3-hydroxypropyl)phosphine, and alkylation with iodoacetamide, acrylamide, N-ethylmaleimide, or 4-vinylpyridine. By using digested peptides from a yeast whole-cell lysate, the number of proteins and peptides identified were very similar using four different reducing reagents. The results from four alkylating reagents, however, were dramatically different with iodoacetamide giving the highest number of peptides with alkylated cysteine and the lowest number of peptides with incomplete cysteine alkylation and side reactions. Alkylation conditions with iodoacetamide were further optimized. To identify more peptides with cysteine, Thiopropyl-Sepharose 6B resins were used to enrich them, and the optimal conditions were employed for the reduction and alkylation. The enrichment resulted in over three times more cysteine-containing peptides than without enrichment. Systematic evaluation of the reduction and alkylation with different reagents can aid in a better design of bottom-up proteomic experiments.

8.2 Future directions

With the current development in extracellular glycoproteomics analysis, several types of samples from different biological systems can be studied. These methods will allow for the determination of the roles of surface and secreted glycoproteins in the previously unknown biological systems. One major utilization is in clinical proteomics, where secreted and surface glycoproteins are critical for the diagnosis and treatment of diseases. While the current methods are mostly limited to within the lab, future development will allow for the faster and more efficient approach to move extracellular glycoproteomics from bench to bedside. Current methods also require a large number of cells due to the low abundance of glycoproteins. Future development will solve this problem and possibly scale down the analysis to the single-cell level, which will come from the advances in both the methods to study the glycoproteins and the instrumentation (particularly the mass spectrometer) to study them. A system-wide analysis that incorporates extracellular glycoproteomics analysis will also further our understanding of biological questions. Moreover, results from the glycoproteomics analysis can be combined with other proteomics and transcriptomics analyses to decipher the biological questions as well. Crosstalk studies between glycosylation and other post-translational modifications of surface and secreted glycoproteins, as well as the modifications of proteins at other locations will also facilitate the understanding.

REFERENCES

1. Spiro, R. G., Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **2002,** *12* (4), 43R-56R.

2. Moremen, K. W.; Tiemeyer, M.; Nairn, A. V., Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **2012,** *13* (7), 448-62.

3. Shental-Bechor, D.; Levy, Y., Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105* (24), 8256-61.

4. Helenius, A.; Aebi, M., Intracellular functions of N-linked glycans. *Science* **2001,** *291* (5512), 2364-9.

5. Munkley, J.; Elliott, D. J., Hallmarks of glycosylation in cancer. *Oncotarget* **2016,** *7* (23), 35478-89.

6. Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease. *Cell* **2006,** *126* (5), 855-67.

7. Carr, S. A.; Huddleston, M. J.; Bean, M. F., Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. *Protein Sci.* **1993,** *2* (2), 183-96.

8. Zhang, H.; Li, X.-j.; Martin, D. B., et al., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **2003,** *21* (6), 660-666.

9. Kaji, H.; Saito, H.; Yamauchi, Y., et al., Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol.* **2003,** *21* (6), 667-72.

10. Xiao, H. P.; Wu, R. H., Global and Site-Specific Analysis Revealing Unexpected and Extensive Protein S-GlcNAcylation in Human Cells. *Anal. Chem.* **2017,** *89* (6), 3656-3663.

11. Wu, R. H.; Haas, W.; Dephoure, N., et al., A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat. Methods* **2011,** *8* (8), 677-683.

12. Zhang, Y. J.; Wang, J. Q.; Ding, M., et al., Site-specific characterization of the Asp- and Glu-ADP-ribosylated proteome. *Nat. Methods* **2013,** *10* (10), 981-+.

13. Desaire, H., Glycopeptide Analysis, Recent Developments and Applications. *Mol. Cell. Proteom.* **2013,** *12* (4), 893-901.

14. Reddy, R. J.; Gajadhar, A. S.; Swenson, E. J., et al., Early signaling dynamics of the epidermal growth factor receptor. *Proc. Natl. Acad. Sci. U.S.A.* **2016,** *113* (11), 3114-3119.

15. Simithy, J.; Sidoli, S.; Yuan, Z. F., et al., Characterization of histone acylations links chromatin modifications with metabolism. *Nat. Commun.* **2017,** *8*, 13.

16. Lin, Z. Q.; Wei, L. M.; Cai, W. X., et al., Simultaneous Quantification of Protein Expression and Modifications by Top-down Targeted Proteomics: A Case of the Sarcomeric Subproteome. *Mol. Cell. Proteom.* **2019,** *18* (3), 594-605.

17. Yang, Y.; Franc, V.; Heck, A. J. R., Glycoproteomics: A Balance between High-Throughput and In-Depth Analysis. *Trends Biotechnol.* **2017,** *35* (7), 598-609.

18. Varki, A., *Essentials of glycobiology*. Third edition. ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 2017.

19. Stanley, P.; Taniguchi, N.; Aebi, M., N-Glycans. In *Essentials of Glycobiology*, rd; Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds. Cold Spring Harbor (NY), 2015; pp 99-111.

20. Brockhausen, I.; Stanley, P., O-GalNAc Glycans. In *Essentials of Glycobiology*, rd; Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds. Cold Spring Harbor (NY), 2015; pp 113-123.

21. Salmi, M.; Jalkanen, S., Cell-surface enzymes in control of leukocyte trafficking. *Nat. Rev. Immunol.* **2005,** *5* (10), 760-71.

22. Shankaran, H.; Resat, H.; Wiley, H. S., Cell surface receptors for signal transduction and ligand transport: a design principles study. *PLOS Comput. Biol.* **2007,** *3* (6), e101.

23. Engel, P.; Boumsell, L.; Balderas, R., et al., CD Nomenclature 2015: Human Leukocyte Differentiation Antigen Workshops as a Driving Force in Immunology. *J. Immunol.* **2015,** *195* (10), 4555-63.

24. DeVeale, B.; Bausch-Fluck, D.; Seaberg, R., et al., Surfaceome Profiling Reveals Regulators of Neural Stem Cell Function. *STEM CELLS* **2014,** *32* (1), 258-268.

25. Hofmann, A.; Thiesler, T.; Gerrits, B., et al., Surfaceome of classical Hodgkin and non-Hodgkin lymphoma. *Proteomics Clin. Appl.* **2015,** *9* (7-8), 661-670.

26. Gundry, R. L.; Riordon, D. R.; Tarasova, Y., et al., A cell surfaceome map for immunophenotyping and sorting pluripotent stem cells. *Mol. Cell. Proteom.* **2012,** *11* (8), 303-16.

27. Vernay, A.; Cosson, P., Immunofluorescence labeling of cell surface antigens in Dictyostelium. *BMC Res. Notes* **2013,** *6*, 317.

28. Zarif, J. C.; Yang, W.; Hernandez, J. R., et al., The Identification of Macrophage-enriched Glycoproteins Using Glycoproteomics. *Mol. Cell. Proteom.* **2017,** *16* (6), 1029-1037.

29. Wollscheid, B.; Bausch-Fluck, D.; Henderson, C., et al., Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat. Biotechnol.* **2009,** *27* (4), 378-86.

30. Bausch-Fluck, D.; Hofmann, A.; Bock, T., et al., A mass spectrometric-derived cell surface protein atlas. *PloS one* **2015,** *10* (3), e0121314.

31. Haverland, N. A.; Waas, M.; Ntai, I., et al., Cell Surface Proteomics of N-Linked Glycoproteins for Typing of Human Lymphocytes. *Proteomics* **2017,** *17* (19).

32. Mallanna, S. K.; Waas, M.; Duncan, S. A., et al., N-glycoprotein surfaceome of human induced pluripotent stem cell derived hepatic endoderm. *Proteomics* **2017,** *17* (5).

33. van Oostrum, M.; Müller, M.; Klein, F., et al., "De novo Classification of Mouse B Cell Types using Surfaceome Proteotype Maps". *bioRxiv* **2019**, 620344.

34. Ramya, T. N.; Weerapana, E.; Cravatt, B. F., et al., Glycoproteomics enabled by tagging sialic acid- or galactose-terminated glycans. *Glycobiology* **2013,** *23* (2), 211-21.

35. Sun, F.; Suttapitugsakul, S.; Wu, R., Enzymatic Tagging of Glycoproteins on the Cell Surface for Their Global and Site-Specific Analysis with Mass Spectrometry. *Anal Chem* **2019,** *91* (6), 4195-4203.

36. Hubbard, S. C.; Boyce, M.; McVaugh, C. T., et al., Cell surface glycoproteomic analysis of prostate cancer-derived PC-3 cells. *Bioorg. Med. Chem. Lett.* **2011,** *21* (17), 4945-4950.

37. Suttapitugsakul, S.; Ulmer, L. D.; Jiang, C. D., et al., Surface Glycoproteomic Analysis Reveals That Both Unique and Differential Expression of Surface Glycoproteins Determine the Cell Type. *Anal. Chem.* **2019,** *91* (10), 6934-6942.

38. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem., Int. Ed. Engl.* **2009,** *48* (38), 6974-98.

39. Devaraj, N. K., The Future of Bioorthogonal Chemistry. *ACS Cent. Sci.* **2018,** *4* (8), 952-959.

40. Chen, W.; Smeekens, J. M.; Wu, R., Systematic and site-specific analysis of N-sialoglycosylated proteins on the cell surface by integrating click chemistry and MS-based proteomics. *Chem. Sci.* **2015,** *6* (8), 4681-4689.

41. Laughlin, S. T.; Bertozzi, C. R., Metabolic labeling of glycans with azido sugars and subsequent glycan-profiling and visualization via Staudinger ligation. *Nat. Protoc.* **2007,** *2* (11), 2930-44.

42. Chang, P. V.; Chen, X.; Smyrniotis, C., et al., Metabolic labeling of sialic acids in living animals with alkynyl sugars. *Angew. Chem., Int. Ed. Engl.* **2009,** *48* (22), 4030-3.

43. Spiciarich, D. R.; Nolley, R.; Maund, S. L., et al., Bioorthogonal Labeling of Human Prostate Cancer Tissue Slice Cultures for Glycoproteomics. *Angew. Chem., Int. Ed. Engl.* **2017,** *56* (31), 8992-8997.

44. Li, Q.; Xie, Y.; Xu, G., et al., Identification of potential sialic acid binding proteins on cell membranes by proximity chemical labeling. *Chem. Sci.* **2019,** *10* (24), 6199-6209.

45. Han, S. S.; Lee, D. E.; Shim, H. E., et al., Physiological Effects of Ac4ManNAz and Optimization of Metabolic Labeling for Cell Tracking. *Theranostics* **2017,** *7* (5), 1164-1176.

46. Han, S. S.; Shim, H. E.; Park, S. J., et al., Safety and Optimization of Metabolic Labeling of Endothelial Progenitor Cells for Tracking. *Sci. Rep.* **2018,** *8* (1), 13212.

47. Xiao, H.; Tang, G. X.; Wu, R., Site-specific quantification of surface N-glycoproteins in statin-treated liver cells. *Anal. Chem.* **2016,** *88* (6), 3324-32.

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

49. Lund, R.; Leth-Larsen, R.; Jensen, O. N., et al., Efficient isolation and quantitative proteomic analysis of cancer cell plasma membrane proteins for identification of metastasis-associated cell surface markers. *J. Proteome Res.* **2009,** *8* (6), 3078-90.

50. Fleming, K. G., Applications of Analytical Ultracentrifugation to Membrane Proteins. In *Analytical Ultracentrifugation: Instrumentation, Software, and Applications*, Uchiyama, S.; Arisaka, F.; Stafford, W. F.; Laue, T., Eds. Springer Japan: Tokyo, 2016; pp 311-327.

51. Hormann, K.; Stukalov, A.; Muller, A. C., et al., A Surface Biotinylation Strategy for Reproducible Plasma Membrane Protein Purification and Tracking of Genetic and Drug-Induced Alterations. *J. Proteome Res.* **2016,** *15* (2), 647-58.

52. Matta, C.; Boocock, D. J.; Fellows, C. R., et al., Molecular phenotyping of the surfaceome of migratory chondroprogenitors and mesenchymal stem cells using biotinylation, glycocapture and quantitative LC-MS/MS proteomic analysis. *Sci. Rep.* **2019,** *9* (1), 9018.

