INTEGRATED STRATEGIES TO DEVELOP POST-TRANSLATIONALLY MODIFIED PROTEINS IN EXTRACELLULAR VESICLES AS CANDIDATE DISEASE MARKERS

by

Hillary Andaluz Aguilar

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Chemistry

West Lafayette, Indiana

December 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. W. Andy Tao, Chair

Department of Biochemistry

Dr. Hilkka Kenttämaa

Department of Chemistry

Dr. Michael Wendt

Department of Medicinal Chemistry and Molecular Pharmacology

Dr. Chang-Deng Hu

Department of Medicinal Chemistry and Molecular Pharmacology

Approved by:

Dr. Christine Hrycyna

Con mucho amor dedicado a mi familia, en especial a mi madre Lérida Aguilar, mi motivación, mi máximo amor y motor en todo momento. A mi padre Manuel Perdomo, mi apoyo y sustento. A mi hermano Himmler Andaluz, mi máxima fuente de risas y momentos inolvidables. A mis abuelos, en especial a mi abuelo Rafael Aguilar, mi modelo a seguir. A mi abuela Lérida, quién desde arriba me sigue inspirando a luchar en su nombre. A mi pareja Carlos Peláez, mi amor sin condiciones. A mi hermana de la vida Cynthia Alvarado, mi amistad más pura e incondicional.

Gracias a Dios y a mi Virgen por siempre guiarme.

Esto es por y para ustedes.

ACKNOWLEDGMENTS

I would first like to thank my advisor, Dr. W. Andy Tao, for his support and guidance throughout this PhD journey, especially in helping me with projects I am passionate and highly motivated about. I have grown so much personally and professionally in his lab. Throughout my journey, I was able to not only think as a scientist but also developed critical skills such as communication, teamwork, problem solving and leadership. I was also able to foster great collaborations among the science and medical community that taught me so much I will forever be grateful. Specifically, I would like to thank Dr. Roberto Pili and Dr. Ronald Boris for giving me the opportunity to start a project in the medical field and learn from them as oncologists. From developing the project ideas to being in the surgery room with them, I will forever appreciate those experiences. I would also like to thank my amazing committee members: Dr. Hilkka Kenttämaa, Dr. Michael Wendt and Dr. Chang-Deng Hu for their guidance through the years.

Without a doubt, I also need to thank the great and wonderful lab mates I have had these five years: without you this would not be possible. I appreciate the mentorship and friendship of Dr. Blair Chen who was there to train me and as a colleague for advice, ideas, and projects. I also appreciate the mentorship from Dr. Anton Iliuk, Dr. Chuan-Chih Hsu, and Dr. Li Li; the friendship, advice, and long discussions with Sebastian Páez and Peipei Zhu, and finally the friendship and help from Leo Kao, Xiaofeng Wu, and Marco Hadisurya. My relationship with my lab mates has strengthened my understanding of projects, challenged my critical thinking and communication, and most importantly, made my time in lab enjoyable and unforgettable.

Most importantly, I would like to thank my loving and supporting family, especially my mom Lérida Aguilar who has been my main motivation. Also, to my dad Manuel Perdomo for his support and warm conversations, and my brother Himmler Andaluz who has been there for me through thick and thin. I will be forever grateful. I would be remiss if I failed to thank one of my main supports throughout these five years, a person who has my back no matter what and the true definition of a friend: Cynthia Alvarado, who I consider family. Last but not least, thank you to my wonderful partner Carlos Peláez, for always being supportive, caring and loving. Thank you all for being my biggest fans all the time. I love you all and I truly cannot wait to see what the future holds.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EVs	Extracellular Vesicles
LC-MS	Liquid chromatography-mass spectrometry
MVs	Microvesicles
PolyMAC	Polymer-based metal-ion affinity capture
PTMs	Post-translational modifications

ABSTRACT

Extracellular vesicles (EVs) are membrane-enclosed nanoparticles containing proteins and nucleic acid cargo. These vesicles are released by almost all cell types and provide an effective and ubiquitous path for intercellular communication and transmission of pathogenic and signaling molecules among cells. Research into potential biomarkers isolated from EV has been propelled by the development of methods and tools to acquire them by minimally and non-invasive means, which reinforces their great diagnostic potential. In the context of cancer, this opens the door to apply EV based liquid biopsy for early detection prior to alternate, more prevailing diagnostic tools like imaging studies. In autoimmune diseases, EVs play a crucial role in immune responses and as immunomodulatory agents as they can modulate the function of a wide variety of immune cells, especially in antigen-presenting cells (APCs). Several efforts have been made to study EVs and their cargo in numerous disease models, but very few in autoimmunity. Autoimmune diseases are chronic, have been underexplored especially in the omics area, and their diagnosis and treatment rely on traditional therapy. Therefore, there is a need for efficient methods to elucidate biomarkers that could provide additional layers of information for treatment, diagnosis, and prognosis. Additionally, protein post-translational modifications (PTMs), such as phosphorylation, glycosylation, and acetylation, are involved in multiple essential cellular processes and represent an important mechanism of regulation for cellular physiological functions, leading to the development of effective and targeted therapeutics. Discovery and profiling PTMs have established the relevance of PTMs in EVs and associated EV functions and novel applications.

This dissertation proposes integrated proteomic strategies to efficiently isolate and analyze EVs in human plasma from different types of pathologies like cancer and autoimmune diseases. The main focus is the development of the platforms, to not only isolate the proteome from EVs, but also PTMs including phosphorylation, glycosylation and acetylation, simultaneously. Chapter one, which is the core of this dissertation, describes the platform to sequentially isolate and analyze the EV proteome, phosphoproteome and glycoproteome from human plasma. Chapters two and three focus on the ongoing application of this platform with slight modifications into different disease models, in this case breast cancer subtypes and autoimmune diseases.

CHAPTER 1. SEQUENTIAL PHOSPHOPROTEOMICS AND N-GLYCOPROTEOMICS OF PLASMA-DERIVED EXTRACELLULLAR VESICLES

A version of this chapter was previously published by *Nature Protocols* **15**, 61-180. Andaluz Aguilar, H., Iliuk, A.B., Chen, I.H. & Tao, W.A. (2020). Sequential phosphoproteomics and N-glycoproteomics of plasma-derived extracellular vesicles. doi.org/10.1038/s41596-019-0260-5.

1.1 Summary

Extracellular vesicles (EVs) are increasingly being recognized as important vehicles for intercellular communication and as promising sources for biomarker discovery. Because the state of protein post-translational modifications (PTMs) such as phosphorylation and glycosylation can be a key determinant of cellular physiology, comprehensive characterization of protein PTMs in EVs can be particularly valuable for early-stage diagnostics and monitoring of disease status. However, the analysis of PTMs in EVs has been complicated by limited amounts of purified EVs, low-abundance PTM proteins, and interference from proteins and metabolites in biofluids. Recently, we developed an approach to isolate phosphoproteins and glycoproteins in EVs from small volumes of human plasma that enabled us to identify nearly 10,000 unique phosphopeptides and 1,500 unique N-glycopeptides. The approach demonstrated the feasibility of using these data to identify potential markers to differentiate disease from healthy states. Here we present an updated workflow to sequentially isolate phosphopeptides and N-glycopeptides, enabling multiple PTM analyses of the same clinical samples. In this updated workflow, we have improved the reproducibility and efficiency of EV isolation, protein extraction, and phosphopeptide/N-glycopeptide enrichment to achieve sensitive analyses of low-abundance PTMs in EVs isolated from 1 mL of plasma. The modularity of the workflow also allows for the characterization of phospho- or glycopeptides only and enables additional analysis of total proteomes and other PTMs of interest. After blood collection, the protocol takes 2 d, including EV isolation, PTM/peptide enrichment, mass spectrometry analysis, and data quantification.

1.2 Introduction

Extracellular vesicles, generally including exosomes and microvesicles[1-3], are membrane-enclosed vesicles containing proteins and nucleic acid cargos. EVs are released by almost all cell types and provide an effective and ubiquitous path for intercellular communication and the transmission of pathogenic and signaling molecules among cells[3-5]. Their potentially important cellular functions in disease onset and progression make them intriguing sources for biomarker discovery and disease diagnosis[6-9]. In particular, these EV-based disease markers can be identified well before the onset of symptoms or physiological detection of an ailment, making them promising candidates for early stage disease detection[9-14]. Recently, proteins in EVs obtained from cell culture media[15, 16] and from biofluids such as plasma[15-17] and urine[18-21] have been reported to contain PTMs. In general, such modifications can alter protein conformation, stability, activity, cellular localization, and interaction with other cellular molecules. However, the effects of these PTMs on EVs and their potential use as biomarkers or in diagnostics has been largely unexplored until now. Alterations in PTMs in proteins are thought to be major determinants in the early onset and progression of diseases such as cancer and neurodegeneration, and they therefore have become actively pursued targets as indicators of cellular states for disease diagnosis and treatment. However, these targets have been largely unexplored because of the limited availability of tools for studying low abundance EV PTMs in highly complex clinically relevant samples such as plasma. Mass spectrometry (MS) is the major tool used to study PTMs, and multiple MS-based strategies and protocols have been introduced to efficiently analyze PTMs on a proteome-wide scale[22-24]. We have recently reported the identification and quantification of a large number of phosphoproteins and glycoproteins in EVs isolated from human plasma[25, 26]. In this protocol, we describe a workflow for the sequential analysis of both type of modifications and total proteome analysis from a limited amount of starting material.

1.2.1 Development of the approach

Biofluids such as plasma are complex, and—owing to the presence of phosphatases in the bloodstream (e.g., alkaline phosphatase secreted by the liver)—previous attempts to identify phosphoproteins from plasma or serum samples identified only a small number of phosphorylation sites, of which the level of phosphorylation could not be connected to the biological status[27, 28]. To

overcome these challenges, we recently developed a sensitive strategy for the isolation of EVs from biobanked human plasma samples, enrichment of EV phosphopeptides or N-glycopetides, and corresponding EV phosphoproteome or N-glycoproteome analyses by nanoflow liquid chromatography (LC)–MS[25, 26]. In these studies, we identified nearly 10,000 phosphorylation sites and 1,500 glycosylation sites from small (1-mL) amounts of plasma sample and demonstrated the possibility of using PTMs on EV proteins from biofluids as potential biomarkers. We reason that because EVs are membrane-encapsulated compartments, their contents are shielded from the activity of external phosphatases and other enzymes[3, 29, 30], resulting in the successful identification of many PTMs from a limited volume of plasma sample. In addition, analyzing the glycoproteome from EVs instead of whole plasma or serum samples minimizes interference from highly abundant plasma components. This circumvents the challenge of analyzing samples with an extremely wide dynamic range of protein abundancies and enabled us to identify hundreds of glycoproteins that had not been previously discovered in blood[25, 26]. Since the publication of our separate approaches for EV phosphoproteomics[25] and glycoproteomics[26], we updated the workflow by developing a combined pipeline for the sequential isolation of both phosphopeptides and N-glycopeptides from the same plasma sample (Figure 1.1). This allows us to efficiently utilize clinical plasma samples, such as plasma from patients, while maintaining comparable sensitivity for the identification of both phosphopeptides and N-glycopeptides (Results and discussion).

1.2.2 Overview of the procedure

Here, we describe a detailed workflow for the analysis of EV total proteome, phosphoproteome and glycoproteome from human plasma. We focus on several critical steps that improve the reproducibility and efficiency of EV isolation, protein extraction, and phosphopeptide/N-glycopeptide enrichment. We first describe how to prepare the plasma from biobanked or freshly collected human blood (Steps 1 and 2) and how to isolate exosomes and microvesicles (MVs) separately using differential centrifugation (Steps 3–11). The isolation efficiency and specificity can be characterized via multiple techniques such as western blot (WB) using antibodies against EV markers, nanoparticle tracking analysis (NTA), and transmission electron microscopy (TEM) (Step 12). EVs are then subjected to lysis, protein extraction, and digestion (Steps 13–21), followed by C18 desalting (Steps 22–29). Phosphopeptides are enriched through metal ion-affinity capture (Steps 30–36), and the unbound material is used to enrich

glycopeptides through hydrazide chemistry capture followed by enzymatic release of formerly N-glycosylated peptides (Steps 37–44) and StageTip desalting (Steps 45–50). Samples are subsequently analyzed by nanoflow LC–MS analysis (Steps 51 and 52) and we provide detailed instructions for quantitative data analysis (Step 53).

1.2.3 Comparison with alternative approaches

Efficient isolation of EVs from cell culture media or biofluids is the first and most critical step of the downstream EV proteome analysis[10, 19]. Several methods for EV isolation have been introduced[19, 31], including differential ultracentrifugation (UC)[32], size-exclusion chromatography (SEC)[33], polymer precipitation[34], sucrose density gradient[35], and affinity purification that can be either antibody[36] or chemical based[37]. None of the existing isolation methods is perfect; each approach has its own advantages and problems related to, for example, efficiency, purity, isolation time, cost, and demands upon the instruments. For example, subpopulations of EVs can be isolated according to their size or specific marker proteins, but high selectivity also leads to lower EV yields, which can result in incomplete coverage of EV proteomes by MS analysis and potential loss of valuable biological information. EV proteome analyses, in particular PTM analyses, are highly sensitive to contamination from high-abundance plasma/serum proteins, lipids, and reagents in the case of polymer-based precipitation. As a result, some of the EV isolation methods may be suitable for downstream analyses such as nucleic acid sequencing or immunoassays but may not work for MS-based analysis. In this workflow, we chose to use UC for the isolation of EVs as an unbiased approach with low contamination of plasma proteins. Although there are EV proteins that can be used as markers (such as CD9, CD81 and CD63) to evaluate isolation efficiency and specificity, these markers were established using relatively homogeneous cell culture systems and may not be applicable to biofluids, in which many more different types of EVs exist[38-40].