53. Kalxdorf, M.; Gade, S.; Eberl, H. C., et al., Monitoring Cell-surface N-Glycoproteome Dynamics by Quantitative Proteomics Reveals Mechanistic Insights into Macrophage Differentiation. *Mol. Cell. Proteom.* **2017,** *16* (5), 770-785.

54. Bausch-Fluck, D.; Goldmann, U.; Muller, S., et al., The in silico human surfaceome. *Proc. Natl. Acad. Sci. U.S.A.* **2018,** *115* (46), E10988-E10997.

55. Arigoni-Affolter, I.; Scibona, E.; Lin, C. W., et al., Mechanistic reconstruction of glycoprotein secretion through monitoring of intracellular N-glycan processing. *Sci Adv* **2019,** *5* (11), eaax8930.

56. Seiradake, E.; Zhao, Y.; Lu, W., et al., Production of cell surface and secreted glycoproteins in mammalian cells. *Methods Mol Biol* **2015,** *1261*, 115-27.

57. Yagi, H.; Yagi-Utsumi, M.; Honda, R., et al., Improved secretion of glycoproteins using an N-glycan-restricted passport sequence tag recognized by cargo receptor. *Nat Commun* **2020,** *11* (1), 1368.

58. Uhlen, M.; Karlsson, M. J.; Hober, A., et al., The human secretome. *Sci Signal* **2019,** *12* (609).

59. Kirwan, A.; Utratna, M.; O'Dwyer, M. E., et al., Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. *Biomed Res Int* **2015,** *2015*, 490531.

60. Ideo, H.; Kondo, J.; Nomura, T., et al., Study of glycosylation of prostate-specific antigen secreted by cancer tissue-originated spheroids reveals new candidates for prostate cancer detection. *Sci Rep* **2020,** *10* (1), 2708.

61. Boersema, P. J.; Geiger, T.; Wisniewski, J. R., et al., Quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples. *Mol Cell Proteomics* **2013,** *12* (1), 158-71.

62. Frenette, G. P.; Carey, T. E.; Varani, J., et al., Biosynthesis and secretion of laminin and laminin-associated glycoproteins by nonmalignant and malignant human keratinocytes: comparison of cell lines from primary and secondary tumors in the same patient. *Cancer Res* **1988,** *48* (18), 5193-202.

63. Gal, S.; Willingham, M. C.; Gottesman, M. M., Processing and lysosomal localization of a glycoprotein whose secretion is transformation stimulated. *J Cell Biol* **1985,** *100* (2), 535-44.

64. Hortin, G. L.; Sviridov, D., The dynamic range problem in the analysis of the plasma proteome. *J Proteomics* **2010,** *73* (3), 629-36.

65. Pare, B.; Deschenes, L. T.; Pouliot, R., et al., An Optimized Approach to Recover Secreted Proteins from Fibroblast Conditioned-Media for Secretomic Analysis. *Front Cell Neurosci* **2016,** *10*, 70.

66. Weaving, G.; Batstone, G. F.; Jones, R. G., Age and sex variation in serum albumin concentration: an observational study. *Ann Clin Biochem* **2016,** *53* (Pt 1), 106-11.

67. Merrell, K.; Southwick, K.; Graves, S. W., et al., Analysis of low-abundance, low-molecular-weight serum proteins using mass spectrometry. *J Biomol Tech* **2004,** *15* (4), 238-248.

68. Kim, H. O.; Kim, H. S.; Youn, J. C., et al., Serum cytokine profiles in healthy young and elderly population assessed using multiplexed bead-based immunoassays. *J Transl Med* **2011,** *9*, 113.

69. Dey, K. K.; Wang, H.; Niu, M., et al., Deep undepleted human serum proteome profiling toward biomarker discovery for Alzheimer's disease. *Clinical Proteomics* **2019,** *16* (1), 16.

70. Totten, S. M.; Adusumilli, R.; Kullolli, M., et al., Multi-lectin Affinity Chromatography and Quantitative Proteomic Analysis Reveal Differential Glycoform Levels between Prostate Cancer and Benign Prostatic Hyperplasia Sera. *Sci Rep* **2018,** *8* (1), 6509.

71. Berven, F. S.; Ahmad, R.; Clauser, K. R., et al., Optimizing performance of glycopeptide capture for plasma proteomics. *J Proteome Res* **2010,** *9* (4), 1706-15.

72. Tu, C.; Rudnick, P. A.; Martinez, M. Y., et al., Depletion of abundant plasma proteins and limitations of plasma proteomics. *J Proteome Res* **2010,** *9* (10), 4982-91.

73. Jankovska, E.; Svitek, M.; Holada, K., et al., Affinity depletion versus relative protein enrichment: a side-by-side comparison of two major strategies for increasing human cerebrospinal fluid proteome coverage. *Clin Proteomics* **2019,** *16*, 9.

74. Merlot, A. M.; Kalinowski, D. S.; Richardson, D. R., Unraveling the mysteries of serum albumin-more than just a serum protein. *Front Physiol* **2014,** *5*, 299.

75. Granger, J.; Siddiqui, J.; Copeland, S., et al., Albumin depletion of human plasma also removes low abundance proteins including the cytokines. *Proteomics* **2005,** *5* (18), 4713-8.

76. Polaskova, V.; Kapur, A.; Khan, A., et al., High-abundance protein depletion: comparison of methods for human plasma biomarker discovery. *Electrophoresis* **2010,** *31* (3), 471-82.

77. de Jesus, J. R.; da Silva Fernandes, R.; de Souza Pessoa, G., et al., Depleting high-abundant and enriching low-abundant proteins in human serum: An evaluation of sample preparation methods using magnetic nanoparticle, chemical depletion and immunoaffinity techniques. *Talanta* **2017,** *170*, 199-209.

78. Zheng, X.; Baker, H.; Hancock, W. S., et al., Proteomic analysis for the assessment of different lots of fetal bovine serum as a raw material for cell culture. Part IV. Application of proteomics to the manufacture of biological drugs. *Biotechnol Prog* **2006,** *22* (5), 1294-300.

79. Chang, Y. H.; Wu, C. C.; Chang, K. P., et al., Cell secretome analysis using hollow fiber culture system leads to the discovery of CLIC1 protein as a novel plasma marker for nasopharyngeal carcinoma. *J Proteome Res* **2009,** *8* (12), 5465-74.

80. Stastna, M.; Van Eyk, J. E., Investigating the secretome: lessons about the cells that comprise the heart. *Circ Cardiovasc Genet* **2012,** *5* (1), o8-o18.

81. Rogers, L. J.; Suchowerska, N.; Khan, A., et al., Profiling of the secretome of human cancer cells: preparation of supernatant for proteomic analysis. *Electrophoresis* **2014,** *35* (18), 2626-33.

82. Eichelbaum, K.; Winter, M.; Berriel Diaz, M., et al., Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nat Biotechnol* **2012,** *30* (10), 984-90.

83. Mechref, Y.; Madera, M.; Novotny, M. V., Glycoprotein enrichment through lectin affinity techniques. *Methods Mol Biol* **2008,** *424*, 373-96.

84. Abbott, K. L.; Pierce, J. M., Lectin-based glycoproteomic techniques for the enrichment and identification of potential biomarkers. *Methods Enzymol* **2010,** *480*, 461-76.

85. Feng, S.; Yang, N.; Pennathur, S., et al., Enrichment of Glycoproteins using Nanoscale Chelating Concanavalin A Monolithic Capillary Chromatography. *Analytical Chemistry* **2009,** *81* (10), 3776-3783.

86. Wei, X.; Dulberger, C.; Li, L. J., Characterization of Murine Brain Membrane Glycoproteins by Detergent Assisted Lectin Affinity Chromatography. *Analytical Chemistry* **2010,** *82* (15), 6329-6333.

87. Zhang, Y.; Tang, X.; Yao, L., et al., Lectin capture strategy for effective analysis of cell secretome. *Proteomics* **2012,** *12* (1), 32-6.

88. Li, X.; Jiang, J.; Zhao, X., et al., N-glycoproteome analysis of the secretome of human metastatic hepatocellular carcinoma cell lines combining hydrazide chemistry, HILIC enrichment and mass spectrometry. *PLoS One* **2013,** *8* (12), e81921.

89. Alpert, A. J., Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal Chem* **2008,** *80* (1), 62-76.

90. Shu, Q.; Li, M.; Shu, L., et al., Large-scale Identification of N-linked Intact Glycopeptides in Human Serum using HILIC Enrichment and Spectral Library Search. *Mol Cell Proteomics* **2020,** *19* (4), 672-689.

91. Alagesan, K.; Khilji, S. K.; Kolarich, D., It is all about the solvent: on the importance of the mobile phase for ZIC-HILIC glycopeptide enrichment. *Anal Bioanal Chem* **2017,** *409* (2), 529-538.

92. Cao, W.; Huang, J.; Jiang, B., et al., Highly Selective Enrichment of Glycopeptides Based on Zwitterionically Functionalized Soluble Nanopolymers. *Sci Rep* **2016,** *6*, 29776.

93. Adav, S. S.; Hwa, H. H.; de Kleijn, D., et al., Improving Blood Plasma Glycoproteome Coverage by Coupling Ultracentrifugation Fractionation to Electrostatic Repulsion-Hydrophilic Interaction Chromatography Enrichment. *J Proteome Res* **2015,** *14* (7), 2828-38.

94. Chen, R.; Williamson, S.; Fulton, K. M., et al., Simultaneous analysis of phosphopeptides and intact glycopeptides from secretome with mode switchable solid phase extraction. *Analytical Methods* **2019,** *11* (41), 5243-5249.

95. Chen, W.; Smeekens, J. M.; Wu, R., A universal chemical enrichment method for mapping the yeast N-glycoproteome by mass spectrometry (MS). *Mol Cell Proteomics* **2014,** *13* (6), 1563-72.

96. Xiao, H.; Chen, W.; Smeekens, J. M., et al., An enrichment method based on synergistic and reversible covalent interactions for large-scale analysis of glycoproteins. *Nat Commun* **2018,** *9* (1), 1692.

97. Zhang, L. J.; Xu, Y. W.; Yao, H. L., et al., Boronic Acid Functionalized Core-Satellite Composite Nanoparticles for Advanced Enrichment of Glycopeptides and Glycoproteins. *Chem.-Eur. J.* **2009,** *15* (39), 10158-10166.

98. Liu, L. T.; Zhang, Y.; Zhang, L., et al., Highly specific revelation of rat serum glycopeptidome by boronic acid-functionalized mesoporous silica. *Anal. Chim. Acta* **2012,** *753*, 64-72.

99. Qu, Y. Y.; Liu, J. X.; Yang, K. G., et al., Boronic Acid Functionalized Core-Shell Polymer Nanoparticles Prepared by Distillation Precipitation Polymerization for Glycopeptide Enrichment. *Chem.-Eur. J.* **2012,** *18* (29), 9056-9062.

100. Xiao, H. P.; Hwang, J. E.; Wu, R. H., Mass spectrometric analysis of the N-glycoproteome in statin-treated liver cells with two lectin-independent chemical enrichment methods. *Int. J. Mass Spectrom.* **2018,** *429*, 66-75.

101. Liu, Z.; He, H., Synthesis and Applications of Boronate Affinity Materials: From Class Selectivity to Biomimetic Specificity. *Accounts Chem. Res.* **2017,** *50* (9), 2185-2193.

102. Xiao, H. P.; Wu, R. H., Simultaneous Quantitation of Glycoprotein Degradation and Synthesis Rates by Integrating Isotope Labeling, Chemical Enrichment, and Multiplexed Proteomics. *Analytical Chemistry* **2017,** *89* (19), 10361-10367.

103. Sparbier, K.; Asperger, A.; Resemann, A., et al., Analysis of glycoproteins in human serum by means of glycospecific magnetic bead separation and LC-MALDI-TOF/TOF analysis with automated glycopeptide detection. *J Biomol Tech* **2007,** *18* (4), 252-8.

104. Smeekens, J. M.; Xiao, H.; Wu, R., Global Analysis of Secreted Proteins and Glycoproteins in Saccharomyces cerevisiae. *J Proteome Res* **2017,** *16* (2), 1039-1049.

105. Goo, Y. A.; Liu, A. Y.; Ryu, S., et al., Identification of secreted glycoproteins of human prostate and bladder stromal cells by comparative quantitative proteomics. *Prostate* **2009,** *69* (1), 49-61.

106. Wang, L.; Aryal, U. K.; Dai, Z., et al., Mapping N-linked glycosylation sites in the secretome and whole cells of Aspergillus niger using hydrazide chemistry and mass spectrometry. *J Proteome Res* **2012,** *11* (1), 143-56.

107. Sajic, T.; Liu, Y.; Arvaniti, E., et al., Similarities and Differences of Blood N-Glycoproteins in Five Solid Carcinomas at Localized Clinical Stage Analyzed by SWATH-MS. *Cell Rep* **2018,** *23* (9), 2819-2831 e5.

108. Roper, S. M.; Zemskova, M.; Neely, B. A., et al., Targeted glycoprotein enrichment and identification in stromal cell secretomes using azido sugar metabolic labeling. *Proteomics Clin Appl* **2013,** *7* (5-6), 367-71.

109. Kuhn, P. H.; Koroniak, K.; Hogl, S., et al., Secretome protein enrichment identifies physiological BACE1 protease substrates in neurons. *EMBO J* **2012,** *31* (14), 3157-68.

110. Tushaus, J.; Muller, S. A.; Kataka, E. S., et al., An optimized quantitative proteomics method establishes the cell type-resolved mouse brain secretome. *EMBO J* **2020,** *39* (20), e105693.

111. Budnik, B.; Levy, E.; Harmange, G., et al., SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol* **2018,** *19* (1), 161.

112. Yi, L.; Tsai, C. F.; Dirice, E., et al., Boosting to Amplify Signal with Isobaric Labeling (BASIL) Strategy for Comprehensive Quantitative Phosphoproteomic Characterization of Small Populations of Cells. *Anal Chem* **2019,** *91* (9), 5794-5801.

113. Russell, C. L.; Heslegrave, A.; Mitra, V., et al., Combined tissue and fluid proteomics with Tandem Mass Tags to identify low-abundance protein biomarkers of disease in peripheral body fluid: An Alzheimer's Disease case study. *Rapid Commun Mass Spectrom* **2017,** *31* (2), 153-159.

114. Suttapitugsakul, S.; Tong, M.; Sun, F., et al., Enhancing Comprehensive Analysis of Secreted Glycoproteins from Cultured Cells without Serum Starvation. *Analytical Chemistry* **2021**.

115. Nagarajah, S., Exosome Secretion - More Than Simple Waste Disposal? Implications for Physiology, Diagnostics and Therapeutics. *J. Circ. Biomark.* **2016,** *5*, 7.

116. Colombo, M.; Raposo, G.; Thery, C., Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles. *Annu. Rev. Cell Dev. Biol.* **2014,** *30*, 255-89.

117. Hessvik, N. P.; Llorente, A., Current knowledge on exosome biogenesis and release. *Cell Mol Life Sci* **2018,** *75* (2), 193-208.

118. Chen, I. H.; Xue, L.; Hsu, C. C., et al., Phosphoproteins in extracellular vesicles as candidate markers for breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **2017,** *114* (12), 3175-3180.

119. Liang, Y.; Eng, W. S.; Colquhoun, D. R., et al., Complex N-linked glycans serve as a determinant for exosome/microvesicle cargo recruitment. *J. Biol. Chem.* **2014,** *289* (47), 32526-37.

120. Pan, D.; Chen, J.; Feng, C., et al., Preferential Localization of MUC1 Glycoprotein in Exosomes Secreted by Non-Small Cell Lung Carcinoma Cells. *Int. J. Mol. Sci.* **2019,** *20* (2).

121. Chauhan, S.; Danielson, S.; Clements, V., et al., Surface Glycoproteins of Exosomes Shed by Myeloid-Derived Suppressor Cells Contribute to Function. *J. Proteome Res.* **2017,** *16* (1), 238-246.

122. Escrevente, C.; Grammel, N.; Kandzia, S., et al., Sialoglycoproteins and N-glycans from secreted exosomes of ovarian carcinoma cells. *PLoS One* **2013,** *8* (10), e78631.

123. Sharma, S.; Alharbi, M.; Kobayashi, M., et al., Proteomic analysis of exosomes reveals an association between cell invasiveness and exosomal bioactivity on endothelial and mesenchymal cell migration in vitro. *Clin. Sci.* **2018,** *132* (18), 2029-2044.

124. Melo, S. A.; Luecke, L. B.; Kahlert, C., et al., Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature* **2015,** *523* (7559), 177-82.

125. Kuang, M.; Tao, X.; Peng, Y., et al., Proteomic analysis of plasma exosomes to differentiate malignant from benign pulmonary nodules. *Clin. Proteomics* **2019,** *16*, 5.

126. Niu, L.; Song, X.; Wang, N., et al., Tumor-derived exosomal proteins as diagnostic biomarkers in non-small cell lung cancer. *Cancer Sci.* **2019,** *110* (1), 433-442.

127. Bai, H.; Pan, Y.; Qi, L., et al., Development a hydrazide-functionalized thermosensitive polymer based homogeneous system for highly efficient N-glycoprotein/glycopeptide enrichment from human plasma exosome. *Talanta* **2018,** *186*, 513-520.

128. Pisitkun, T.; Shen, R. F.; Knepper, M. A., Identification and proteomic profiling of exosomes in human urine. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101* (36), 13368-73.

129. Lim, J. H.; Lee, C. H.; Kim, K. Y., et al., Novel urinary exosomal biomarkers of acute T cell-mediated rejection in kidney transplant recipients: A cross-sectional study. *PLoS One* **2018,** *13* (9), e0204204.

130. Zou, G.; Benktander, J. D.; Gizaw, S. T., et al., Comprehensive Analytical Approach toward Glycomic Characterization and Profiling in Urinary Exosomes. *Anal. Chem.* **2017,** *89* (10), 5364-5372.

131. Zhang, H.; Freitas, D.; Kim, H. S., et al., Identification of distinct nanoparticles and subsets of extracellular vesicles by asymmetric flow field-flow fractionation. *Nat. Cell Biol.* **2018,** *20* (3), 332-343.

132. Zhang, Q.; Higginbotham, J. N.; Jeppesen, D. K., et al., Transfer of Functional Cargo in Exomeres. *Cell Rep.* **2019,** *27* (3), 940-954 e6.

133. Freitas, D.; Balmaña, M.; Poças, J., et al., Different isolation approaches lead to diverse glycosylated extracellular vesicle populations. *J. Extracell. Vesicles* **2019,** *8* (1), 1621131.

134. Thery, C.; Witwer, K. W.; Aikawa, E., et al., Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J. Extracell. Vesicles* **2018,** *7* (1), 1535750.

135. Mathivanan, S.; Fahner, C. J.; Reid, G. E., et al., ExoCarta 2012: database of exosomal proteins, RNA and lipids. *Nucleic Acids Res.* **2012,** *40* (Database issue), D1241-4.

136. Keerthikumar, S.; Chisanga, D.; Ariyaratne, D., et al., ExoCarta: A Web-Based Compendium of Exosomal Cargo. *J. Mol. Biol.* **2016,** *428* (4), 688-692.

137. Pathan, M.; Fonseka, P.; Chitti, S. V., et al., Vesiclepedia 2019: a compendium of RNA, proteins, lipids and metabolites in extracellular vesicles. *Nucleic Acids Res.* **2019,** *47* (D1), D516-D519.

138. Kim, D. K.; Lee, J.; Simpson, R. J., et al., EVpedia: A community web resource for prokaryotic and eukaryotic extracellular vesicles research. *Semin. Cell Dev. Biol.* **2015,** *40*, 4-7.

139. Stanley, P.; Schachter, H.; Taniguchi, N., N-Glycans. In *Essentials of Glycobiology*, nd; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor (NY), 2009.

140. Ekblom, P.; Vestweber, D.; Kemler, R., Cell-matrix interactions and cell adhesion during development. *Annu. Rev. Cell Biol.* **1986,** *2*, 27-47.

141. Josic, D.; Clifton, J. G., Mammalian plasma membrane proteomics. *Proteomics* **2007,** *7* (16), 3010-29.

142. Rosenbaum, D. M.; Rasmussen, S. G.; Kobilka, B. K., The structure and function of G-protein-coupled receptors. *Nature* **2009,** *459* (7245), 356-63.

143. le Coutre, J.; Kaback, H. K., Structure-function relationships of integral membrane proteins: membrane transporters vs channels. *Biopolymers* **2000,** *55* (4), 297-307.

144. Gahmberg, C. G.; Tolvanen, M., Why mammalian cell surface proteins are glycoproteins. *Trends Biochem. Sci.* **1996,** *21* (8), 308-11.

145. Drake, P. M.; Cho, W.; Li, B., et al., Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin. Chem.* **2010,** *56* (2), 223-36.

146. Kimiz-Gebologlu, I.; Gulce-Iz, S.; Biray-Avci, C., Monoclonal antibodies in cancer immunotherapy. *Mol. Biol. Rep.* **2018,** *45* (6), 2935-2940.

147. Craig, F. E.; Foon, K. A., Flow cytometric immunophenotyping for hematologic neoplasms. *Blood* **2008,** *111* (8), 3941-67.

148. Witze, E. S.; Old, W. M.; Resing, K. A., et al., Mapping protein post-translational modifications with mass spectrometry. *Nat. Methods* **2007,** *4* (10), 798-806.

149. Zhu, Z. K.; Desaire, H., Carbohydrates on Proteins: Site-Specific Glycosylation Analysis by Mass Spectrometry. In *Annual Review of Analytical Chemistry, Vol 8*, Cooks, R. G.; Pemberton, J. E., Eds. Annual Reviews: Palo Alto, 2015; Vol. 8, pp 463-483.

150. Yang, Y.; Liu, F.; Franc, V., et al., Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat. Comm.* **2016,** *7*, 10.

151. Shi, T. J.; Fillmore, T. L.; Sun, X. F., et al., Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (38), 15395-15400.

152. Yu, C.; Mao, H. B.; Novitsky, E. J., et al., Gln40 deamidation blocks structural reconfiguration and activation of SCF ubiquitin ligase complex by Nedd8. *Nat. Comm.* **2015,** *6*, 12.

153. Wu, J.; Xie, X. L.; Liu, Y. S., et al., Identification and confirmation of differentially expressed fucosylated glycoproteins in the serum of ovarian cancer patients using a lectin array and LC-MS/MS. *J. Proteome Res.* **2012,** *11* (9), 4541-4552.

154. Wang, X. S.; Yuan, Z. F.; Fan, J., et al., A novel quantitative mass spectrometry platform for determining protein O-GlcNAcylation dynamics. *Mol. Cell. Proteomics* **2016,** *15* (7), 2462-2475.

155. Banazadeh, A.; Veillon, L.; Wooding, K. M., et al., Recent advances in mass spectrometric analysis of glycoproteins. *Electrophoresis* **2017,** *38* (1), 162-189.

156. Zheng, J. N.; Xiao, H. P.; Wu, R. H., Specific identification of glycoproteins bearing the Tn antigen in human cells. *Angew. Chem.-Int. Edit.* **2017,** *56* (25), 7107-7111.

157. Zhang, H.; Li, X. J.; Martin, D. B., et al., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **2003,** *21* (6), 660-6.

158. Rillahan, C. D.; Paulson, J. C., Glycan microarrays for decoding the glycome. In *Annual Review of Biochemistry, Vol 80*, Kornberg, R. D.; Raetz, C. R. H.; Rothman, J. E.; Thorner, J. W., Eds. Annual Reviews: Palo Alto, 2011; Vol. 80, pp 797-823.

159. Zauner, G.; Deelder, A. M.; Wuhrer, M., Recent advances in hydrophilic interaction liquid chromatography (HILIC) for structural glycomics. *Electrophoresis* **2011,** *32* (24), 3456-3466.

160. Suttapitugsakul, S.; Xiao, H.; Smeekens, J., et al., Evaluation and optimization of reduction and alkylation methods to maximize peptide identification with MS-based proteomics. *Mol Biosyst* **2017,** *13* (12), 2574-2582.

161. Rappsilber, J.; Mann, M.; Ishihama, Y., Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2007,** *2* (8), 1896-906.

162. Xiao, H.; Wu, R., Global and Site-Specific Analysis Revealing Unexpected and Extensive Protein S-GlcNAcylation in Human Cells. *Anal Chem* **2017,** *89* (6), 3656-3663.