The workflow also includes EV lysis, protein digestion, and PTM-peptide enrichment steps. Alternative strategies have been developed for each step; for example, EV lysis and the extraction of proteins can be achieved using methods developed for cultured cells, with or without any detergent[41-45]. Considering the high membrane content of EVs relative to cells, in this workflow we have adopted a lysis and extraction method for membrane protein analysis to improve EV lysis

and protein extraction[46]. Once EV proteins have been extracted, different protocols can be applied to achieve efficient protein digestion[47-49].

Multiple approaches have also been developed to enrich peptides with PTMs, such as phosphopeptides or glycopeptides, from a complex sample typically dominated by peptides without any modification. For instance, phosphopeptides can be efficiently enriched using immobilized-metal affinity chromatography (IMAC)[50-52], metal oxides (e.g.,TiO2)[51-53], or polymer-based affinity capture[51]. In this workflow, we used polymer-based metal-ion affinity capture (polyMAC)[51], which improves the enrichment efficiency through homogeneous isolation of phosphopeptides from a complex mixture. For more complex samples such as a whole-cell extract, a fractionation step before or after the phosphopeptide enrichment has been shown to be particularly useful. However, we did not find that a fractionation step is needed for EV extracts and instead observed that a single-run EV phosphoproteomics workflow is sufficient. Using a single-run workflow can greatly improve sample throughput and is particularly suitable for label-free quantitation for EV-based biomarker discovery. Similarly, there are continuous efforts in the development of glycopeptide enrichment, which include approaches using classic hydrazide chemistry–based N-glycopeptide capture followed by enzymatic release for MS analysis[54, 55] and, more recently, the isolation of intact glycopeptides using affinity chromatography to identify peptide sequence, glycan structure, and glycosylation sites on the basis of multiple dissociation methods[56-60]. The latter approach can provide rich molecular information and determine the structure of intact glycopeptide, but the technique is still evolving, and the throughput is relatively low, which might be further developed for the identification of protein biomarkers. In this protocol, by primarily focusing on the protein portion, not on the glycan portion, we applied the hydrazide chemistry to enrich N-glycopeptides.

1.2.4 Limitations of the approach

The major limitation of the analysis of EVs obtained from plasma is the complexity of the samples. This complexity is a result of EV heterogeneity, possible contamination from high-abundance plasma components such as plasma proteins, lipids, apoptotic bodies, platelets, and so on, and incomplete understanding of EV biogenesis. Compared to EVs isolated from cell culture media, plasma-derived EVs can be released by any type of cell, and quantitative analysis of the heterogeneity of the EVs will be critical to extracting important and relevant information from the

high background of EVs released by all types of cells. If desired, different EV subpopulations can be isolated and their proteins and modifications can be analyzed separately, thereby allowing one to determine whether the presence of EV PTMs is restricted to specific subtypes of EVs. However, methods for isolating specific EV subtypes are often not efficient, and the amount of resulting EVs may be too low to be processed for measurement by MS.

The majority of EV isolation approaches (including in this protocol) are based on separation by UC. However, there are several drawbacks of this approach: the EV isolation time using UC is long; it requires the operation of delicate equipment; and the yield is typically low and is subject to the sample conditions, such as temperature and dilution factor. Together, these limitations make UC not the ideal method for EV isolation in clinical settings. Among different attempts to develop robust EV isolation methods, we have recently introduced a chemical affinity–based method to readily isolate EVs from urine samples on functionalized magnetic beads with the potential for future clinical applications[37].

Finally, the relatively long procedure of the protocol, technical variability in plasma collection and EV purification, and the low-abundance, high-complexity, and highly dynamic nature of protein PTMs all greatly influence quantitative MS analysis and subsequent EV-based biomarker discovery. Because an increasing number of research groups are currently interested in measuring functional proteins that are specifically present in different EV subtypes, we expect that future studies may focus on targeted approaches such as multiple-reaction monitoring (MRM)[61, 62] for greater sensitivity and more accurate quantification.

1.3 Experimental design

1.3.1 Evaluating the specificity and reproducibility of EV isolation from plasma

The EV isolation needs to be specific, with minimal contamination from aggregated plasma/serum proteins, lipids, and nucleic acids that may affect the PTM enrichment, MS analysis and data interpretation. The isolation of EVs also requires a relatively high yield—because of the limited amount of clinical sample—for successful identification of low-abundance proteins and PTMs from EVs. Furthermore, EVs have to remain intact during the isolation process to prevent their internal cargo from being released. In this workflow, we use UC for EV isolation, which is the most commonly used approach and is considered a “gold standard” approach. The overall UC method is relatively straightforward and does not require additional chemicals or reagents.

Characterization of isolated EVs on the basis of their shape, size, particle concentration and purity is typically required because of the imperfect nature of existing EV isolation methods[63]. Commonly used EV characterization approaches include antibody-based methods to demonstrate the presence of specific EV markers, NTA[64, 65] for EV size distribution and concentration, and electron microscopy (EM)[65] for visual morphology and shape and size of EVs. Typical EV characterization requires at least two orthogonal techniques to obtain comprehensive information on the isolated EVs[66]. To verify whether the UC isolation method is suitable for downstream proteomics and phosphoproteomics, we analyzed two types of samples, one from breast cancer patients and another from healthy controls. Starting with 1 mL of plasma each, we applied the workflow in Figure 1.1 to obtain EV phosphoproteomes. These phosphoproteomes were compared against an EV database downloaded from Vesiclepedia (http://microvesicles.org/). The results showed a 77% overlap between our identified proteins and the EV database (Figure 1.6), with relatively low contamination from plasma proteins. We also carried out Pearson correlation to examine the reproducibility of the isolation method. As shown in Figure 1.7, the reproducibility of the EV isolation method across different samples is relatively high, with Pearson correlations between 0.96 and 0.98.

1.3.2 Extracting EV proteins

The workflow also includes instructions for extracting and digesting EV proteins. With a membrane structure similar to that of cells, standard cell lysis and protein extraction methods have been applied to EVs. To improve the protein extraction efficiency for EVs (which contain a higher proportion of membrane contents as compared to cells), we adopted the phase-transfer surfactant (PTS) protocol[67] for EV lysis and protein extraction. The protocol using sodium deoxycholate (SDC) and N-lauroylsarcosine sodium salt (SLS) as surfactants improves the number of identified peptides with PTMs[46]. In addition, the enzymatic activity is enhanced in the SDC–SLS buffer as compared to other digestion buffers, which results in a lower percentage of missed cleavages in peptides[46, 67].

1.3.3 Sequential enrichment of EV phosphoproteome and N-glycoproteome

In our original approaches[25, 26], EV phosphoproteomics and N-glycoproteomics were performed separately, using different samples of human plasma. Although the published methods work well for diverse biological and clinical samples, the quantities of valuable clinical samples are often limited. Instead of dividing clinical samples into several aliquots or increasing the number of samples by expanding cohorts of patients and corresponding controls, we therefore sought to simultaneously analyze several PTMs in EVs from a single clinical source through sequential enrichment of PTMs. The approach could also provide a unique opportunity to integrate studies of EV protein networks involving multiple PTMs.

In our previous studies[25, 26], we examined MVs and exosomes separately. Our results based on MS and WB analyses identified several exosome or MV protein markers in both MV and exosome fractions, suggesting that high- and ultra-high-speed centrifugation did not isolate plasma EV subpopulations specifically. In the updated protocol, we do not perform the isolation of MVs and exosome for separate MS analyses. This is consistent with the most recent recommendation from the International Society for Extracellular Vesicles[63], which does not propose the use of molecular markers for characterizing a single EV population. We also updated the protocol by isolating phosphopeptides in spintips (StageTips)[68], which can be made in-house by packing Ti(IV)-based solid phase into ordinary pipette tips or can be purchased from commercial sources (e.g., SigmaAldrich), enabling us to carry out single-run phosphoproteomics[69] with low amounts of starting materials (30–50 μg) isolated from plasma EVs. Typically, an 'enhancer', a high concentration of glycolic acid or lactic acid, is used to improve the selectivity of phosphopeptide enrichment. The reagent, however, is not directly compatible with the downstream N-glycopeptide enrichment, because glycolic acid or lactic acid can also be oxidized to generate aldehyde when N-glycopeptides are oxidized by $NaIO_4$ in the N-glycopeptide enrichment protocol[55]. It is possible to add one desalting step to remove the reagent, but the extra step will lead to potentially severe sample loss when minimal material is available. Instead, we introduce a low-pH condition using a high concentration of trifluoroacetic acid (TFA) to obtain similar selectivity, and the TFA can be later removed under vacuum or neutralized with base.

1.3.4 Analyzing total proteomes and additional PTMs

In addition to the analysis of phospho- and glycopeptides, we also provide guidance on how to perform total-proteome analysis (Figure 1.1). A small portion of peptides (typically <1 μg) can be taken after the desalting step (Step 28) for the analysis of the total EV proteome, and the rest is subjected to sequential enrichment of PTM peptides. Furthermore, we expect it would be feasible to include the analysis of additional PTMs in the workflow described in this protocol. Such analysis could include immunoprecipitation with antibodies specific to acetylation or methylation to obtain multi-PTM –omics data using the same clinical samples. For example, EV peptide mixtures (obtained in Step 29) could be first incubated with beads immobilized with antibodies to enrich the peptides by acetylation or methylation, and the flow-through could be subjected to phosphopeptide and/or glycopeptide enrichment.

1.3.5 Plasma samples

The plasma samples used in this protocol were obtained from the Susan G. Komen Breast Cancer Foundation Tissue Bank under Standard Operating Procedure (SOP) 004V6.0 and from Indiana Biobank. We anticipate that this protocol can also be adapted for cell lines, urine or other biofluids. Plasma sample processing was initiated within 30 min of blood draw to a Vacutainer K_2EDTA tube. Samples were spun for 10 min at 1,300*g* to remove blood cells, debris, and large apoptotic bodies. Supernatant was collected and centrifuged at 4,000*g* for 15 min at RT to deplete platelets. Supernatant was collected and stored in -80 ℃ until ready to use.

1.3.6 EV purification

Plasma samples were centrifuged at 20,000 xg at 4 ℃ for 1hr. Pellets were washed with cold PBS and centrifuged again at 20,000 xg at 4 ℃ for 1 hr, the pellets were microvesicles. Supernatant of the first centrifugation was further centrifuged at 100,000 xg at 4 ℃ for 1hr. Pellets were washed with cold PBS and centrifuged at 100,000 xg for 1hr again. The pellets from ultra-high-speed centrifugations were exosome. Two separate isolated EVs were combined during sample lysis. ready to use.

1.3.7 Characterization of EVs

EVs can be characterized based on the heterogeneity, composition and quantity using a method of choice. How extensive the characterization of the EVs should be depends on the specific scientific question asked and on the downstream applications used[66]. General characterization of EVs includes WB using antibodies against known EV markers such as CD9, CD63, CD81, α-actinin-4, CD40 and mitofilin[64, 70]. The size distribution and concentration of EVs can be estimated using NTA[64, 65], tunable resistive pulse sensing (TRPS)[64, 71], or dynamic light scattering (DLS)[64, 65]. In addition, the size and shape can be visualized using electron microscopy (TEM, cryo-EM)[64, 65, 72], or atomic force microscopy (AFM)[64, 65].

1.3.8 EV lysis, reduction, alkylation, and enzymatic digestion

EVs were solubilized in lysis buffer containing 12mM sodium deoxycholate (SDC), 12mM sodium lauroyl sarcosinate (SLS) and phosphatase inhibitor cocktail in 100mM Tris-HCl, pH 8.5. Proteins were reduced and alkylated with 10 mM tris-(2-carboxyethyl)phosphine (TCEP) and 40 mM chloroacetamide (CAA) at 95 °C for 5 min. Alkylated proteins were diluted to 5 fold by 50mM triethylammonium bicarbonate (TEAB) and digested with Lys-C in a 1:100 (w/w) enzyme-to-protein ratio for 3 hr at 37 °C. Trypsin was added to a final 1:50 (w/w) enzyme-to-protein ratio for overnight digestion. The digested peptides were acidified with trifluoroacetic acid (TFA) to final concentration of 0.5% TFA, and 250ul of ethyl acetate was added to 250ul digested solution. The mixture was shaken for 2 min, then centrifuged at 13,200 rpm for 2 min to obtain aqueous and organic phases. The aqueous phase was collected and desalted using a Sep-Pak C18 column.