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

164. Cox, J.; Neuhauser, N.; Michalski, A., et al., Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **2011,** *10* (4), 1794-805.

165. Tyanova, S.; Temu, T.; Sinitcyn, P., et al., The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **2016,** *13* (9), 731-40.

166. Kall, L.; Krogh, A.; Sonnhammer, E. L., A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **2004,** *338* (5), 1027-36.

167. Bendtsen, J. D.; Jensen, L. J.; Blom, N., et al., Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel* **2004,** *17* (4), 349-56.

168. Petersen, B.; Petersen, T. N.; Andersen, P., et al., A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct. Biol.* **2009,** *9*, 51.

169. Schwanhausser, B.; Busse, D.; Li, N., et al., Global quantification of mammalian gene expression control. *Nature* **2011,** *473* (7347), 337-42.

170. Huttlin, E. L.; Jedrychowski, M. P.; Elias, J. E., et al., A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **2010,** *143* (7), 1174-89.

171. Geiger, T.; Wehner, A.; Schaab, C., et al., Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* **2012,** *11* (3), M111 014050.

172. Shannon, P.; Markiel, A.; Ozier, O., et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003,** *13* (11), 2498-504.

173. Szklarczyk, D.; Franceschini, A.; Wyder, S., et al., STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **2015,** *43* (Database issue), D447-52.

174. Bindea, G.; Mlecnik, B.; Hackl, H., et al., ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **2009,** *25* (8), 1091-3.

175. Mahal, L. K.; Yarema, K. J.; Bertozzi, C. R., Engineering chemical reactivity on cell surfaces through oligosaccharide biosynthesis. *Science* **1997,** *276* (5315), 1125-1128.

176. Chen, W. X.; Smeekens, J. M.; Wu, R. H., Systematic and site-specific analysis of N-sialoglycosylated proteins on the cell surface by integrating click chemistry and MS-based proteomics. *Chem. Sci.* **2015,** *6* (8), 4681-4689.

177. Smeekens, J. M.; Chen, W. X.; Wu, R. H., Mass spectrometric analysis of the cell surface N-glycoproteome by combining metabolic labeling and click chemistry. *J. Am. Soc. Mass Spectrom.* **2015,** *26* (4), 604-614.

178. Baskin, J. M.; Prescher, J. A.; Laughlin, S. T., et al., Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (43), 16793-16797.

179. Deeb, S. J.; Cox, J.; Schmidt-Supprian, M., et al., N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes. *Mol. Cell. Proteomics* **2014,** *13* (1), 240-51.

180. Sun, S.; Shah, P.; Eshghi, S. T., et al., Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat. Biotechnol.* **2016,** *34* (1), 84-8.

181. Xiao, H. P.; Suttapitugsakul, S.; Sun, F. X., et al., Mass spectrometry-based chemical and enzymatic methods for global analysis of protein glycosylation. *Acc. Chem. Res.* **2018,** *51* (8), 1796-1806.

182. Laughlin, S. T.; Baskin, J. M.; Amacher, S. L., et al., In vivo imaging of membrane-associated glycans in developing zebrafish. *Science* **2008,** *320* (5876), 664-7.

183. Xiong, D. C.; Zhu, J. J.; Han, M. J., et al., Rapid probing of sialylated glycoproteins in vitro and in vivo via metabolic oligosaccharide engineering of a minimal cyclopropene reporter. *Org. Biomol. Chem.* **2015,** *13* (13), 3911-3917.

184. Feo, S.; Arcuri, D.; Piddini, E., et al., ENO1 gene product binds to the c-myc promoter and acts as a transcriptional repressor: relationship with Myc promoter-binding protein 1 (MBP-1). *FEBS Lett.* **2000,** *473* (1), 47-52.

185. Hsu, L. C.; Liu, S.; Abedinpour, F., et al., The murine G+C-rich promoter binding protein mGPBP is required for promoter-specific transcription. *Mol. Cell. Biol.* **2003,** *23* (23), 8773-85.

186. Zielinska, D. F.; Gnad, F.; Wisniewski, J. R., et al., Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **2010,** *141* (5), 897-907.

187. Zajonc, D. M.; Striegl, H.; Dascher, C. C., et al., The crystal structure of avian CD1 reveals a smaller, more primordial antigen-binding pocket compared to mammalian CD1. *Proc. Natl. Acad. Sci. USA* **2008,** *105* (46), 17925-30.

188. Lam, P. V.; Goldman, R.; Karagiannis, K., et al., Structure-based comparative analysis and prediction of N-linked glycosylation sites in evolutionarily distant eukaryotes. *Genomics Proteomics Bioinformatics* **2013,** *11* (2), 96-104.

189. Chen, X.; VanValkenburgh, C.; Liang, H., et al., Signal peptidase and oligosaccharyltransferase interact in a sequential and dependent manner within the endoplasmic reticulum. *J. Biol. Chem.* **2001,** *276* (4), 2411-6.

190. Walmsley, A. R.; Hooper, N. M., Distance of sequons to the C-terminus influences the cellular N-glycosylation of the prion protein. *Biochem. J.* **2003,** *370* (Pt 1), 351-5.

191. Nilsson, I.; von Heijne, G., Glycosylation efficiency of Asn-Xaa-Thr sequons depends both on the distance from the C terminus and on the presence of a downstream transmembrane segment. *J. Biol. Chem.* **2000,** *275* (23), 17338-43.

192. Park, D. D.; Xu, G.; Wong, M., et al., Membrane glycomics reveal heterogeneity and quantitative distribution of cell surface sialylation. *Chem. Sci.* **2018,** *9* (29), 6271-6285.

193. Van Acker, H. H.; Capsomidis, A.; Smits, E. L., et al., CD56 in the immune system: more than a marker for cytotoxicity? *Front. Immunol.* **2017,** *8*, 892.

194. Bernstein, H. B.; Plasterer, M. C.; Schiff, S. E., et al., CD4 expression on activated NK cells: ligation of CD4 induces cytokine expression and cell migration. *J. Immunol.* **2006,** *177* (6), 3669-76.

195. Soma, L.; Wu, D.; Chen, X., et al., Apparent CD19 expression by natural killer cells: a potential confounder for minimal residual disease detection by flow cytometry in B lymphoblastic leukemia. *Cytometry B Clin. Cytom.* **2015,** *88* (2), 145-7.

196. Mogensen, T. H., Pathogen recognition and inflammatory signaling in innate immune defenses. *Clin Microbiol Rev* **2009,** *22* (2), 240-73, Table of Contents.

197. Turvey, S. E.; Broide, D. H., Innate immunity. *J Allergy Clin Immunol* **2010,** *125* (2 Suppl 2), S24-32.

198. Akira, S.; Uematsu, S.; Takeuchi, O., Pathogen recognition and innate immunity. *Cell* **2006,** *124* (4), 783-801.

199. Bonilla, F. A.; Oettgen, H. C., Adaptive immunity. *J Allergy Clin Immunol* **2010,** *125* (2 Suppl 2), S33-40.

200. Masopust, D.; Picker, L. J., Hidden memories: frontline memory T cells and early pathogen interception. *J Immunol* **2012,** *188* (12), 5811-7.

201. Zhu, Y.; Yao, S.; Chen, L., Cell surface signaling molecules in the control of immune responses: a tide model. *Immunity* **2011,** *34* (4), 466-78.

202. Rabinovich, G. A.; van Kooyk, Y.; Cobb, B. A., Glycobiology of immune responses. *Ann N Y Acad Sci* **2012,** *1253*, 1-15.

203. Iwasaki, A.; Medzhitov, R., Control of adaptive immunity by the innate immune system. *Nat Immunol* **2015,** *16* (4), 343-53.

204. Osborn, M. J.; Rosen, S. M.; Rothfield, L., et al., Lipopolysaccharide of the Gram-Negative Cell Wall. *Science* **1964,** *145* (3634), 783-9.

205. Lu, Y. C.; Yeh, W. C.; Ohashi, P. S., LPS/TLR4 signal transduction pathway. *Cytokine* **2008,** *42* (2), 145-151.

206. Alexander, C.; Rietschel, E. T., Bacterial lipopolysaccharides and innate immunity. *J Endotoxin Res* **2001,** *7* (3), 167-202.

207. Laupeze, B.; Fardel, O.; Onno, M., et al., Differential expression of major histocompatibility complex class Ia, Ib, and II molecules on monocytes-derived dendritic and macrophagic cells. *Hum Immunol* **1999,** *60* (7), 591-7.

208. Chanput, W.; Mes, J.; Vreeburg, R. A., et al., Transcription profiles of LPS-stimulated THP-1 monocytes and macrophages: a tool to study inflammation modulating effects of food-derived compounds. *Food Funct* **2010,** *1* (3), 254-61.

209. Chanput, W.; Mes, J. J.; Wichers, H. J., THP-1 cell line: an in vitro cell model for immune modulation approach. *Int Immunopharmacol* **2014,** *23* (1), 37-45.

210. Meijer, K.; Weening, D.; de Vries, M. P., et al., Quantitative proteomics analyses of activation states of human THP-1 macrophages. *J Proteomics* **2015,** *128*, 164-72.

211. Tarasova, N. K.; Ytterberg, A. J.; Lundberg, K., et al., Establishing a proteomics-based monocyte assay to assess differential innate immune activation responses. *J. Proteome Res.* **2016,** *15* (7), 2337-2345.

212. Genin, M.; Clement, F.; Fattaccioli, A., et al., M1 and M2 macrophages derived from THP-1 cells differentially modulate the response of cancer cells to etoposide. *BMC Cancer* **2015,** *15*, 577.

213. Gadgil, H. S.; Pabst, K. M.; Giorgianni, F., et al., Proteome of monocytes primed with lipopolysaccharide: analysis of the abundant proteins. *Proteomics* **2003,** *3* (9), 1767-80.

214. Chanput, W.; Mes, J. J.; Savelkoul, H. F., et al., Characterization of polarized THP-1 macrophages and polarizing ability of LPS and food compounds. *Food Funct.* **2013,** *4* (2), 266-76.

215. Smeekens, J. M.; Chen, W.; Wu, R., Mass spectrometric analysis of the cell surface N-glycoproteome by combining metabolic labeling and click chemistry. *J Am Soc Mass Spectrom* **2015,** *26* (4), 604-14.

216. Suttapitugsakul, S.; Ulmer, L. D.; Jiang, C., et al., Surface Glycoproteomic Analysis Reveals That Both Unique and Differential Expression of Surface Glycoproteins Determine the Cell Type. *Anal Chem* **2019,** *91* (10), 6934-6942.

217. Eng, J. K.; McCormack, A. L.; Yates, J. R., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **1994,** *5* (11), 976-89.

218. Elias, J. E.; Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **2007,** *4* (3), 207-14.

219. Beausoleil, S. A.; Villen, J.; Gerber, S. A., et al., A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **2006,** *24* (10), 1285-92.

220. Huang, D. W.; Sherman, B. T.; Lempicki, R. A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009,** *4* (1), 44-57.

221. Yates, J. R.; Ruse, C. I.; Nakorchevsky, A., Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **2009,** *11*, 49-79.

222. Aebersold, R.; Mann, M., Mass-spectrometric exploration of proteome structure and function. *Nature* **2016,** *537* (7620), 347-355.

223. Tran, J. C.; Zamdborg, L.; Ahlf, D. R., et al., Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **2011,** *480* (7376), 254-U141.

224. Zhang, H.; Li, X. J.; Martin, D. B., et al., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotech.* **2003,** *21* (6), 660-666.

225. Yang, Y.; Franc, V.; Heck, A. J. R., Glycoproteomics: a balance between high-throughput and in-depth analysis. *Trends Biotechnol.* **2017,** *35* (7), 598-609.

226. Zhou, C.; Schulz, B. L., Glycopeptide variable window SWATH for improved data independent acquisition glycoprotein analysis. *Anal. Biochem.* **2020,** *597*, 7.

227. Thaysen-Andersen, M.; Packer, N. H.; Schulz, B. L., Maturing glycoproteomics technologies provide unique structural insights into the N-glycoproteome and its regulation in health and disease. *Mol. Cell. Proteomics* **2016,** *15* (6), 1773-1790.

228. Wollscheid, B.; Bausch-Fluck, D.; Henderson, C., et al., Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat. Biotech.* **2009,** *27* (4), 378-386.

229. Bausch-Fluck, D.; Hofmann, A.; Bock, T., et al., A mass spectrometric-derived cell surface protein atlas. *Plos One* **2015,** *10* (4), 22.

230. Haverland, N. A.; Waas, M.; Ntai, I., et al., Cell surface proteomics of n-linked glycoproteins for typing of human lymphocytes. *Proteomics* **2017,** *17* (19), 5.

231. Leung, K. K.; Wilson, G. M.; Kirkemo, L. L., et al., Broad and thematic remodeling of the surfaceome and glycoproteome on isogenic cells transformed with driving proliferative oncogenes. *Proc. Natl. Acad. Sci. U. S. A.* **2020,** *117* (14), 7764-7775.

232. Spiciarich, D. R.; Nolley, R.; Maund, S. L., et al., Bioorthogonal labeling of human prostate cancer tissue slice cultures for glycoproteomics. *Angewandte Chemie-International Edition* **2017,** *56* (31), 8992-8997.

233. Elgueta, R.; Benson, M. J.; de Vries, V. C., et al., Molecular mechanism and function of CD40/CD40L engagement in the immune system. *Immunol Rev* **2009,** *229* (1), 152-72.

234. Prechtel, A. T.; Steinkasserer, A., CD83: an update on functions and prospects of the maturation marker of dendritic cells. *Arch Dermatol Res* **2007,** *299* (2), 59-69.

235. Cao, W.; Lee, S. H.; Lu, J., CD83 is preformed inside monocytes, macrophages and dendritic cells, but it is only stably expressed on activated dendritic cells. *Biochem. J.* **2005,** *385* (Pt 1), 85-93.

236. Hayashida, K.; Bartlett, A. H.; Chen, Y., et al., Molecular and cellular mechanisms of ectodomain shedding. *Anat Rec (Hoboken)* **2010,** *293* (6), 925-37.

237. Woo, C. H.; Lim, J. H.; Kim, J. H., Lipopolysaccharide induces matrix metalloproteinase-9 expression via a mitochondrial reactive oxygen species-p38 kinase-activator protein-1 pathway in Raw 264.7 cells. *J Immunol* **2004,** *173* (11), 6973-80.

238. Jung, I. H.; Choi, J. H.; Jin, J., et al., CD137-inducing factors from T cells and macrophages accelerate the destabilization of atherosclerotic plaques in hyperlipidemic mice. *FASEB J* **2014,** *28* (11), 4779-91.

239. Tedesco, S.; Bolego, C.; Toniolo, A., et al., Phenotypic activation and pharmacological outcomes of spontaneously differentiated human monocyte-derived macrophages. *Immunobiology* **2015,** *220* (5), 545-54.

240. Lawson, C.; Wolf, S., ICAM-1 signaling in endothelial cells. *Pharmacol Rep* **2009,** *61* (1), 22-32.

241. Jeon, K. I.; Xu, X.; Aizawa, T., et al., Vinpocetine inhibits NF-kappaB-dependent inflammation via an IKK-dependent but PDE-independent mechanism. *Proc Natl Acad Sci U S A* **2010,** *107* (21), 9795-800.

242. Sawa, Y.; Ueki, T.; Hata, M., et al., LPS-induced IL-6, IL-8, VCAM-1, and ICAM-1 expression in human lymphatic endothelium. *J Histochem Cytochem* **2008,** *56* (2), 97-109.

243. Takashiba, S.; Van Dyke, T. E.; Amar, S., et al., Differentiation of monocytes to macrophages primes cells for lipopolysaccharide stimulation via accumulation of cytoplasmic nuclear factor kappaB. *Infect Immun* **1999,** *67* (11), 5573-8.

244. Prieto, J.; Eklund, A.; Patarroyo, M., Regulated expression of integrins and other adhesion molecules during differentiation of monocytes into macrophages. *Cell Immunol* **1994,** *156* (1), 191-211.

245. Messaritou, G.; East, L.; Roghi, C., et al., Membrane type-1 matrix metalloproteinase activity is regulated by the endocytic collagen receptor Endo180. *J Cell Sci* **2009,** *122* (Pt 22), 4042-8.

246. Kramer, A.; Green, J.; Pollard, J., Jr., et al., Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **2014,** *30* (4), 523-30.

247. Yi, J. H.; Park, S. W.; Kapadia, R., et al., Role of transcription factors in mediating post-ischemic cerebral inflammation and brain damage. *Neurochem Int* **2007,** *50* (7-8), 1014-27.

248. Chiao, P. J.; Ling, J., Kras, Pten, NF-kappaB, and inflammation: dangerous liaisons. *Cancer Discov* **2011,** *1* (2), 103-5.

249. Portillo, J. A.; Greene, J. A.; Schwartz, I., et al., Blockade of CD40-TRAF2,3 or CD40-TRAF6 is sufficient to inhibit pro-inflammatory responses in non-haematopoietic cells. *Immunology* **2015,** *144* (1), 21-33.

250. Ramos, T. N.; Bullard, D. C.; Barnum, S. R., ICAM-1: isoforms and phenotypes. *J Immunol* **2014,** *192* (10), 4469-74.

251. Miller, J.; Knorr, R.; Ferrone, M., et al., Intercellular adhesion molecule-1 dimerization and its consequences for adhesion mediated by lymphocyte function associated-1. *J Exp Med* **1995,** *182* (5), 1231-41.

252. Jimenez, D.; Roda-Navarro, P.; Springer, T. A., et al., Contribution of N-linked glycans to the conformation and function of intercellular adhesion molecules (ICAMs). *J Biol Chem* **2005,** *280* (7), 5854-61.

253. Ngkelo, A.; Meja, K.; Yeadon, M., et al., LPS induced inflammatory responses in human peripheral blood mononuclear cells is mediated through NOX4 and Gialpha dependent PI-3kinase signalling. *J. Inflamm.* **2012,** *9* (1), 1.

254. Munford, R. S., Endotoxemia-menace, marker, or mistake? *J. Leukoc. Biol.* **2016,** *100* (4), 687-698.

255. Magdaleno, F.; Blajszczak, C. C.; Nieto, N., Key events participating in the pathogenesis of alcoholic liver disease. *Biomolecules* **2017,** *7* (1).

256. Magro, D. O.; Kotze, P. G.; Martinez, C. A. R., et al., Changes in serum levels of lipopolysaccharides and CD26 in patients with Crohn's disease. *Intest Res* **2017,** *15* (3), 352-357.

257. Pasternak, B. A.; D'Mello, S.; Jurickova, II, et al., Lipopolysaccharide exposure is linked to activation of the acute phase response and growth failure in pediatric Crohn's disease and murine colitis. *Inflamm. Bowel Dis.* **2010,** *16* (5), 856-69.

258. Pretorius, E.; Akeredolu, O. O.; Soma, P., et al., Major involvement of bacterial components in rheumatoid arthritis and its accompanying oxidative stress, systemic inflammation and hypercoagulability. *Exp. Biol. Med. (Maywood)* **2017,** *242* (4), 355-373.

259. Liang, H.; Hussey, S. E.; Sanchez-Avila, A., et al., Effect of lipopolysaccharide on inflammation and insulin action in human muscle. *PLoS One* **2013,** *8* (5), e63983.

260. Cani, P. D.; Amar, J.; Iglesias, M. A., et al., Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes* **2007,** *56* (7), 1761-72.

261. Liu, M.; Bing, G., Lipopolysaccharide animal models for Parkinson's disease. *Parkinsons Dis.* **2011,** *2011*, 327089.

262. Zhan, X.; Stamova, B.; Sharp, F. R., Lipopolysaccharide associates with amyloid plaques, neurons and oligodendrocytes in Alzheimer's disease brain: a review. *Front. Aging Neurosci.* **2018,** *10*, 42.

263. Kzhyshkowska, J.; Gudima, A.; Moganti, K., et al., Perspectives for monocyte/macrophage-based diagnostics of chronic inflammation. *Transfus. Med. Hemother.* **2016,** *43* (2), 66-77.

264. Gjelstrup, M. C.; Stilund, M.; Petersen, T., et al., Subsets of activated monocytes and markers of inflammation in incipient and progressed multiple sclerosis. *Immunol. Cell Biol.* **2018,** *96* (2), 160-174.

265. Yang, J.; Zhang, L.; Yu, C., et al., Monocyte and macrophage differentiation: circulation inflammatory monocyte as biomarker for inflammatory diseases. *Biomark. Res.* **2014,** *2* (1), 1.

266. Davignon, J. L.; Hayder, M.; Baron, M., et al., Targeting monocytes/macrophages in the treatment of rheumatoid arthritis. *Rheumatology (Oxford)* **2013,** *52* (4), 590-8.

267. Kalluri, R.; Weinberg, R. A., The basics of epithelial-mesenchymal transition. *J Clin Invest* **2009,** *119* (6), 1420-8.

268. Kim, D. H.; Xing, T.; Yang, Z., et al., Epithelial Mesenchymal Transition in Embryonic Development, Tissue Repair and Cancer: A Comprehensive Overview. *J Clin Med* **2017,** *7* (1).

269. Thiery, J. P.; Acloque, H.; Huang, R. Y., et al., Epithelial-mesenchymal transitions in development and disease. *Cell* **2009,** *139* (5), 871-90.

270. Haensel, D.; Dai, X., Epithelial-to-mesenchymal transition in cutaneous wound healing: Where we are and where we are heading. *Dev Dyn* **2018,** *247* (3), 473-480.

271. Stone, R. C.; Pastar, I.; Ojeh, N., et al., Epithelial-mesenchymal transition in tissue repair and fibrosis. *Cell Tissue Res* **2016,** *365* (3), 495-506.

272. Heerboth, S.; Housman, G.; Leary, M., et al., EMT and tumor metastasis. *Clin Transl Med* **2015,** *4*, 6.

273. Mittal, V., Epithelial Mesenchymal Transition in Tumor Metastasis. *Annu Rev Pathol* **2018,** *13*, 395-412.

274. Yang, J.; Antin, P.; Berx, G., et al., Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **2020,** *21* (6), 341-352.

275. Zeisberg, M.; Neilson, E. G., Biomarkers for epithelial-mesenchymal transitions. *J Clin Invest* **2009,** *119* (6), 1429-37.

276. Cook, D. P.; Vanderhyden, B. C., Context specificity of the EMT transcriptional response. *Nat Commun* **2020,** *11* (1), 2142.

277. Xu, J.; Lamouille, S.; Derynck, R., TGF-beta-induced epithelial to mesenchymal transition. *Cell Res* **2009,** *19* (2), 156-72.

278. Jolly, M. K.; Tripathi, S. C.; Jia, D., et al., Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget* **2016,** *7* (19), 27067-84.

279. Jordan, N. V.; Johnson, G. L.; Abell, A. N., Tracking the intermediate stages of epithelial-mesenchymal transition in epithelial stem cells and cancer. *Cell Cycle* **2011,** *10* (17), 2865-73.

280. Font-Clos, F.; Zapperi, S.; La Porta, C. A. M., Topography of epithelial-mesenchymal plasticity. *Proc Natl Acad Sci U S A* **2018,** *115* (23), 5902-5907.

281. Saitoh, M., Involvement of partial EMT in cancer progression. *J Biochem* **2018,** *164* (4), 257-264.

282. Zhang, J.; Tian, X. J.; Zhang, H., et al., TGF-beta-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Sci Signal* **2014,** *7* (345), ra91.

283. Aiello, N. M.; Maddipati, R.; Norgard, R. J., et al., EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. *Dev Cell* **2018,** *45* (6), 681-695 e4.

284. Riahi, R.; Gogoi, P.; Sepehri, S., et al., A novel microchannel-based device to capture and analyze circulating tumor cells (CTCs) of breast cancer. *Int J Oncol* **2014,** *44* (6), 1870-8.

285. Suresh, D.; Ghoshdastidar, S.; Gangula, A., et al., Magnetic Iron Nanocubes Effectively Capture Epithelial and Mesenchymal Cancer Cells. *ACS Omega* **2020,** *5* (37), 23724-23735.

286. Mani, S. A.; Guo, W.; Liao, M. J., et al., The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **2008,** *133* (4), 704-15.

287. Song, J.; Wang, W.; Wang, Y., et al., Epithelial-mesenchymal transition markers screened in a cell-based model and validated in lung adenocarcinoma. *BMC Cancer* **2019,** *19* (1), 680.