1.3.9 Phosphopeptide enrichment

We typically enrich phosphopeptides using a Spin-Tip PolyMAC-Ti kit[73]. However, several other suitable phosphopeptide enrichment reagents are available[74, 75]. Peptides were resuspended in 200 µL of loading buffer containing 1% TFA, and 80% ACN for the sequential experiment. For the parallel/separate experiments, samples were resuspended in 200 µL of loading buffer from the commercially available kit which contains glycolic acid. Samples were incubated with PolyMAC-Ti silica beads for 15 min. The beads were loaded into the tip with frit to remove the flow-through. The beads were washed with 200 µL washing buffer containing 100uM glycolic acid, 1% TFA,

and 50% ACN, using centrifuge at 20g for 2 min at RT, followed by 100g for 1 min at RT. Then a second wash with washing buffer containing 25 mM glycolic acid, 80% ACN, and 0.2% TFA and a third time with washing buffer containing 80% ACN in DI water. The phosphopeptides were then eluted from the beads twice with 50 µL of 400 mM ammonium hydroxide, 50%ACN, using centrifuge at 100 rcf. The eluates were collected and dried under vacuum. The flow-throughs were dried for glycopeptide enrichment.

1.3.10 Glycopeptide enrichment

Dried peptides were oxidized with 10 mM sodium periodate in 50% ACN, 0.1% TFA at room temperature with shaking in the dark for 30 minutes. Excess sodium periodate was quenched by using 50 mM sodium sulfite for 15 minutes at room temperature with shaking in the dark. The samples were mixed with 50µL hydrazide magnetic beads slurry and incubated with vigorous shaking at room temperature overnight for the coupling reaction. Magnetic beads were washed sequentially with 400 µL of 50% ACN, 0.1% TFA and 1.5 M NaCl, three times per solution for 1 minute per wash for the removal of non-coupled peptides. Beads were rinsed once with 100 µL of 1x GlycoBuffer 2 (NEB) and incubated with 3 µL of PNGase F (NEB) in 100 µL 1x GlycoBuffer 2 (NEB). N-glycans were cleaved by PNGase F and N-glycopeptides were desalted using SDB-XC StageTip.

1.3.11 LC-MS/MS

The PTM peptides were dissolved in 4 µL of 0.3% formic acid (FA) with 3% ACN and injected into an Easy-nLC 1200 (Thermo Fisher Scientific). Peptides were separated on a 45 cm in-house packed column (360 µm OD × 75 µm ID) containing C18 resin (2.2 µm, 100Å, Michrom Bioresources) with a 30 cm column heater (Analytical Sales and Services) set to 50 °C. The mobile phase buffer consisted of 0.1% FA in ultra-pure water (buffer A) with an eluting buffer of 0.1% FA in 80% ACN (buffer B) run over either with a 45 min or 60 min linear gradient of 5%-25% buffer B at flow rate of 300 nL/min. The Easy-nLC 1200 was coupled online with a Thermo Scientific™ mass spectrometer. Thermo LTQ Orbitrap Velos Pro was used to profile PTMs in plasma-derived EVs and a QExactive hybrid quadrupole-Orbitrap mass spectrometer for parallel reaction monitoring (PRM) additional experiments. The mass spectrometer was operated in the

data-dependent mode in where the 10 most intense ions were subjected to high-energy collisional dissociation (HCD) fragmentation (normalized collision energy (NCE) 30%, AGC 3e4, max injection time 100 ms) for each full MS scan (from m/z 350-1500 with a resolution of 120,000 at m/z 200).

1.3.12 Data processing

The raw files were searched directly UniprotKB database with no redundant entries using MaxQuant software with the Andromeda search engine. Initial precursor mass tolerance was set to 20 p.p.m. and the final tolerance was set to 6 p.p.m., and ITMS MS/MS tolerance was set at 0.6 Da. Search criteria included a static carbamidomethylation of cysteines (+57.0214 Da) and variable modifications of oxidation (+15.9949 Da) on methionine residues, acetylation (+42.011 Da) at N-terminus of protein, and phosphorylation (+79.996 Da) on serine, threonine or tyrosine residues for phosphorylation, and deamidation (+0.984Da) on asparagine residues for glycosylation were searched. Search was performed with Trypsin/P digestion and allowed a maximum of two missed cleavages on the peptides analyzed from the sequence database. The false discovery rates of proteins, peptides and PTMs sites were set at 0.01. The minimum peptide length was six amino acids, and a minimum Andromeda score was set at 40 for modified peptides. The glycosylation sites were selected based on the matching to the N-X-S/T (X not Pro) motif. A site localization probability of 0.75 was used as the cut-off for localization of glycosylation and phosphorylation sites. All the peptide spectral matches and MS/MS spectra can be viewed through MaxQuant viewer.

1.4 Results and discussion

We previously identified nearly 10,000 unique phosphopeptides[25] and 1,500 unique glycopeptides[26] in EVs using several milliliters of human plasma by analyzing the samples separately. Here, we have described an updated protocol that allows simultaneous analysis of the EV phosphoproteome and N-glycoproteome. This combined workflow is especially suitable for situations in which the amount of starting material is limited, as is often the case for clinical samples. To compare the performance of the separate phosphoproteomics and N-glycoproteomics approaches to the sequential workflow, we carried out label-free quantitation of the sequential and

separate enrichment procedures using the same plasma samples. We started with a total of 9 mL of human plasma (Figure 1.2): 3 mL for three technical replicates of sequential enrichment of phosphopeptides and N-glycopeptides, 3 mL for three technical replicates of the enrichment of phosphopeptides only, and 3 mL for three technical replicates of the enrichment of N-glycopeptides only. Using a 60-min LC gradient on a high-resolution Orbitrap mass spectrometer, we identified ~4,000 unique EV phosphopeptides for each 1 mL of human plasma, representing ~1,300 phosphoproteins in each sequential or separate replicate (Figure 1.3). Similar to the previous study[25], the selectivity (the percentage of PTM peptides over total peptides) for EV phosphopeptide identification from plasma samples (65–70%) was lower than phosphopeptide identification from a typical whole-cell extract (>90%), which is probably due to the small amount of starting materials and potential interference from EV components. The low-pH buffer condition for sequential enrichment led to only a 7% decrease in phosphopeptide identification number, and the correlation coefficients decreased from 0.94 (average) among technical replicates to 0.87 (average) between sequential and separate enrichments. Similarly, we identified virtually the same number of unique EV N-glycopeptides for each 1 mL of human plasma, representing ~650 glycoproteins in either sequential or separate replicates. We saw a moderate decrease in the selectivity (from 54 to 42%; Figure 1.4) and correlation coefficients (from 0.97 on average to 0.85 on average between sequential and separate workflows; Figure 1.5). Overall, our data indicate that the sequential enrichment of phosphopeptides and N-glycopeptides offers greater benefit by achieving similar numbers of identified PTMs with half the amount of starting material. We further evaluated the level of variation in the composition of the EVs among individual samples from healthy donors. To do this, we carried out targeted analyses on five selected EV phosphopetides using PRM[76-78] (Table 1.2 and Figure 1.8). The measured coefficient of variation (CV) values between the samples from the five individuals are relatively high, indicating either a large variation in the concentration of these proteins in the EV and/or variation in the extent to which these peptides are modified between individuals.

1.5 Data availability

The raw MS data from this study have been deposited into the ProteomeXchange Consortium via the PRIDE Archive with the identifier PXD013893 (https://www.ebi.ac.uk/pride/archive/projects/ PXD013893).

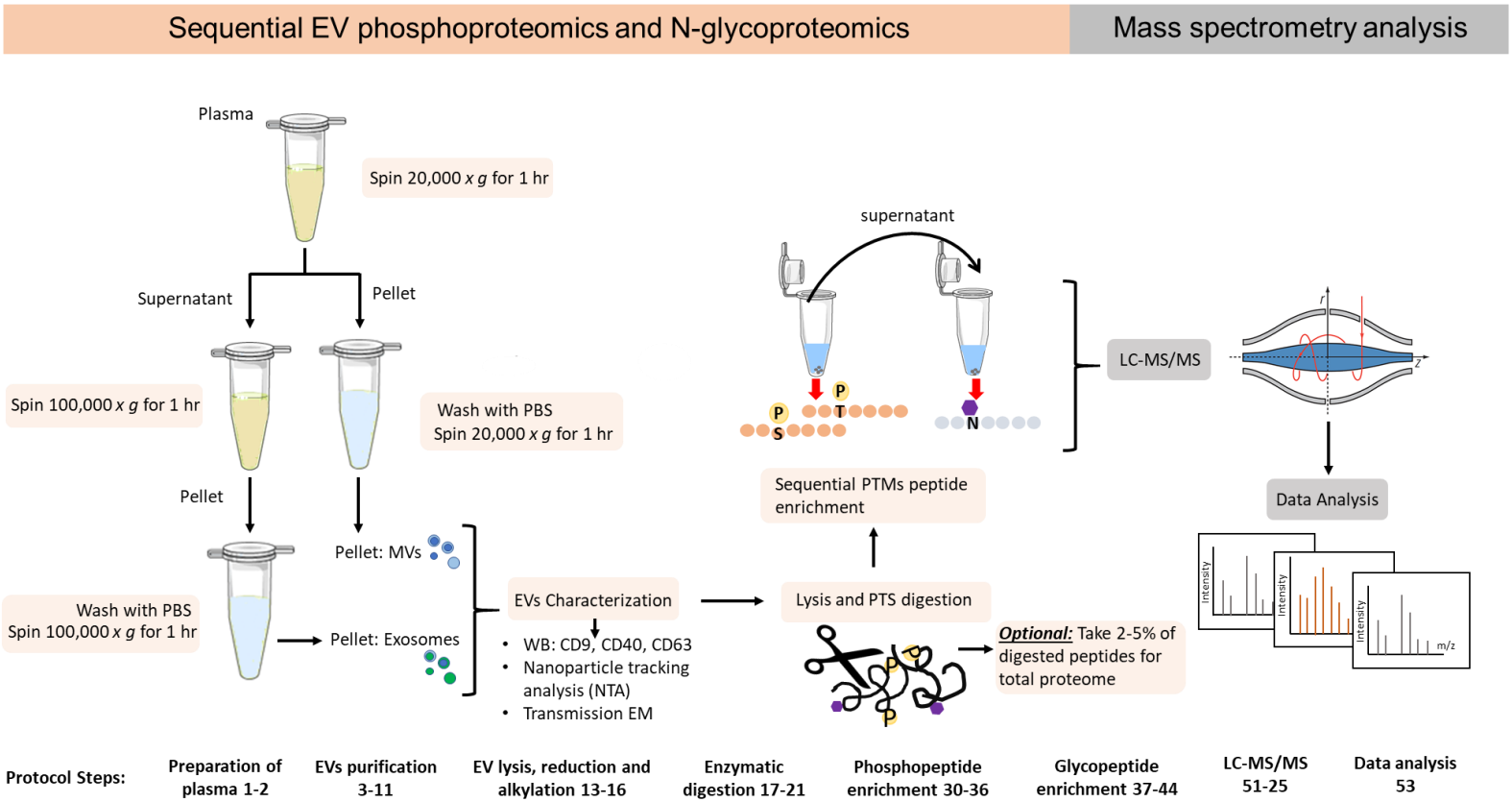

Figure 1.1. Workflow for sequential EV phosphoproteomics and glycoproteomics.
The workflow for the isolation of EVs, enrichment of phosphopeptides and glycopeptides, and nanoflow LC–MS analysis. Microvesicles and exosomes are isolated from human plasma through sequential high-speed and ultra-high-speed centrifugation. EVs are lysed and proteins are extracted and digested in SDC buffer. Phosphopeptides and N-glycopeptides are sequentially enriched, followed by LC–MS/MS analysis. Protocol steps are indicated. N, asparagine; P, phosphorylation; S, serine; T, threonine.