288. Melissaridou, S.; Wiechec, E.; Magan, M., et al., The effect of 2D and 3D cell cultures on treatment response, EMT profile and stem cell features in head and neck cancer. *Cancer Cell Int* **2019,** *19*, 16.

289. Konge, J.; Leteurtre, F.; Goislard, M., et al., Breast cancer stem cell-like cells generated during TGFbeta-induced EMT are radioresistant. *Oncotarget* **2018,** *9* (34), 23519-23531.

290. Camerlingo, R.; Miceli, R.; Marra, L., et al., Conditioned medium of primary lung cancer cells induces EMT in A549 lung cancer cell line by TGF-ss1 and miRNA21 cooperation. *PLoS One* **2019,** *14* (7), e0219597.

291. Lourenco, A. R.; Ban, Y.; Crowley, M. J., et al., Differential Contributions of Pre- and Post-EMT Tumor Cells in Breast Cancer Metastasis. *Cancer Res* **2020,** *80* (2), 163-169.

292. Puram, S. V.; Parikh, A. S.; Tirosh, I., Single cell RNA-seq highlights a role for a partial EMT in head and neck cancer. *Mol Cell Oncol* **2018,** *5* (3), e1448244.

293. Dong, J.; Hu, Y.; Fan, X., et al., Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. *Genome Biol* **2018,** *19* (1), 31.

294. Chae, Y. K.; Chang, S.; Ko, T., et al., Epithelial-mesenchymal transition (EMT) signature is inversely associated with T-cell infiltration in non-small cell lung cancer (NSCLC). *Sci Rep* **2018,** *8* (1), 2918.

295. Xu, R.; Won, J. Y.; Kim, C. H., et al., Roles of the Phosphorylation of Transcriptional Factors in Epithelial-Mesenchymal Transition. *J Oncol* **2019,** *2019*, 5810465.

296. Lucena, M. C.; Carvalho-Cruz, P.; Donadio, J. L., et al., Epithelial Mesenchymal Transition Induces Aberrant Glycosylation through Hexosamine Biosynthetic Pathway Activation. *J Biol Chem* **2016,** *291* (25), 12917-29.

297. Voutsadakis, I. A., Ubiquitination and the Ubiquitin-Proteasome System as regulators of transcription and transcription factors in epithelial mesenchymal transition of cancer. *Tumour Biol* **2012,** *33* (4), 897-910.

298. Mobley, R. J.; Abell, A. N., Controlling Epithelial to Mesenchymal Transition through Acetylation of Histone H2BK5. *J Nat Sci* **2017,** *3* (9).

299. Serrano-Gomez, S. J.; Maziveyi, M.; Alahari, S. K., Regulation of epithelial-mesenchymal transition through epigenetic and post-translational modifications. *Mol Cancer* **2016,** *15*, 18.

300. Suttapitugsakul, S.; Sun, F.; Wu, R., Recent Advances in Glycoproteomic Analysis by Mass Spectrometry. *Anal Chem* **2020,** *92* (1), 267-291.

301. Xiao, H.; Sun, F.; Suttapitugsakul, S., et al., Global and site-specific analysis of protein glycosylation in complex biological systems with Mass Spectrometry. *Mass Spectrom Rev* **2019,** *38* (4-5), 356-379.

302. Huang, D. W.; Sherman, B. T.; Tan, Q., et al., The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* **2007,** *8* (9), R183.

303. Kim, J. H.; Jang, Y. S.; Eom, K. S., et al., Transforming growth factor beta1 induces epithelial-to-mesenchymal transition of A549 cells. *J Korean Med Sci* **2007,** *22* (5), 898-904.

304. Kawata, M.; Koinuma, D.; Ogami, T., et al., TGF-beta-induced epithelial-mesenchymal transition of A549 lung adenocarcinoma cells is enhanced by pro-inflammatory cytokines derived from RAW 264.7 macrophage cells. *J Biochem* **2012,** *151* (2), 205-16.

305. Zhang, C.; Zhu, X.; Hua, Y., et al., YY1 mediates TGF-beta1-induced EMT and pro-fibrogenesis in alveolar epithelial cells. *Respir Res* **2019,** *20* (1), 249.

306. Kasai, H.; Allen, J. T.; Mason, R. M., et al., TGF-beta1 induces human alveolar epithelial to mesenchymal cell transition (EMT). *Respir Res* **2005,** *6*, 56.

307. Xue, Y.; Li, L.; Zhang, D., et al., Twisted epithelial-to-mesenchymal transition promotes progression of surviving bladder cancer T24 cells with hTERT-dysfunction. *PLoS One* **2011,** *6* (11), e27748.

308. Mikesh, L. M.; Kumar, M.; Erdag, G., et al., Evaluation of molecular markers of mesenchymal phenotype in melanoma. *Melanoma Res* **2010,** *20* (6), 485-95.

309. Chang, H.; Liu, Y.; Xue, M., et al., Synergistic action of master transcription factors controls epithelial-to-mesenchymal transition. *Nucleic Acids Res* **2016,** *44* (6), 2514-27.

310. Simone, T. M.; Higgins, C. E.; Czekay, R. P., et al., SERPINE1: A Molecular Switch in the Proliferation-Migration Dichotomy in Wound-"Activated" Keratinocytes. *Adv Wound Care (New Rochelle)* **2014,** *3* (3), 281-290.

311. Perez-Moreno, P.; Indo, S.; Niechi, I., et al., Endothelin-converting enzyme-1c promotes stem cell traits and aggressiveness in colorectal cancer cells. *Mol Oncol* **2020,** *14* (2), 347-362.

312. Du, J.; Hong, S.; Dong, L., et al., Dynamic Sialylation in Transforming Growth Factor-beta (TGF-beta)-induced Epithelial to Mesenchymal Transition. *J Biol Chem* **2015,** *290* (19), 12000-13.

313. Turner, M. D.; Nedjai, B.; Hurst, T., et al., Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease. *Biochim. Biophys. Acta* **2014,** *1843* (11), 2563-2582.

314. Wahl-Jensen, V.; Kurz, S. K.; Hazelton, P. R., et al., Role of Ebola virus secreted glycoproteins and virus-like particles in activation of human macrophages. *J. Virol.* **2005,** *79* (4), 2413-9.

315. Chandler, K.; Goldman, R., Glycoprotein disease markers and single protein-omics. *Mol. Cell. Proteomics* **2013,** *12* (4), 836-45.

316. Thaysen-Andersen, M.; Thogersen, I. B.; Lademann, U., et al., Investigating the biomarker potential of glycoproteins using comparative glycoprofiling - application to tissue inhibitor of metalloproteinases-1. *BBA-Proteins Proteomics* **2008,** *1784* (3), 455-463.

317. Song, E.; Mechref, Y., Defining glycoprotein cancer biomarkers by MS in conjunction with glycoprotein enrichment. *Biomark. Med.* **2015,** *9* (9), 835-844.

318. Kailemia, M. J.; Park, D.; Lebrilla, C. B., Glycans and glycoproteins as specific biomarkers for cancer. *Anal. Bioanal. Chem.* **2017,** *409* (2), 395-410.

319. Zhu, J. H.; Warner, E.; Parikh, N. D., et al., Glycoproteomic markers of hepatocellular carcinoma-mass spectrometry based approaches. *Mass Spectrom. Rev.* **2019,** *38* (3), 265-290.

320. Lilja, H.; Ulmert, D.; Vickers, A. J., Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. *Nat. Rev. Cancer* **2008,** *8* (4), 268-78.

321. Llop, E.; Ferrer-Batalle, M.; Barrabes, S., et al., Improvement of prostate cancer diagnosis by detecting PSA glycosylation-specific changes. *Theranostics* **2016,** *6* (8), 1190-204.

322. Hsiao, C. J.; Tzai, T. S.; Chen, C. H., et al., Analysis of urinary prostate-specific antigen glycoforms in samples of prostate cancer and benign prostate hyperplasia. *Dis. Markers* **2016,** *2016*, 8915809.

323. Popa, C.; Netea, M. G.; van Riel, P. L., et al., The role of TNF-alpha in chronic inflammatory conditions, intermediary metabolism, and cardiovascular risk. *J. Lipid Res.* **2007,** *48* (4), 751-62.

324. Terrando, N.; Monaco, C.; Ma, D., et al., Tumor necrosis factor-alpha triggers a cytokine cascade yielding postoperative cognitive decline. *Proc. Natl. Acad. Sci. USA* **2010,** *107* (47), 20518-22.

325. Shin, J.; Rhim, J.; Kwon, Y., et al., Comparative analysis of differentially secreted proteins in serum-free and serum-containing media by using BONCAT and pulsed SILAC. *Sci. Rep.* **2019,** *9* (1), 3096.

326. Boersema, P. J.; Geiger, T.; Wisniewski, J. R., et al., Quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples. *Molecular & Cellular Proteomics* **2013,** *12* (1), 158-171.

327. Yen, T. Y.; Haste, N.; Timpe, L. C., et al., Using a cell line breast cancer progression system to identify biomarker candidates. *J. Proteomics* **2014,** *96*, 173-183.

328. Yang, W.; Zhou, J. Y.; Chen, L., et al., Glycoproteomic analysis identifies human glycoproteins secreted from HIV latently infected T cells and reveals their presence in HIV+ plasma. *Clin. Proteom.* **2014,** *11*, 11.

329. Suttapitugsakul, S.; Sun, F. X.; Wu, R. H., Recent advances in glycoproteomic analysis by mass spectrometry. *Analytical Chemistry* **2020,** *92* (1), 267-291.

330. Bjorhall, K.; Miliotis, T.; Davidsson, P., Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics* **2005,** *5* (1), 307-17.

331. Witzke, K. E.; Rosowski, K.; Muller, C., et al., Quantitative secretome analysis of activated Jurkat cells using click chemistry-based enrichment of secreted glycoproteins. *J. Proteome Res.* **2017,** *16* (1), 137-146.

332. Liu, X. N.; Wang, S.; Yang, Q., et al., ESC reverses epithelial mesenchymal transition induced by transforming growth factor-beta via inhibition of Smad signal pathway in HepG2 liver cancer cells. *Cancer Cell Int.* **2015,** *15*, 114.

333. Shajahan, A.; Supekar, N. T.; Wu, H., et al., Mass spectrometric method for the unambiguous profiling of cellular dynamic glycosylation. *ACS Chem. Biol.* **2020,** *15* (10), 2692-2701.

334. Milner, C. M.; Day, A. J., TSG-6: a multifunctional protein associated with inflammation. *J. Cell Sci.* **2003,** *116* (Pt 10), 1863-73.

335. Mittal, M.; Tiruppathi, C.; Nepal, S., et al., TNFalpha-stimulated gene-6 (TSG6) activates macrophage phenotype transition to prevent inflammatory lung injury. *Proc. Natl. Acad. Sci. USA* **2016,** *113* (50), E8151-E8158.

336. Benigni, A.; Cassis, P.; Remuzzi, G., Angiotensin II revisited: new roles in inflammation, immunology and aging. *EMBO Mol. Med.* **2010,** *2* (7), 247-57.

337. Peng, J. C.; Abu Bakar, S.; Richardson, M. M., et al., IL10 and IL12B polymorphisms each influence IL-12p70 secretion by dendritic cells in response to LPS. *Immunol. Cell Biol.* **2006,** *84* (2), 227-32.

338. Zhao, Q.; Du, Q.; Wei, F., et al., An infectious disease-associated il12b polymorphism regulates IL-12/23 p40 transcription involving poly(ADP-ribose) polymerase 1. *J. Immunol.* **2017,** *198* (7), 2935-2942.

339. Frank, P. G.; Lisanti, M. P., ICAM-1: role in inflammation and in the regulation of vascular permeability. *Am. J. Physiol. Heart Circ. Physiol.* **2008,** *295* (3), H926-H927.

340. Videm, V.; Albrigtsen, M., Soluble ICAM-1 and VCAM-1 as markers of endothelial activation. *Scand. J. Immunol.* **2008,** *67* (5), 523-31.

341. Starr, T.; Bauler, T. J.; Malik-Kale, P., et al., The phorbol 12-myristate-13-acetate differentiation protocol is critical to the interaction of THP-1 macrophages with Salmonella Typhimurium. *PLoS One* **2018,** *13* (3), e0193601.

342. Horvatinovich, J. M.; Grogan, E. W.; Norris, M., et al., Soluble CD83 inhibits T cell activation by binding to the TLR4/MD-2 complex on CD14(+) monocytes. *J. Immunol.* **2017,** *198* (6), 2286-2301.

343. Lin, X. L.; Liu, M.; Liu, Y., et al., Transforming growth factor beta1 promotes migration and invasion in HepG2 cells: Epithelialtomesenchymal transition via JAK/STAT3 signaling. *Int. J. Mol. Med.* **2018,** *41* (1), 129-136.

344. Enomoto, Y.; Matsushima, S.; Shibata, K., et al., LTBP2 is secreted from lung myofibroblasts and is a potential biomarker for idiopathic pulmonary fibrosis. *Clin. Sci.* **2018,** *132* (14), 1565-1580.

345. Park, S. A.; Kim, M. J.; Park, S. Y., et al., TIMP-1 mediates TGF-beta-dependent crosstalk between hepatic stellate and cancer cells via FAK signaling. *Sci. Rep.* **2015,** *5*, 16492.

346. Busnadiego, O.; Gonzalez-Santamaria, J.; Lagares, D., et al., LOXL4 is induced by transforming growth factor beta1 through Smad and JunB/Fra2 and contributes to vascular matrix remodeling. *Mol. Cell. Biol.* **2013,** *33* (12), 2388-401.

347. Maris, J. M., Recent advances in neuroblastoma. *N Engl J Med* **2010,** *362* (23), 2202-11.

348. Matthay, K. K.; Maris, J. M.; Schleiermacher, G., et al., Neuroblastoma. *Nat Rev Dis Primers* **2016,** *2*, 16078.

349. The American Cancer Society medical and editorial content team. Key Statistics About Neuroblastoma,. https://www.cancer.org/cancer/neuroblastoma/about/key-statistics.html (accessed 01/16/2019).

350. Braekeveldt, N.; von Stedingk, K.; Fransson, S., et al., Patient-Derived Xenograft Models Reveal Intratumor Heterogeneity and Temporal Stability in Neuroblastoma. *Cancer Res* **2018,** *78* (20), 5958-5969.

351. Boeva, V.; Louis-Brennetot, C.; Peltier, A., et al., Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat Genet* **2017,** *49* (9), 1408-1413.

352. Borriello, L.; Seeger, R. C.; Asgharzadeh, S., et al., More than the genes, the tumor microenvironment in neuroblastoma. *Cancer Lett* **2016,** *380* (1), 304-14.

353. Pistoia, V.; Morandi, F.; Bianchi, G., et al., Immunosuppressive microenvironment in neuroblastoma. *Front Oncol* **2013,** *3*, 167.

354. Garner, E. F.; Beierle, E. A., Cancer Stem Cells and Their Interaction with the Tumor Microenvironment in Neuroblastoma. *Cancers (Basel)* **2015,** *8* (1).

355. DuBois, S. G.; Kalika, Y.; Lukens, J. N., et al., Metastatic sites in stage IV and IVS neuroblastoma correlate with age, tumor biology, and survival. *J Pediatr Hematol Oncol* **1999,** *21* (3), 181-9.

356. Chen, Y.; Takita, J.; Choi, Y. L., et al., Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* **2008,** *455* (7215), 971-4.

357. Huang, M.; Weiss, W. A., Neuroblastoma and MYCN. *Cold Spring Harb Perspect Med* **2013,** *3* (10), a014415.

358. Cook, M. N.; Olshan, A. F.; Guess, H. A., et al., Maternal medication use and neuroblastoma in offspring. *Am J Epidemiol* **2004,** *159* (8), 721-31.

359. McCall, E. E.; Olshan, A. F.; Daniels, J. L., Maternal hair dye use and risk of neuroblastoma in offspring. *Cancer Causes Control* **2005,** *16* (6), 743-8.

360. Daniels, J. L.; Olshan, A. F.; Teschke, K., et al., Residential pesticide exposure and neuroblastoma. *Epidemiology* **2001,** *12* (1), 20-7.

361. Cohn, S. L.; Pearson, A. D.; London, W. B., et al., The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report. *J Clin Oncol* **2009,** *27* (2), 289-97.

362. The American Cancer Society medical and editorial content team. Neuroblastoma Survival Rates by Risk Group. https://www.cancer.org/cancer/neuroblastoma/detection-diagnosis-staging/survival-rates.html (accessed 01/16/2019).

363. Esposito, M. R.; Aveic, S.; Seydel, A., et al., Neuroblastoma treatment in the post-genomic era. *J Biomed Sci* **2017,** *24* (1), 14.

364. Michaelis, M.; Selt, F.; Rothweiler, F., et al., Aurora kinases as targets in drug-resistant neuroblastoma cells. *PLoS One* **2014,** *9* (9), e108758.

365. London, W. B.; Frantz, C. N.; Campbell, L. A., et al., Phase II randomized comparison of topotecan plus cyclophosphamide versus topotecan alone in children with recurrent or refractory neuroblastoma: a Children's Oncology Group study. *J Clin Oncol* **2010,** *28* (24), 3808-15.

366. Pommier, Y., Topoisomerase I inhibitors: camptothecins and beyond. *Nat Rev Cancer* **2006,** *6* (10), 789-802.

367. Emadi, A.; Jones, R. J.; Brodsky, R. A., Cyclophosphamide and cancer: golden anniversary. *Nat Rev Clin Oncol* **2009,** *6* (11), 638-47.

368. Emens, L. A.; Middleton, G., The interplay of immunotherapy and chemotherapy: harnessing potential synergies. *Cancer Immunol Res* **2015,** *3* (5), 436-43.

369. Abu Eid, R.; Razavi, G. S.; Mkrtichyan, M., et al., Old-School Chemotherapy in Immunotherapeutic Combination in Cancer, A Low-cost Drug Repurposed. *Cancer Immunol Res* **2016,** *4* (5), 377-82.

370. Duan, C.; Wang, H.; Chen, Y., et al., Whole exome sequencing reveals novel somatic alterations in neuroblastoma patients with chemotherapy. *Cancer Cell Int* **2018,** *18*, 21.

371. Theruvath, J.; Russo, A.; Kron, B., et al., Next-generation sequencing reveals germline mutations in an infant with synchronous occurrence of nephro- and neuroblastoma. *Pediatr Hematol Oncol* **2016,** *33* (4), 264-75.

372. Imperiale, A.; Elbayed, K.; Moussallieh, F. M., et al., Metabolomic pattern of childhood neuroblastoma obtained by (1)H-high-resolution magic angle spinning (HRMAS) NMR spectroscopy. *Pediatr Blood Cancer* **2011,** *56* (1), 24-34.

373. Egler, R. A.; Li, Y.; Dang, T. A., et al., An integrated proteomic approach to identifying circulating biomarkers in high-risk neuroblastoma and their potential in relapse monitoring. *Proteomics Clin Appl* **2011,** *5* (9-10), 532-41.

374. Kim, P. Y.; Tan, O.; Diakiw, S. M., et al., Identification of plasma complement C3 as a potential biomarker for neuroblastoma using a quantitative proteomic approach. *J Proteomics* **2014,** *96*, 1-12.

375. Harris, W.; Sachana, M.; Flaskos, J., et al., Proteomic analysis of differentiating neuroblastoma cells treated with sub-lethal neurite inhibitory concentrations of diazinon: identification of novel biomarkers of effect. *Toxicol Appl Pharmacol* **2009,** *240* (2), 159-65.

376. Werner, T.; Sweetman, G.; Savitski, M. F., et al., Ion coalescence of neutron encoded TMT 10-plex reporter ions. *Anal Chem* **2014,** *86* (7), 3594-601.

377. McAlister, G. C.; Huttlin, E. L.; Haas, W., et al., Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal Chem* **2012,** *84* (17), 7469-78.

378. Huang da, W.; Sherman, B. T.; Lempicki, R. A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **2009,** *4* (1), 44-57.

379. Huang da, W.; Sherman, B. T.; Lempicki, R. A., Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **2009,** *37* (1), 1-13.

380. Bailey, T. L.; Boden, M.; Buske, F. A., et al., MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **2009,** *37* (Web Server issue), W202-8.

381. Schwartz, D.; Gygi, S. P., An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* **2005,** *23* (11), 1391-8.

382. Cao, Q.; Chen, X.; Wu, X., et al., Inhibition of UGT8 suppresses basal-like breast cancer progression by attenuating sulfatide-alphaVbeta5 axis. *J Exp Med* **2018,** *215* (6), 1679-1692.

383. Bauden, M.; Kristl, T.; Sasor, A., et al., Histone profiling reveals the H1.3 histone variant as a prognostic biomarker for pancreatic ductal adenocarcinoma. *BMC Cancer* **2017,** *17* (1), 810.

384. Sprowl-Tanio, S.; Habowski, A. N.; Pate, K. T., et al., Lactate/pyruvate transporter MCT-1 is a direct Wnt target that confers sensitivity to 3-bromopyruvate in colon cancer. *Cancer Metab* **2016,** *4*, 20.

385. Krishnan, C.; Higgins, J. P.; West, R. B., et al., Microtubule-associated protein-2 is a sensitive marker of primary and metastatic neuroblastoma. *Am J Surg Pathol* **2009,** *33* (11), 1695-704.

386. Haber, M.; Smith, J.; Bordow, S. B., et al., Association of high-level MRP1 expression with poor clinical outcome in a large prospective study of primary neuroblastoma. *J Clin Oncol* **2006,** *24* (10), 1546-53.

387. Tang, X. X.; Robinson, M. E.; Riceberg, J. S., et al., Favorable neuroblastoma genes and molecular therapeutics of neuroblastoma. *Clin Cancer Res* **2004,** *10* (17), 5837-44.

388. Riley, R. D.; Heney, D.; Jones, D. R., et al., A systematic review of molecular and biological tumor markers in neuroblastoma. *Clin Cancer Res* **2004,** *10* (1 Pt 1), 4-12.

389. Xie, Y.; Xu, H.; Fang, F., et al., A 3-Protein Expression Signature of Neuroblastoma for Outcome Prediction. *Am J Surg Pathol* **2018,** *42* (8), 1027-1035.

390. Desai, S. D.; Zhang, H.; Rodriguez-Bauman, A., et al., Transcription-dependent degradation of topoisomerase I-DNA covalent complexes. *Mol Cell Biol* **2003,** *23* (7), 2341-50.

391. Pfister, T. D.; Reinhold, W. C.; Agama, K., et al., Topoisomerase I levels in the NCI-60 cancer cell line panel determined by validated ELISA and microarray analysis and correlation with indenoisoquinoline sensitivity. *Mol Cancer Ther* **2009,** *8* (7), 1878-84.

392. Whitacre, C. M.; Zborowska, E.; Gordon, N. H., et al., Topotecan increases topoisomerase IIalpha levels and sensitivity to treatment with etoposide in schedule-dependent process. *Cancer Res* **1997,** *57* (8), 1425-8.

393. Ogu, C. C.; Maxa, J. L., Drug interactions due to cytochrome P450. *Proc (Bayl Univ Med Cent)* **2000,** *13* (4), 421-3.

394. Tomita, H.; Tanaka, K.; Tanaka, T., et al., Aldehyde dehydrogenase 1A1 in stem cells and cancer. *Oncotarget* **2016,** *7* (10), 11018-32.

395. van der Waals, L. M.; Borel Rinkes, I. H. M.; Kranenburg, O., ALDH1A1 expression is associated with poor differentiation, 'right-sidedness' and poor survival in human colorectal cancer. *PLoS One* **2018,** *13* (10), e0205536.

396. Ohneseit, P. A.; Prager, D.; Kehlbach, R., et al., Cell cycle effects of topotecan alone and in combination with irradiation. *Radiother Oncol* **2005,** *75* (2), 237-45.

397. Rogakou, E. P.; Pilch, D. R.; Orr, A. H., et al., DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* **1998,** *273* (10), 5858-68.

398. Branzei, D.; Foiani, M., Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol* **2008,** *9* (4), 297-308.

399. Podhorecka, M.; Skladanowski, A.; Bozko, P., H2AX Phosphorylation: Its Role in DNA Damage Response and Cancer Therapy. *J Nucleic Acids* **2010,** *2010*.

400. Matsuoka, S.; Ballif, B. A.; Smogorzewska, A., et al., ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **2007,** *316* (5828), 1160-6.