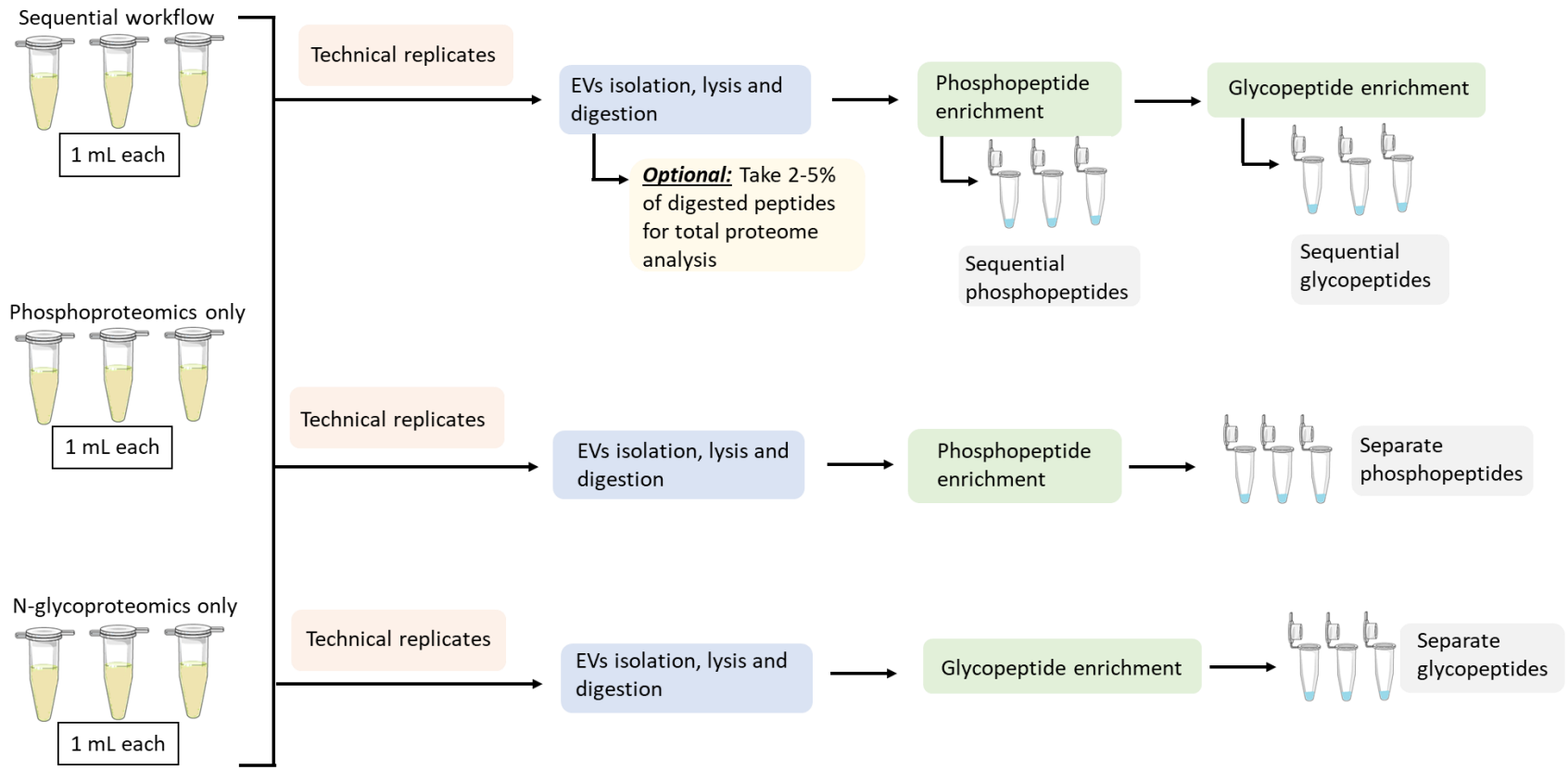

Figure 1.2. Schematic of experimental setup for sequential and separate phosphopeptide and glycopeptide enrichments. For the sequential analysis of both phosphopeptides and glycopeptides, the workflow as shown in Figure 1.1 is used. For the separate enrichment procedures, only one modification is enriched, by either phosphorylation (by skipping Steps 37–44) or glycosylation (by skipping Steps 30–36).

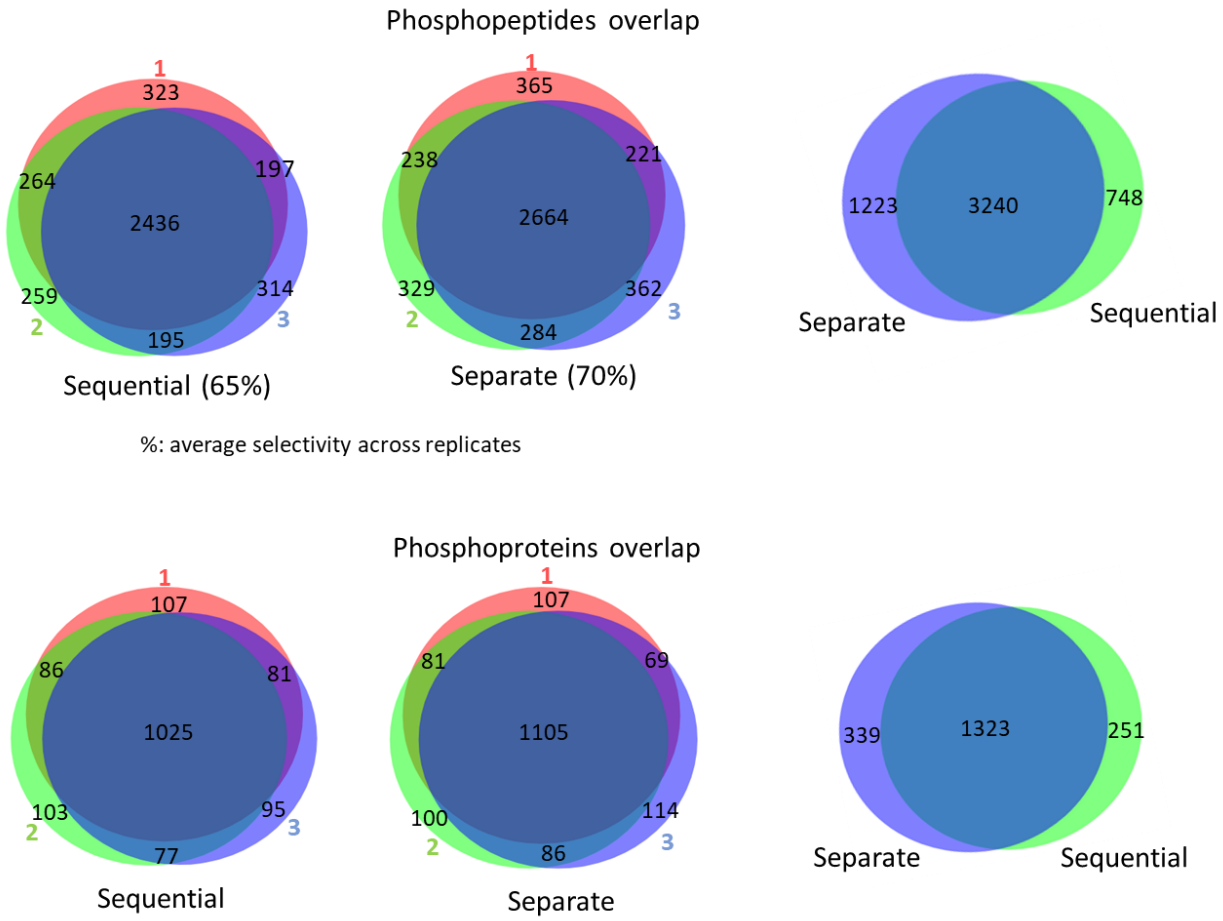

Figure 1.3. Venn diagrams showing the phosphopeptide and phosphoprotein overlap between replicates and the overlap between the separate and sequential workflows.
The overlaps between replicates are shown at the phosphoprotein and phosphopeptide level (left-hand and center diagrams). Selectivity across replicates and the overlap between the results obtained from the separate workflows and the sequential workflows (right-hand diagrams) are depicted.

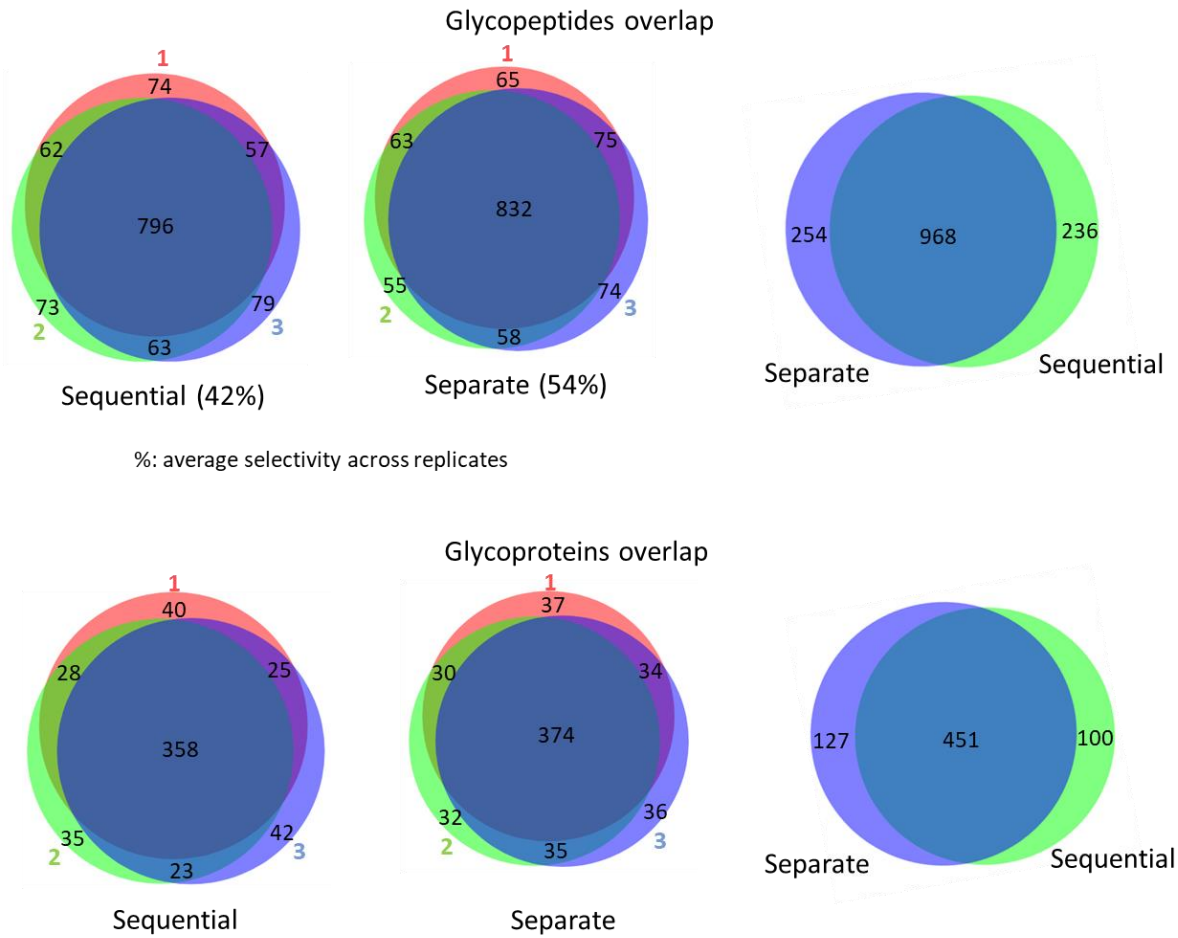

Figure 1.4. Venn diagrams showing the glycopeptide and glycoprotein overlap between replicates and the overlap between the separate and sequential workflows.
The overlaps between replicates are shown at the glycoprotein and glycopeptide level (left-hand and center diagrams). Selectivity across replicates and the overlap between the results obtained from the separate workflows and the sequential workflows (right-hand diagrams) are depicted.

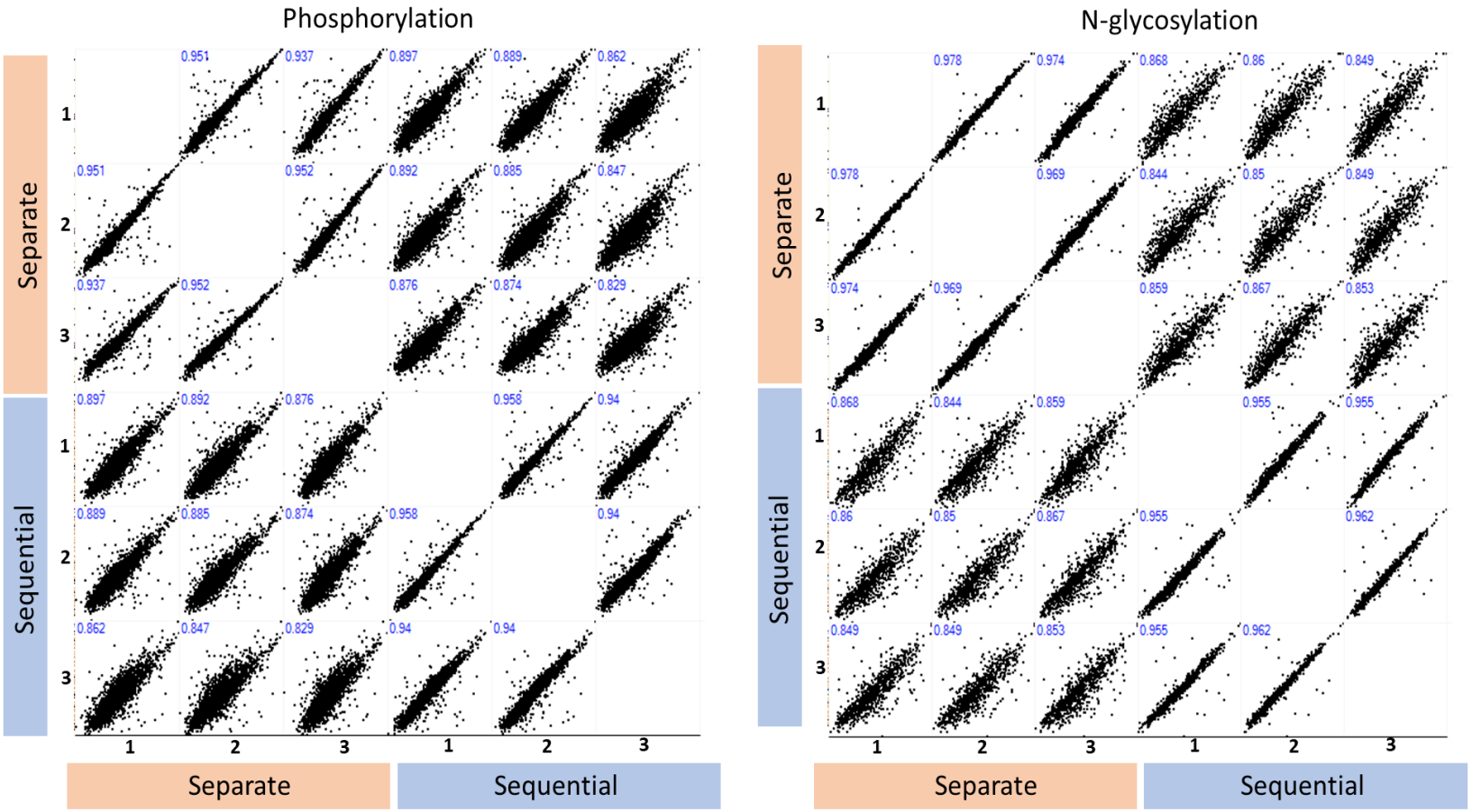

Figure 1.5. Quantitation results from MaxQuant and Perseus showing Pearson correlations across each condition and replicate. Scatterplots and Pearson correlation coefficients depicting the log2-tranformed intensities of phosphopeptides and glycopeptides across both workflows (sequential versus separate) and three technical replicates.

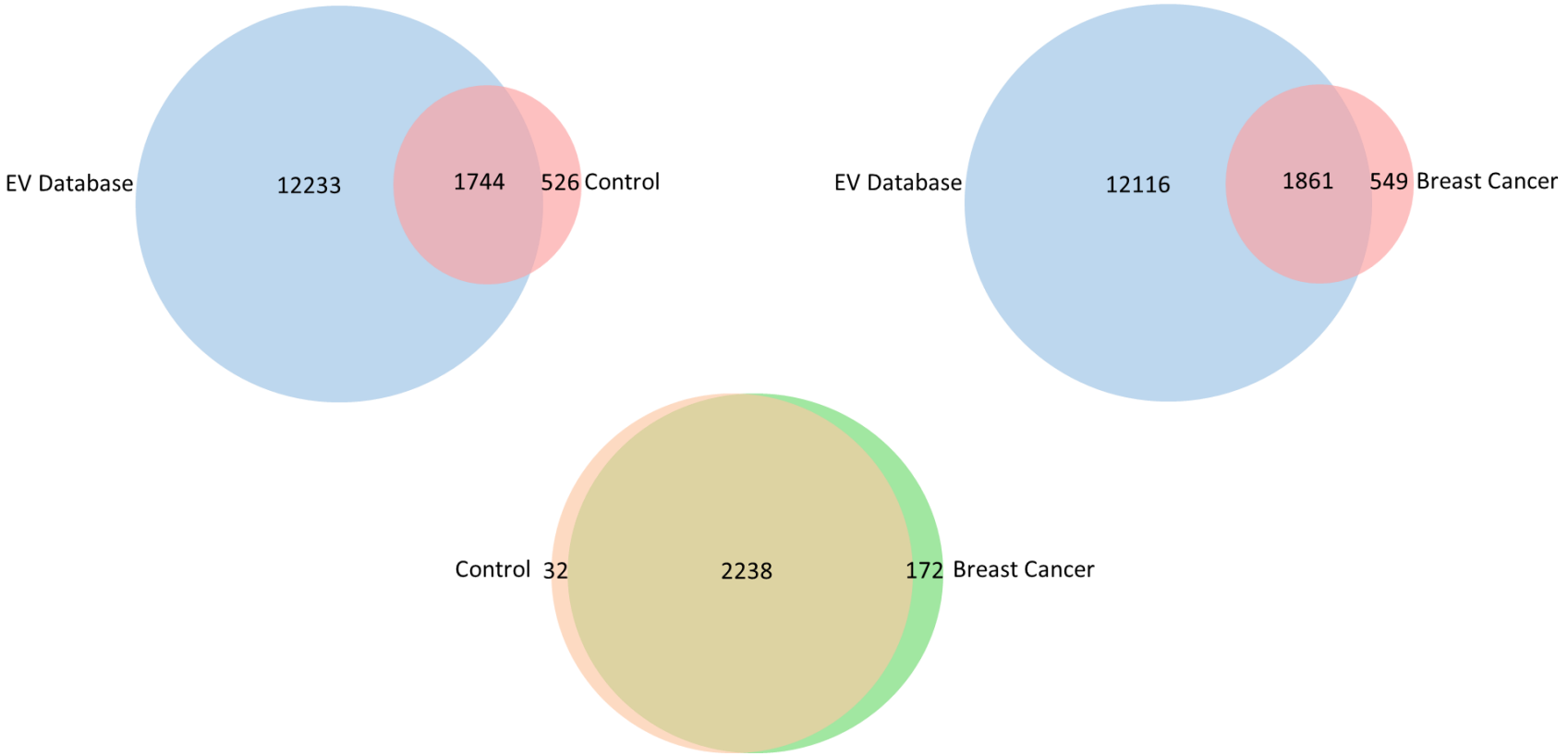

Figure 1.6. Venn diagrams showing the overlap between the EV database from Vesiclepedia and EV data from control and breast cancer.

The Venn diagrams show a 77% overlap between both, the EV database and control EV proteins, and the EV database and EV breast cancer proteins. Overlap of the proteome from the control and the breast cancer EV data is also shown.

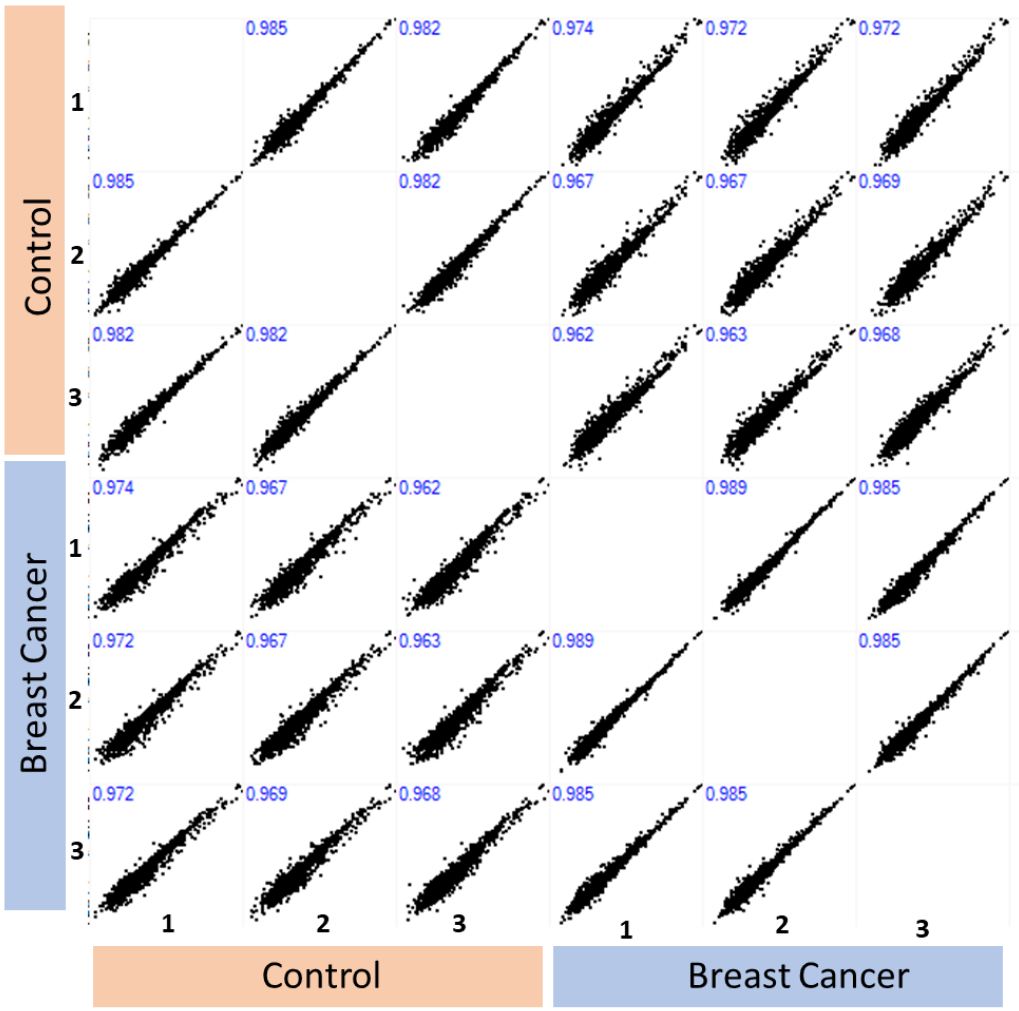

Figure 1.7. Quantitation results from MaxQuant and Perseus showing Pearson correlations from proteome analysis across each condition and replicate.
Scatterplots and Pearson correlation coefficients depicting the log2-transformed intensities from proteome analysis of control and breast cancer samples, each in triplicate.

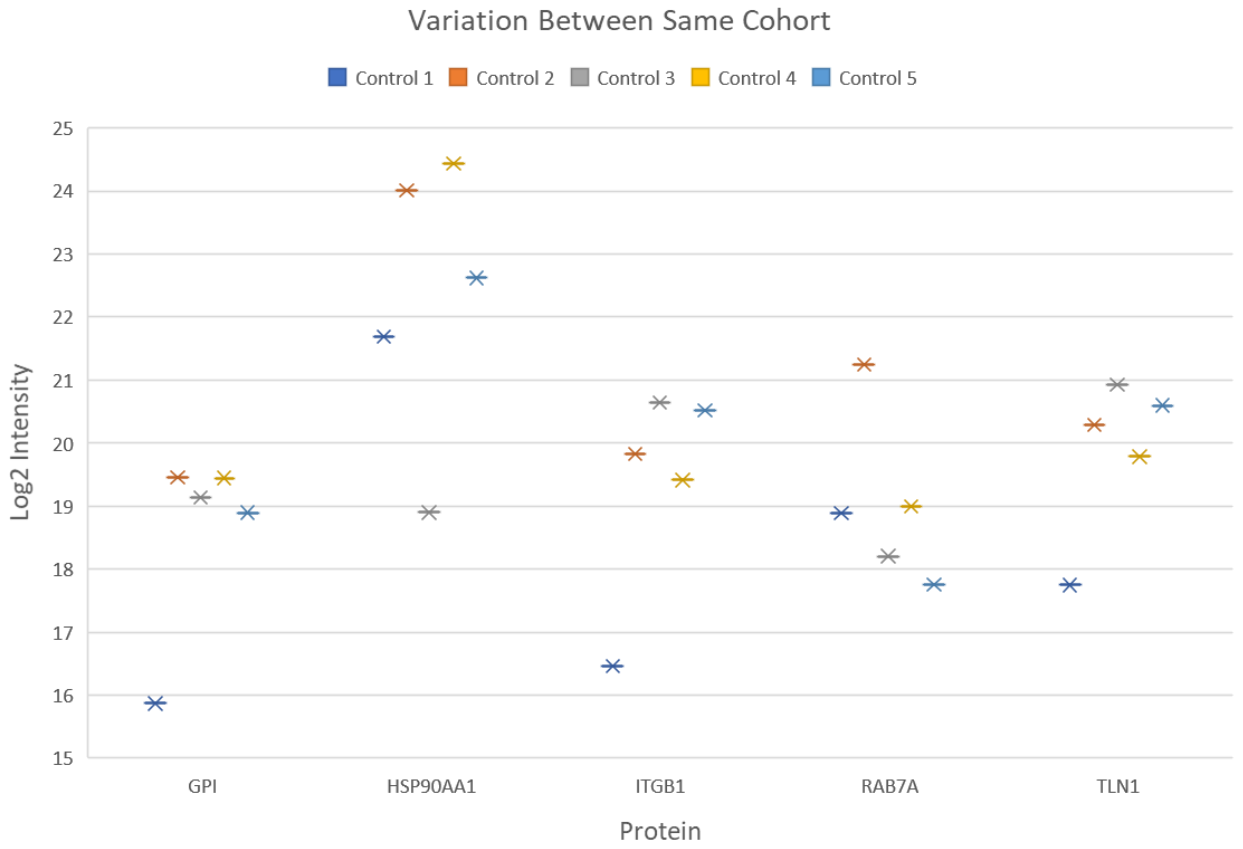

Figure 1.8. Scatterplots representing the targeted proteomic (PRM) analysis.
Raw intensities from five individual control samples corresponding to five phosphopeptides from five EV proteins were plotted.
See Table 1.2 for details.