401. Peng, Z.; Liao, Z.; Dziegielewska, B., et al., Phosphorylation of serine 51 regulates the interaction of human DNA ligase I with replication factor C and its participation in DNA replication and repair. *J Biol Chem* **2012,** *287* (44), 36711-9.

402. Lasek, A. L.; McPherson, B. M.; Trueman, N. G., et al., The Functional Significance of Posttranslational Modifications on Polo-Like Kinase 1 Revealed by Chemical Genetic Complementation. *PLoS One* **2016,** *11* (2), e0150225.

403. Kanehisa, M.; Sato, Y.; Kawashima, M., et al., KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **2016,** *44* (D1), D457-62.

404. Liu, P.; Cheng, H.; Roberts, T. M., et al., Targeting the phosphoinositide 3-kinase pathway in cancer. *Nat Rev Drug Discov* **2009,** *8* (8), 627-44.

405. Li, Y. Y.; Popivanova, B. K.; Nagai, Y., et al., Pim-3, a proto-oncogene with serine/threonine kinase activity, is aberrantly expressed in human pancreatic cancer and phosphorylates bad to block bad-mediated apoptosis in human pancreatic cancer cell lines. *Cancer Res* **2006,** *66* (13), 6741-7.

406. Villen, J.; Beausoleil, S. A.; Gerber, S. A., et al., Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci U S A* **2007,** *104* (5), 1488-93.

407. Wang, Y. Y.; Zhao, R.; Zhe, H., The emerging role of CaMKII in cancer. *Oncotarget* **2015,** *6* (14), 11725-34.

408. Lucic, V.; Greif, G. J.; Kennedy, M. B., Detailed state model of CaMKII activation and autophosphorylation. *Eur Biophys J* **2008,** *38* (1), 83-98.

409. Hafner, J.; Mayr, M. I.; Mockel, M. M., et al., Pre-anaphase chromosome oscillations are regulated by the antagonistic activities of Cdk1 and PP1 on Kif18A. *Nat Commun* **2014,** *5*, 4397.

410. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** *422* (6928), 198-207.

411. Baldwin, M. A., Protein identification by mass spectrometry: issues to be considered. *Mol. Cell. Proteomics* **2004,** *3* (1), 1-9.

412. Cottrell, J. S., Protein identification using MS/MS data. *J. Proteomics* **2011,** *74* (10), 1842-51.

413. Washburn, M. P.; Wolters, D.; Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001,** *19* (3), 242-247.

414. Keshishian, H.; Addona, T.; Burgess, M., et al., Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* **2007,** *6* (12), 2212-2229.

415. Moradian, A.; Kalli, A.; Sweredoski, M. J., et al., The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics* **2014,** *14* (4-5), 489-97.

416. Resing, K. A.; Ahn, N. G., Proteomics strategies for protein identification. *FEBS Lett.* **2005,** *579* (4), 885-9.

417. Catherman, A. D.; Skinner, O. S.; Kelleher, N. L., Top Down proteomics: facts and perspectives. *Biochem. Biophys. Res. Commun.* **2014,** *445* (4), 683-93.

418. Heck, A. J. R., Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* **2008,** *5* (11), 927-933.

419. Li, H. L.; Sheng, Y. W.; McGee, W., et al., Structural Characterization of Native Proteins and Protein Complexes by Electron Ionization Dissociation-Mass Spectrometry. *Anal. Chem.* **2017,** *89* (5), 2731-2738.

420. Cai, W. X.; Tucholski, T.; Chen, B. F., et al., Top-Down Proteomics of Large Proteins up to 223 kDa Enabled by Serial Size Exclusion Chromatography Strategy. *Anal. Chem.* **2017,** *89* (10), 5467-5475.

421. Mann, M., Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell. Biol.* **2006,** *7* (12), 952-8.

422. Ong, S. E.; Blagoev, B.; Kratchmarova, I., et al., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002,** *1* (5), 376-86.

423. Feist, P. E.; Sidoli, S.; Liu, X., et al., Multicellular Tumor Spheroids Combined with Mass Spectrometric Histone Analysis To Evaluate Epigenetic Drugs. *Anal. Chem.* **2017,** *89* (5), 2773-2781.

424. Wang, X. R.; Cimermancic, P.; Yu, C., et al., Molecular Details Underlying Dynamic Structures and Regulation of the Human 26S Proteasome. *Mol. Cell. Proteomics* **2017,** *16* (5), 840-854.

425. Aebersold, R.; Goodlett, D. R., Mass spectrometry in proteomics. *Chem. Rev.* **2001,** *101* (2), 269-295.

426. Sun, L. L.; Hebert, A. S.; Yan, X. J., et al., Over 10000 Peptide Identifications from the HeLa Proteome by Using Single-Shot Capillary Zone Electrophoresis Combined with Tandem Mass Spectrometry. *Angew. Chem.-Int. Edit.* **2014,** *53* (50), 13931-13933.

427. Miah, S.; Banks, C. A. S.; Adams, M. K., et al., Advancement of mass spectrometry-based proteomics technologies to explore triple negative breast cancer. *Mol. Biosyst.* **2017,** *13* (1), 42-55.

428. Su, D.; Gaffrey, M. J.; Guo, J., et al., Proteomic identification and quantification of S-glutathionylation in mouse macrophages using resin-assisted enrichment and isobaric labeling. *Free Radic. Biol. Med.* **2014,** *67*, 460-70.

429. McLachlin, D. T.; Chait, B. T., Analysis of phosphorylated proteins and peptides by mass spectrometry. *Curr. Opin. Chem. Biol.* **2001,** *5* (5), 591-602.

430. Xiao, H. P.; Wu, R. H., Quantitative investigation of human cell surface N-glycoprotein dynamics. *Chem. Sci.* **2017,** *8* (1), 268-277.

431. Chen, Y.; Sprung, R.; Tang, Y., et al., Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol. Cell. Proteomics* **2007,** *6* (5), 812-819.

432. Shi, T. J.; Gao, Y. Q.; Gaffrey, M. J., et al., Sensitive Targeted Quantification of ERK Phosphorylation Dynamics and Stoichiometry in Human Cells without Affinity Enrichment. *Anal. Chem.* **2015,** *87* (2), 1103-1110.

433. Smits, A. H.; Vermeulen, M., Characterizing Protein-Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. *Trends Biotechnol.* **2016,** *34* (10), 825-34.

434. Abu-Farha, M.; Elisma, F.; Figeys, D., Identification of protein-protein interactions by mass spectrometry coupled techniques. *Adv. Biochem. Eng. Biotechnol.* **2008,** *110*, 67-80.

435. Vasilescu, J.; Guo, X.; Kast, J., Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. *Proteomics* **2004,** *4* (12), 3845-54.

436. Xiao, Y. S.; Guo, L.; Wang, Y. S., Isotope-Coded ATP Probe for Quantitative Affinity Profiling of ATP-Binding Proteins. *Anal. Chem.* **2013,** *85* (15), 7478-7486.

437. Bulaj, G., Formation of disulfide bonds in proteins and peptides. *Biotechnol. Adv.* **2005,** *23* (1), 87-92.

438. Woycechowsky, K. J.; Raines, R. T., Native disulfide bond formation in proteins. *Curr. Opin. Chem. Biol.* **2000,** *4* (5), 533-9.

439. Feige, M. J.; Hendershot, L. M., Disulfide bonds in ER protein folding and homeostasis. *Curr. Opin. Cell. Biol.* **2011,** *23* (2), 167-75.

440. Gundry, R. L.; White, M. Y.; Murray, C. I., et al., Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Curr. Protoc. Mol. Biol.* **2009,** *Chapter 10*, Unit10 25.

441. Zhang, Y.; Fonslow, B. R.; Shan, B., et al., Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* **2013,** *113* (4), 2343-94.

442. Sechi, S.; Chait, B. T., Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Ana.l Chem.* **1998,** *70* (24), 5150-8.

443. Singh, R.; Lamoureux, G. V.; Lees, W. J., et al., Reagents for Rapid Reduction of Disulfide Bonds. *Method Enzymol* **1995,** *251*, 167-173.

444. Burns, J. A.; Butler, J. C.; Moran, J., et al., Selective Reduction of Disulfides by Tris(2-Carboxyethyl)Phosphine. *J. Org. Chem.* **1991,** *56* (8), 2648-2650.

445. Paulech, J.; Solis, N.; Cordwell, S. J., Characterization of reaction conditions providing rapid and specific cysteine alkylation for peptide-based mass spectrometry. *Bba-Proteins Proteom* **2013,** *1834* (1), 372-379.

446. Hill, B. G.; Reily, C.; Oh, J. Y., et al., Methods for the determination and quantification of the reactive thiol proteome. *Free Radic. Biol. Med.* **2009,** *47* (6), 675-83.

447. Righetti, P. G., Real and imaginary artefacts in proteome analysis via two-dimensional maps. *J Chromatogr B Analyt Technol Biomed Life Sci* **2006,** *841* (1-2), 14-22.

448. Crankshaw, M. W.; Grant, G. A., Modification of cysteine. *Curr. Protoc. Protein Sci.* **2001,** *Chapter 15*, Unit15 1.

449. Zoeller, R. T.; Fletcher, D. L.; Butnariu, O., et al., N-ethylmaleimide (NEM) can significantly improve in situ hybridization results using 35S-labeled oligodeoxynucleotide or complementary RNA probes. *J. Histochem. Cytochem.* **1997,** *45* (7), 1035-41.

450. Boja, E. S.; Fales, H. M., Overalkylation of a protein digest with iodoacetamide. *Ana.l Chem.* **2001,** *73* (15), 3576-82.

451. Guo, M. Z.; Weng, G. F.; Yin, D. Y., et al., Identification of the over alkylation sites of a protein by IAM in MALDI-TOF/TOF tandem mass spectrometry. *Rsc Adv.* **2015,** *5* (125), 103662-103668.

452. Muller, T.; Winter, D., Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents. *Mol. Cell. Proteomics* **2017**.

453. Guo, J.; Gaffrey, M. J.; Su, D., et al., Resin-assisted enrichment of thiols as a general strategy for proteomic profiling of cysteine-based reversible modifications. *Nat. Protoc.* **2014,** *9* (1), 64-75.

454. Chen, W. X.; Smeekens, J. M.; Wu, R. H., A Universal Chemical Enrichment Method for Mapping the Yeast N-glycoproteome by Mass Spectrometry (MS). *Mol. Cell. Proteomics* **2014,** *13* (6), 1563-1572.

455. Xiao, H. P.; Tang, G. X.; Wu, R. H., Site-Specific Quantification of Surface N-Glycoproteins in Statin-Treated Liver Cells. *Anal. Chem.* **2016,** *88* (6), 3324-3332.

456. Peng, J. M.; Elias, J. E.; Thoreen, C. C., et al., Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome. *J. Proteome Res.* **2003,** *2* (1), 43-50.

457. Huttlin, E. L.; Jedrychowski, M. P.; Elias, J. E., et al., A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. *Cell* **2010,** *143* (7), 1174-1189.

458. Getz, E. B.; Xiao, M.; Chakrabarty, T., et al., A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal Biochem* **1999,** *273* (1), 73-80.

459. Stevens, R.; Stevens, L.; Price, N. C., The Stabilities of Various Thiol Compounds Used in Protein Purifications. *Biochem Educ* **1983,** *11* (2), 70-70.

460. McNulty, J.; Krishnamoorthy, V.; Amoroso, D., et al., Tris(3-hydroxypropyl) phosphine (THPP): A mild, air-stable reagent for the rapid, reductive cleavage of small-molecule disulfides. *Bioorg Med Chem Lett* **2015,** *25* (19), 4114-4117.

461. Fruchter, R. G.; Crestfield, A. M., The specific alkylation by iodoacetamide of histidine-12 in the active site of ribonuclease. *J. Biol. Chem.* **1967,** *242* (23), 5807-12.

462. Yang, Z.; Attygalle, A. B., LC/MS characterization of undesired products formed during iodoacetamide derivatization of sulfhydryl groups of peptides. *J. Mass Spectrom.* **2007,** *42* (2), 233-43.

463. Paulsen, C. E.; Carroll, K. S., Cysteine-mediated redox signaling: chemistry, biology, and tools for discovery. *Chem Rev* **2013,** *113* (7), 4633-79.

464. Huang, D. W.; Sherman, B. T.; Lempicki, R. A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **2009,** *4* (1), 44-57.