Table 1.1. Comparison of EV phosphopeptides, phosphoproteins, N-glycopeptides and N-glycoproteins using sequential or separate procedures

Sample	**#glycopeptides**	**#total peptides**	**Selectivity**	**#glycoproteins**
Sequential 1	1003	2389	42%	653
Sequential 2	1005	2439	41%	638
Sequential 3	1004	2407	42%	638
Separate 1	1042	1960	53%	673
Separate 2	1013	1893	54%	656
Separate 3	1043	1943	54%	660

Sample	**#phosphopeptides**	**#total peptides**	**Selectivity**	**#phosphoproteins**
Sequential 1	3947	6110	65%	1298
Sequential 2	3864	5963	65%	1289
Sequential 3	3845	5948	65%	1276
Separate 1	4156	6022	69%	1361
Separate 2	4176	5997	70%	1371
Separate 3	4186	5973	70%	1372

Table 1.2. Raw intensities from five individual control samples monitoring five peptides corresponding to five EV proteins, including mean, SD, and CV (%). Proteins selected are in the top 100 EV proteins from Vesiclepedia.

Protein Gene	Peptide	Peptide Modified Sequence Average Masses	Precursor Charge	Control 1	Control 2	Control 3	Control 4	Control 5	Mean	SD	CV (%)
GPI	SNTPILVDGK	SNT[+79.979901]PILVDGK	2	59865	718569	575007	712412	487559	510682.4	241586.6	47.30662
HSP90AA1	EVSDDEAEEK	EVS[+79.979901]DDEAEEK	2	3382085	16903736	488350	22667768	6458404	9980069	8425475	84.42301
ITGB1	WDTGENPIYK	WDT[+79.979901]GENPIYK	2	90336	930164	1637321	698009	1506258	972417.6	562536.3	57.84925
RAB7A	FQSLGVAFYR	FQS[+79.979901]LGVAFYR	2	486314	2491084	302097	521000	221576	804414.2	850692.6	105.7531
TLN1	VLVQNAAGSQEK	VLVQNAAGS[+79.979901]QEK	2	219752	1279873	1986830	902402	1582806	1194333	603541.5	50.53379

CHAPTER 2. INTEGRATED PTMS IN PLASMA-DERIVED EXTRACELLULAR VESICLES AS FINGERPRINTS FOR BREAST CANCER SUBTYPES

2.1 Summary

EVs are emerging as important biological carriers for immune regulation and intercellular communications. As protein post-translational modifications (PTMs) such as phosphorylation, acetylation, and glycosylation regulate cell signaling events essential for cellular functions, PTMs in EVs may be determined to assess cellular physiological status. Here we establish a liquid biopsy platform to simultaneously measure multiple PTMs in plasma EVs from the same clinical samples and show that modified proteins including known oncogenes could be identified in EVs and can be used to distinguish breast cancer subtypes. Protein phosphorylation, acetylation, and glycosylation were concurrently analyzed through sequentially isolating acetylated-, phosphorylated-, and N-glycosylated peptides in EVs from more than 100 individual plasma samples from luminal A/B, triple negative breast cancer patients, and healthy controls. We identified more than 10,000 phosphopeptides, 900 acetylated peptides, and 1,000 glycopeptides in plasma EVs through data-dependent acquisition and prioritized 135 phosphopeptides, 47 acetylated peptides, and 98 glycopeptides for quantification in individual plasma EV samples. A total of 132 scheduled parallel reaction monitoring (PRM) analysis was performed, through which we successfully generated a panel of specific PTM sites that has the capability of potentially differentiating breast cancer subtypes. The findings reveal that the integrative determination of protein PTMs in plasma-derived EVs provide a new strategy for biomarker discovery and present a clinically convenient approach to diagnose disease status.

2.2 Introduction

EVs, including exosomes and microvesicles, contain a wealth of nucleic acids, proteins and signaling molecules vital for intercellular communication[1, 3, 79]. With growing evidence that EVs reflect the molecular signature of the parent cell, the increase in the understanding of EV compositions is critical for their establishment as valuable repertoires of biomarkers for human diseases[4, 80]. Research into potential biomarkers isolated from EV has been propelled by the

development of methods and tools to acquire them by minimally and non-invasive means, which reinforces their great diagnostic potential[6-8]. In the context of cancer, this opens the door to apply EV based liquid biopsy for early detection ahead of imaging studies currently prevailing.

Protein post-translational modifications (PTMs), such as phosphorylation, glycosylation and acetylation, are involved in multiple essential cellular processes and represent an important mechanism of regulation for cellular physiological functions, leading to the development of effective and targeted therapeutics[81]. Discovery and profiling PTMs have established the relevance of PTMs in EVs and associated EV functions and novel applications. Tandem mass spectrometry (MS/MS) has been the main tool to study PTMs on a global proteome level. Several studies including some from our lab have previously demonstrated MS-based methods to profile phosphoproteome and glycoproteome in EVs from biofluids such as plasma and urine[25, 26, 37, 82, 83]. These studies assert the PTMomic approach is a viable path towards developing non-invasive diagnostic tools as each different pathology will have specific and dynamic molecular signatures carried by EVs.

Breast cancer (BC) has been extensively characterized and its molecular complexity and heterogeneity is well recognized. The molecular subtypes of BC can affect clinical outcomes and treatment response in patients[84]. Therefore, a potential diagnostic modality would be greatly beneficial if it could distinguish between molecular subtypes. These subtypes, defined by expression of estrogen receptor (ER), progesterone receptor (PR) or the epidermal growth factor receptor (ErbB2/Her2), breakdown into three major groups, luminal A/B, Her2 positive, and triple negative (TN)[84]. The subtype luminal A is characterized by having the highest expression of ER and PR receptor genes and the lowest expression of proliferation-related genes[84]. The subtype luminal B is characterized by lower expression of ER and PR receptor genes and high expression of proliferation-related genes. Importantly, some luminal B subtypes can also be Her2 positive. However, despite having the best clinical outcome and survival rate, most deaths from metastatic breast cancer (MBC) come from patients with luminal A/B subtypes[85]. Although endocrine therapy remains the main treatment, it can lose effectiveness due to primary or acquired endocrine resistance[85]. Endocrine therapies include tamoxifen, which is the “gold standard” and oldest treatment in ER positive breast cancers, endocrine therapies for ovarian suppression (Lupron, Zoladex), and aromatase inhibition (Anastrozole, Letrozole)[85].

In the case of Her2 positive subtype, which is characterized by the amplification of the HER2 gene and overexpression of the Her2 protein, trastuzumab is used as the main therapy as it exerts anti-tumor effects. Trastuzumab is a humanized recombinant monoclonal antibody that binds selectively to the extracellular domain of HER2. In conjunction with endocrine therapies, if PR or ER receptor genes present, and traditional chemotherapy, trastuzumab remains the most effective treatment in Her2 subtype breast cancer[86]. On the other hand, TNBC subtype, where none of the three markers is expressed, has an especially poor prognosis[84].

The recurrence rate in BC is high due to primary or acquired resistance, or lack of pharmacological treatment resulting in a high rate of therapeutic or prophylactic mastectomies[85, 87, 88]. It is now known that EV-based disease markers can be identified before the onset of the disease or as markers to aid in detection of relapse in breast cancer[25, 89]. Distinguishing between subtypes at the early stages of breast cancer will allow for patients to receive correct therapies sooner and will universally increase the survival rate. Moreover, the opportunity to evaluate plasma EVs in patients that undergo mastectomy procedures could be favorable, as patients with resections lack breast tissue for biopsies.

The integration of multiple PTM signatures has already been demonstrated as a viable method to outline lung cancer signaling networks[90]. Moreover, previous work has shown the profiling of breast cancer allows for distinction between molecular subtypes using an omics style approach[89, 91]. However, all previous multi-PTM analyses have been achieved either with cell culture systems or with tissue samples. We reason that integrating PTM biosignatures in plasma EVs greatly enhance the detection or prognosis of malignancies, in this case profiling breast cancer subtypes, in clinical applications. EVs carry a wealth of biological information from parent cells, which makes their cargo, including proteins, nucleic acids, and metabolites, viable biological read-outs of the parent cells. Analyzing proteins with PTMs from EV cargo provides a read of cellular regulation and processes associated with signaling pathways that reflect disease status. The combination of different PTMs isolated from plasma EV in breast cancer patients allows for delineation of breast cancer subtypes suggesting a novel way forward for monitoring breast cancer patient pathophysiology.

In this study, we present a novel strategy to sequentially isolate and analyze PTMs from plasma-derived EVs from the same biological sample to differentiate breast cancer subtypes and aid in treatment assignment. Our study focuses on luminal A/B and triple negative due to their

unique profile as predominantly metastatic, and lacking pharmacological treatment, respectively. Interestingly, our data reveals very distinct profiles when analyzing EV PTMs. Our EV proteome data alone did not provide as extensive information on subtype specific targets as our PTM data further highlighting the important role of PTMs. Overall, this study shows the feasibility of a plasma-derived EV pipeline as a promising alternative for elucidating clinically relevant targets in disease pathology.

2.3 Experimental design

2.3.1 Plasma samples

The Iowa University Institutional Review Board approved the use of human plasma samples. In the global PTM-ome experiment, blood samples were collected from healthy females obtained through Susan G. Komen Tissue Bank and from breast cancer subtypes obtained through the University of Iowa biobank. Plasma samples were collected by standard protocol, in brief, plasma sample processing was initiated within 30 min of blood draw to an ethylenediaminetetraacetic acid (EDTA) containing tube. Samples were spun for 30 min at 3500 rpm to remove all cell debris and platelets. Samples were stored in -80 °C until ready to use.

2.3.2 Extracellular vesicles isolation

The EVs isolation and digestion were performed according to the reported protocol[67, 82]. A total of 5 ml pooled plasma samples was collected from both healthy individuals and patients diagnosed with breast cancer for the global PTMs experiment as technical replicates. Plasma samples were centrifuged at 20,000 xg at 4 °C for 1hr. Pellets were washed with cold PBS and centrifuged again at 20,000 xg at 4 °C for 1 hr, the pellets were microvesicles. Supernatant of the first centrifugation was further centrifuged at 100,000 xg at 4 °C for 1hr. Pellets were washed with cold PBS and centrifuged at 100,000 xg for 1hr again. The pellets from ultra-high speed centrifugations were exosome. Two separate isolated EVs were combined during sample lysis.

2.3.3 Enzymatic digestion

The enzymatic digestion was performed with phase transfer surfactant aided (PTS) digestion[67]. Extracellular vesicles were solubilized in lysis buffer containing 12mM sodium

deoxycholate (SDC), 12mM sodium lauroyl sarcosinate (SLS) and phosphatase inhibitor cocktail in 100mM Tris-HCl, pH8.5. Proteins were reduced and alkylated with 10 mM tris-(2-carboxyethyl)phosphine (TCEP) and 40 mM chloroacetamide (CAA) at 95 °C for 5 min. Alkylated proteins were diluted to 5 fold by 50mM triethylammonium bicarbonate (TEAB) and digested with Lys-C in a 1:100 (w/w) enzyme-to-protein ratio for 3 hr at 37 °C. Trypsin was added to a final 1:50 (w/w) enzyme-to-protein ratio for overnight digestion. The digested peptides were acidified with trifluoroacetic acid (TFA) to final concentration of 0.5% TFA, and 250ul of Ethyl acetate was added to 250ul digested solution. The mixture was shaken for 2 min, then centrifuged at 13,200 rpm for 2 min to obtain aqueous and organic phases. The aqueous phase was collected and desalted using a 100 mg of Sep-Pak C18 column.

2.3.4 Tyrosine phosphopeptides enrichment

Desalted peptides were resuspended in 50 mM Tris-HCl pH 7.5 and incubated with 20 uL PT66 beads with rotation overnight at 4^{o}C. The PT66 beads were washed sequentially with lysis buffer (50mM Tris-HCl, 50mM NaCl, 1%NP40 pH7.5) and water, three times per solution for 10 mins rotation to wash off non-specific binding. Tyrosine phosphopeptides were sequential eluted twice by 0.1%TFA and once with 0.1%TFA/50%ACN. The eluent was dried under vacuum and then subjected to PolyMAC enrichment.

2.3.5 Lysine acetylation peptides enrichment

Immunoaffinity enrichment of lysine acetylated peptides from EVs was performed according to manufacturer's instructions (PTMScan). In brief, 20ul of lysine acetylation antibody conjugated beads were washed extensively with PBS. The flow-through from tyrosine phosphopeptides were mixed with lysine acetylation antibody beads and incubated for 2hr at 4°C. The beads were washed twice with IAP buffer (50 mM MOPS, pH 7.2, 10 mM sodium phosphate, 50 mM NaCl) and three times with water. Peptides were eluted from beads with 0.15% TFA (sequential elutions of 55 µl followed by 50 µl, 10 min each elution at room temperature). Eluted peptides were desalted by SDB-XC stage tip and eluted with 40% acetonitrile in 0.1% TFA. Eluted peptides were dried under vacuum. The flow-through were desalted by SDB-XC stage tip and dried under vacuum.

2.3.6 PolyMAC phosphopeptides enrichment

Peptides were resuspended in 200 µL of loading buffer containing 1% TFA (trifluroacetic acid), and 80% acetonitrile and incubated with PolyMAC-Ti silica beads for 15 min. The beads were loaded into the tip with frit to remove the flow-through. The beads were washed twice with 200 µL washing buffer containing 100 mM glycolic acid, 1% TFA, and 50% ACN and once with 80% ACN, using centrifuge at 100 rcf. The phosphopeptides were then eluted from the beads by twice with 50 µL of 400 mM ammonium hydroxide, 50%ACN, using centrifuge at 100 rcf. The eluates were collected and dried under vacuum. The flow-through were dried for glycopeptides enrichment.

2.3.7 N-Glycopeptides enrichment

The glycopeptide enrichment was performed according to the reported protocol[54]. Desalted peptides were oxidized with 10 mM sodium periodate in 50% ACN, 0.1% TFA at room temperature with shaking in the dark for 30 minutes. Excess sodium periodate was quenched by using 50 mM sodium sulfite for 15 minutes at room temperature with shaking in the dark. The samples were mixed with 50 µL/100 µL hydrazide magnetic beads for individual and pooled samples respectively and incubated with vigorous shaking at room temperature overnight for the coupling reaction. Magnetic beads were washed sequentially with 400 µL/800 µL of 50% ACN, 0.1% TFA and 1.5 M NaCl for individual and pooled samples respectively, three times per solution for 1 minute per wash for the removal of non-coupled peptides. Beads were rinsed once with 100 µL/200 µL of 1x GlycoBuffer 2 (NEB) for individual and pooled samples respectively and incubated with 3 µL/4 µL of PNGase F (NEB) in 100 µL/200 µL for individual and pooled samples, respectively. N-glycans were cleaved by PNGase F. After desalting, the released former N-glycopeptides were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS).

2.3.8 LC-MS/MS

The PTM peptides were dissolved in 4 µL of 0.3% formic acid (FA) with 3% ACN and injected into an Easy-nLC 1200 (Thermo Fisher Scientific). Peptides were separated on a 45 cm in-house packed column (360 µm OD × 75 µm ID) containing C18 resin (2.2 µm, 100Å, Michrom

Bioresources) with a 30 cm column heater (Analytical Sales and Services) set to 50 °C. The mobile phase buffer consisted of 0.1% FA in ultra-pure water (buffer A) with an eluting buffer of 0.1% FA in 80% ACN (buffer B) run over either with a 45 min or 60 min linear gradient of 5%-25% buffer B at flow rate of 300 nL/min. The Easy-nLC 1200 was coupled online with a Thermo Scientific™ Orbitrap Fusion™ Tribrid™ mass spectrometer. The mass spectrometer was operated in the data-dependent mode in where the 10 most intense ions were subjected to high-energy collisional dissociation (HCD) fragmentation (normalized collision energy (NCE) 30%, AGC 3e4, max injection time 100 ms) for each full MS scan (from m/z 350-1500 with a resolution of 120,000 at m/z 200).

2.3.9 Data processing

The raw files were searched directly UniprotKB database version Aug2017 with no redundant entries using MaxQuant software (version 1.5.6.1)[92, 93] with the Andromeda search engine. Initial precursor mass tolerance was set to 20 p.p.m. and the final tolerance was set to 6 p.p.m., and ITMS MS/MS tolerance was set at 0.6 Da. Search criteria included a static carbamidomethylation of cysteines (+57.0214 Da) and variable modifications of (1) oxidation (+15.9949 Da) on methionine residues, (2) acetylation (+42.011 Da) at N-terminus of protein, and (3) phosphorylation(+79.996 Da) on serine, threonine or tyrosine residues for phosphorylation, acetylation (+42.011 Da) on lysine residue for acetylation and deamidation (+0.984 Da) on asparagine residues for glycosylation were searched. Search was performed with Trypsin/P digestion and allowed a maximum of two missed cleavages on the peptides analyzed from the sequence database. The false discovery rates of proteins, peptides and PTMs sites were set at 0.01. The minimum peptide length was six amino acids, and a minimum Andromeda score was set at 40 for modified peptides. The glycosylation sites were selected based on the matching to the N-X-S/T (X not Pro) motif. A site localization probability of 0.75 was used as the cut-off for localization of glycosylation sites. All the peptide spectral matches and MS/MS spectra can be viewed through MaxQuant viewer.

2.3.10 Quantitative data analysis

Data from the pooled and PRM experiments was analyzed using the Perseus software (version 1.5.4.1)[94] and Skyline[95]. For quantification of both proteomic and PTM-omic datasets, the

intensities of proteins and PTMs sites were derived from MaxQuant, and the missing values of intensities were replaced by normal distribution with a downshift of 1.8 standard deviations and a width of 0.3 standard deviations. The significantly increased PTMs sites or proteins in patient samples were identified by ANOVA multi-test with a permutation-based FDR cut-off 0.05 for all of data sets. For heatmap, the changed sites or proteins were used, the imputed data set was normalized by z-score within each dataset. For the PRM data, after differential intensities at the modification site level were selected, the corresponding precursor peptides were selected and imported to Skyline, in addition to the search results from MaxQuant that were used to generate a spectral library and the raw data was used to visualize the extracted ion chromatograms. After removing peptides that were detected as differential due to erroneous extracted ion chromatogram integration (XIC) by MaxQuant, the list of target peptides was refined by removing the sequences that contained more than 3 amino acids that could carry the respective modification (STY for phosphorylation, K for acetylation and N for glycosylation), ragged tryptic ends, sequences longer than 25 amino acids or any histidine.

Analysis of the precursor ions m/z's were collapsed for phospho-isoforms and additional filtering was done to fit the desired duty cycle of 2.5 seconds with an injection time of 100 ms for each MS2 scan for glycopeptides and acetylated peptides or 50 ms for phosphopeptides, with a retention time window of 5 minutes before the minimum observed retention time and 5 minutes after the highest observed. This process was done manually, giving priority to the peptides that showed the highest observed difference in intensities between conditions. Variable importance for classification was calculated by using the implementation of random forest in the ranger package[96] for the R language and environment for statistical computing. The metric used to calculate the importance was the impurity-corrected method[97] on a forest fit with 20000 trees.

2.4 Results

2.4.1 Identification of 20,788, 11,181, 1,035, and 914 unique peptides, phosphopeptides, glycopeptides, and acetopeptides in plasma EVs

To identify PTM biosignatures in EVs from human plasma, we developed a robust workflow for the isolation of EVs from plasma, sequential enrichment of phosphorylated-, N-glycosylated-, and acetylated- peptides analyses (Figure 2.1). Human plasma was centrifuged first at low speed to remove cell debris and EVs were isolated through high-speed (20k xg) and ultra-high-speed

centrifugations (100k xg), an approach that has been used in our previous studies[25, 82]. For the initial screening, plasma samples were collected and pooled from healthy individuals (n = 20) and from patients diagnosed with luminal A/B breast cancer (LAB) (n = 20) and triple negative breast cancer (TNBC) (n = 20). Each pool had a final volume of 5 mL for three technical replicates, among which 0.250 mL were taken from each patient. After lysis of EVs, proteins were extracted, denatured, alkylated, and enzymatically digested using Lys-C and trypsin with the aid of phase-transfer surfactants for fewer missed tryptic sites and better digestion[67]. Sequential enrichment was performed for each sample starting with tyrosine phosphorylation using PT66 antibody, followed by lysine acetylation, S/T phosphorylation by PolyMAC[73] and N-glycopeptide enrichment using hydrazide chemistry approach[54]. Label-free quantitation was performed by LC-MS/MS on a high-speed and high-resolution mass spectrometer to determine the differential PTM proteins in plasma-derived EVs from control and breast cancer subtypes patients. Based on a pipeline that allowed for the enrichment of three PTMs from the same clinical sample, the platform enabled us to identify 20,788, 11,181, 1,035, and 914 unique peptides, phosphopeptides, glycopeptides, acetopeptides, representing 2,693, 1,764, 504, and 331 proteins, phosphoproteins, N-glycoproteins and acetylated proteins with 1% FDR, respectively. We overlapped our EV proteome data against an EV curated database extracted from Vesiclepedia. Results showed a 70% overlap between our identified proteins and the EV database, indicating overall selective and efficient isolation and identification of EV proteins (Figure 2.2). Moreover, we performed a Pearson correlation to examine the reproducibility of the approach. As shown in Figure 2.2, reproducibility of the EVs across each condition and replicate was considerably high, with Pearson correlations between 0.95-0.98.

2.4.2 Quantitative analysis of EV proteome, phosphoproteome, N-glycoproteome and acetylproteome in BC subtypes

Label-free quantitative analysis of EV proteome, phosphoproteome, N-glycoproteome, and acetylproteome between healthy controls and breast cancer subtypes is illustrated in Figure 2.3. Quantitative analyses of EV proteomes revealed similar expressions between LAB and TNBC compared to controls whereas better distinctions across subtypes were visualized in the phosphoproteome, N-glycoproteome and acetylproteome. This supports the claim that proteomics analysis alone might not provide adequate information about differentiation among breast cancer subtypes. The data also showed that the PTM differences among breast cancer subtypes and

controls were not merely a result of differences in protein expression. Therefore, this justifies the need to develop PTMomics approaches to deeper analyze specific signaling and regulation events with breast cancer patients. As a notable example, programmed death-ligand 1 (PD-L1) was identified in both EV phosphoproteome and N-glycoproteome, and its glycoform was significantly higher in TNBC patients compared to LAB and controls. PD-L1 was significantly expressed in cancer cells and in breast cancer specifically, in TNBC patients according to several studies[98-102]. This further supports our claim to develop PTM-omics approaches and the potential of EVs as relevant biomarkers. Recently, monoclonal antibody targeting of glycosylated PD-L1 was found to inhibit PD-L1/PD-1 interaction[102]. This manipulation produced a novel response in which targeted and adjacent TNBC cancer cells were killed via internalization and degradation of PD-L1/PD-1 complex[102]. Furthermore, this monoclonal antibody also induced a bystander-killing effect of adjacent TNBC cells lacking PD-L1 expression without detectable toxicity demonstrating the therapeutic potential of targeting protein glycosylation[102]. This evidence further reinforces the importance of analyzing glycosylation events in disease models, which is highly emphasized in our methodological approach.

We built gene ontology (GO) functional enrichment networks of the phosphoproteome, N-glycoproteome, and acetylproteome among healthy controls and breast cancer subtypes (Figures 2.4 and 2.5). The enrichment analysis showed upregulated proteins in each modification and connections between genes. Notably, the phosphoproteome in plasma derived EVs better distinguished LAB and TNBC with signaling pathways related to each subtype. These pathway relationships were also confirmed by several studies performed in tissue or cell culture[103-109]. ErbB signaling pathway was more prominent in LAB phosphoproteome, while signaling by receptor tyrosine kinase family related pathways were more prominent in TNBC. In the EV acetyl- and N-glyco-proteomes, the enrichment analysis mainly identified metabolic pathways and PI3K-Akt signaling pathways, respectively. This is supported by literature which suggests metabolic pathways are strongly associated with lysine acetylation, and abnormal protein glycosylation is activated in PI3K-Akt signaling pathways and Notch signaling pathways[103-108]. We also carried out principal component analysis (PCA) on each modification (Figure 2.6). The multi-valent analyses allowed us to reveal the distance and relatedness between sample replicates and conditions, with the overall purpose being to reduce the dimensionality of a dataset containing multiple variables. Interestingly, PCA for the proteome analysis did not distinguish LAB and

TNBC as accurately as the phosphoproteome analysis did. For the acetylproteome, variables were more distanced, particularly in the LAB, and in the N-glycoproteome, we observed a separation comparable to the one in the phosphoproteome. Overall, this indicates the phosphoproteome could better distinguish between breast cancer subtypes compared to the other two modifications and the proteome.

2.4.3 Integrated EV PTMs to distinguish BC subtypes

After the first screening phase performed with the pool samples between healthy controls, LAB and TNBC, we proceeded to select a group of target-specific markers per modification and breast cancer subtypes. Overall, 135, 98 and 47, phosphopeptides, N-glycopeptides and acetylated peptides, respectively, were selected for the targeted approach. We performed a total of 132 scheduled PRM runs with 44 individual samples to quantify individual EV modifications using 700 µL plasma per sample according to the same sequential approach as in the pool experiment.

Variable importance classification (Figure 2.7) showed the top 30 classifiers to distinguish breast cancer subtypes according to our random forest analysis: 22 are phosphorylated targets, 6 acetylated and 2 glycosylated. The top 2 target classifiers were Bruton's Tyrosine Kinase (BTK) which is phosphorylated and myosin-9 (MYH9) which is acetylated. BTK expression has been reported in breast cancer cells and to protect these cells from apoptosis[109-112]. Moreover, there are recently developed inhibitors of BTK such as ibrutinib (PC1-32765), AVL-292, and CGI-1746 which prevent drug-resistant clones from arising and reduce breast cancer cell survival[110, 112].

On the other hand, MYH9 has been associated with cancer in several studies[113-116]. Animal studies have shown that defective MYH9 expression is correlated with oncogenesis and tumor progression in human squamous cell carcinoma (SCC) and invasive lobular breast carcinoma (ILBC)[114]. Other relevant markers, which are listed in our top 30 candidate targets, and have been related to cancer and/or breast cancer[117-122] include tensin-1 (TNS1), stathmin, also known as metablastin and oncoprotein 18 (STMN1), and Msx2-interacting protein (SPEN). STMN1 has been highly associated with aggressive phenotypes of breast cancer[117, 120, 121]. Its overexpression has been correlated with breast cancer proliferation in low estrogen and progesterone receptors and high histological grade in human breast cancer[120, 121]. Additional, SPEN has been found to be implicated in hormone-dependent breast cancers as it functions as a tumor suppressor and

candidate biomarker of drug responsiveness[122]. Overall, our study suggests that EV markers in plasma are, to some extent, reflective of candidate markers that have been studied in cell culture systems or tissue samples.

The top 2 best classifiers for breast cancer subtypes, BTK and MYH9, showed great potential distinguishing between controls from LAB and TNBC (Figure 2.7). The scatterplot depicting log-2 intensities of these two targets could segregate to some extent our PRM individual samples into their respective category. Moreover, Figure 2.8 shows the top 30 proteins per modification better suited for discriminating BC subtypes with the percentages of detectable levels on the individual samples according to our random forest analysis. Altogether, the selected EV proteins signature show prominent specificity for the respective breast cancer subtype as stated before. More importantly, this shows the capability of using EVs as feasible tools to develop clinical diagnostics that could enable better BC therapeutic decisions.

2.5 Discussion

The data generated for this study offers a unique opportunity to evaluate and analyze the proteome and three different PTMs from the same biological sample without compromising the quality of the data. This tested the feasibility of a PTM-omic pipeline to elucidate targets that could better distinguish BC subtypes. Our bioinformatic analysis approach was capable of integrating all targets from the 3 modifications to come up with a panel of 30 candidate markers to best distinguish BC subtypes. These candidate markers have been previously studied in cell culture systems and/or tissue samples and have been associated with BC progression. The strong potential of these EVs as circulating biomarkers offers great benefits for developing diagnostic tools for better decision making in therapeutics for BC.

Moreover, we also demonstrated the relevance of evaluating PTMs, as proteomic analysis alone did not provide as extensive information on subtype specific targets as PTMs did, especially phosphorylation. Moreover, we detected targets such as glycosylated PD-L1 significantly increased in TNBC which agrees with recent studies that suggest monoclonal targeting of glycosylated PD-L1 in TNBC patients as a potential novel therapy[102]. Additionally, we showed significantly upregulated candidate markers in our top 30 targets such as BTK, MYH9, TNS1, STMN1, and SPEN which have been reported in the BC literature. Furthermore, our 2 top classifiers, BTK and MYH9, were able to discriminate among BC subtypes in our 44-individual

cohort (Figure 2.7) and we were also able to generate a candidate target list per modification (Figure 2.8).

Overall, our study displays the utility of EV biomarkers in disease diagnostics. One of the biggest challenges of this approach—and of the liquid biopsy approaches in general—lies in implementing the identified candidate markers in large prospective studies for validation and further clinical use. Nonetheless, this study provides pivotal insight about possible platforms for better therapeutics. Significant effort needs to be spent in the future to establish EV related clinical assays that can provide true clinical benefits and can be implemented in clinical settings.

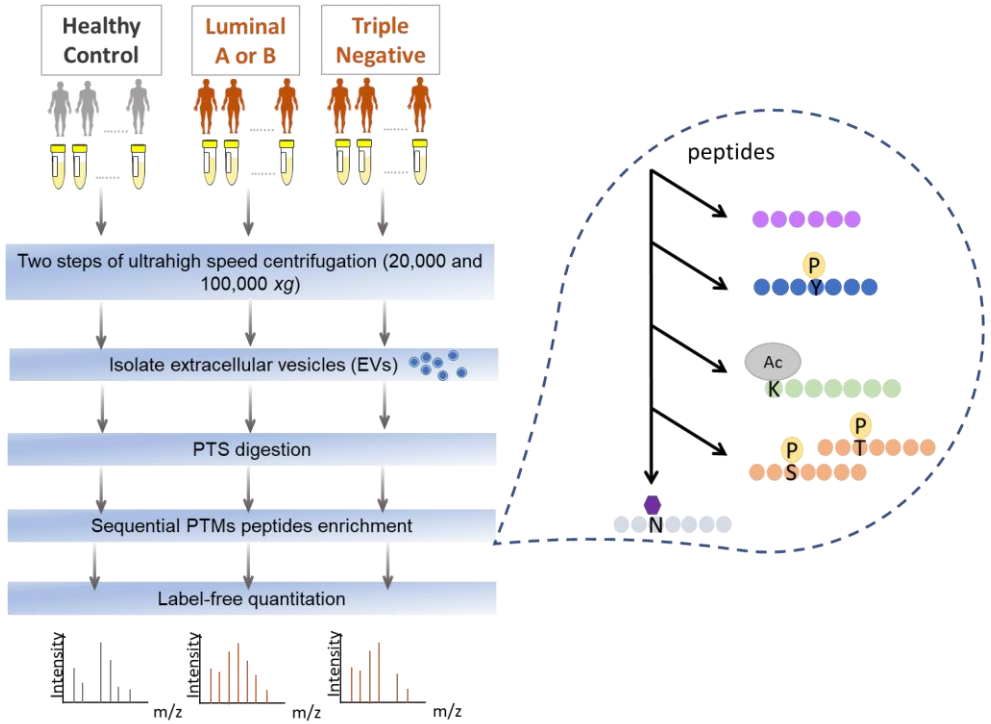

Figure 2.1. Workflow of the PTM-omics pipeline for plasma-derived EVs in healthy control and breast cancer patients. Plasma samples were pooled from healthy individuals (n=20), luminal A/B (n=20), and triple negative patients (n=20). EVs were isolated from human plasma through high speed and ultra-high-speed centrifugation. After isolation, EVs were lysed, proteins were extracted and enzymatically digested with LysC and trypsin. Sequential enrichment was done starting with tyrosine phosphorylation using PT66 antibody, followed by lysine acetylation, S/T phosphorylation by PolyMAC and glycopeptide enrichment with hydrazide chemistry. Samples were analyzed by LC-MS/MS on a high-speed and high-resolution mass spectrometer with technical replicates. Label-free quantitation was performed to determine the differential PTM proteins in the plasma of control and two subtypes of breast cancer patient samples. Finally, targeted proteomics (PRM) was performed to verify possible candidates using 44 individual samples.

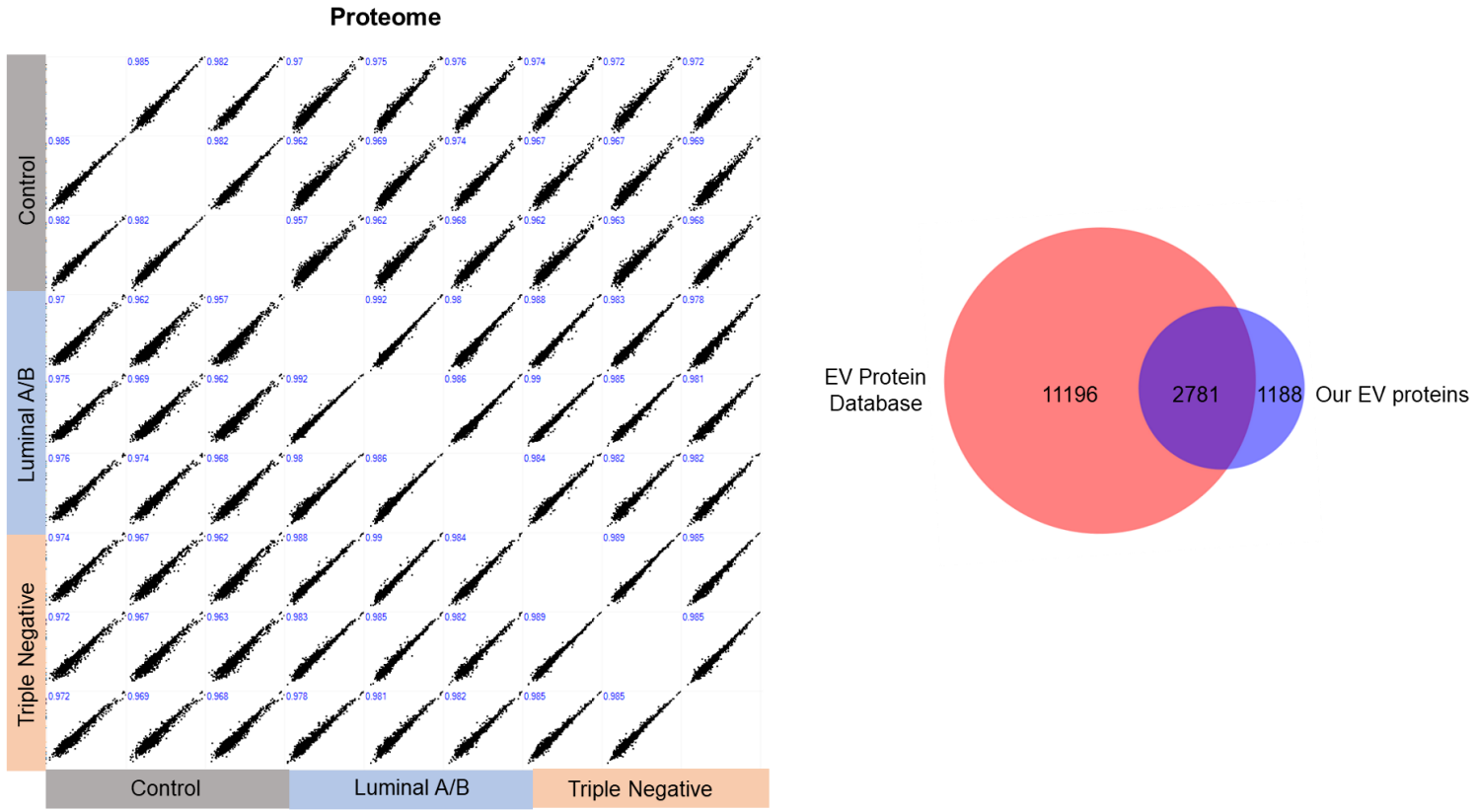

Figure 2.2. Quantitation results from MaxQuant and Perseus showing Pearson correlations from the proteome analysis across each condition and replicate (left) and Venn diagram showing the overlap between our EV dataset and Vesiclepedia, a manually curated EV database (right).
Scatterplots depicting the log2-transformed intensities from the proteome analysis with Pearson correlations between 0.95-0.98 and Venn diagram showing a 70% overlap between our EV dataset with Vesiclepedia.

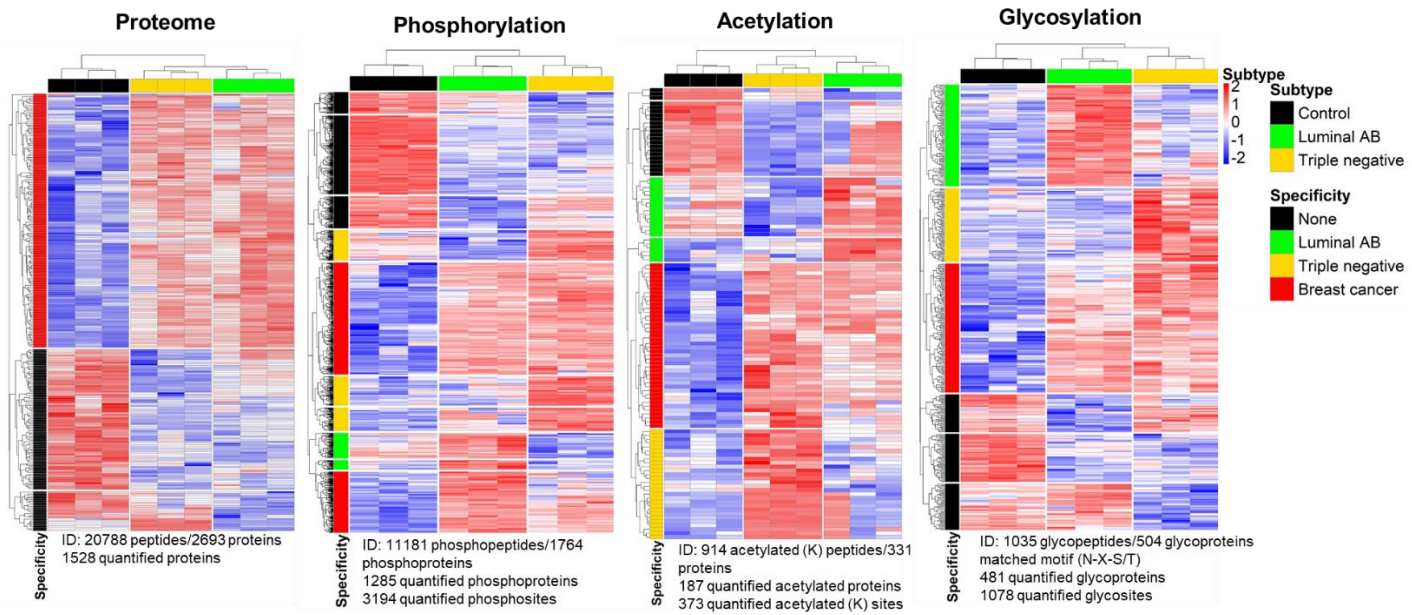

Figure 2.3. Heatmaps showing quantitative analysis PTM-omics between breast cancer subtypes and healthy controls. Heatmaps representing the quantitative analysis of the proteome, phosphoproteome, acetylproteome and glycoproteome depicting each condition. The imputed data set was normalized by z-score which shows red as 2 and blue as -2.

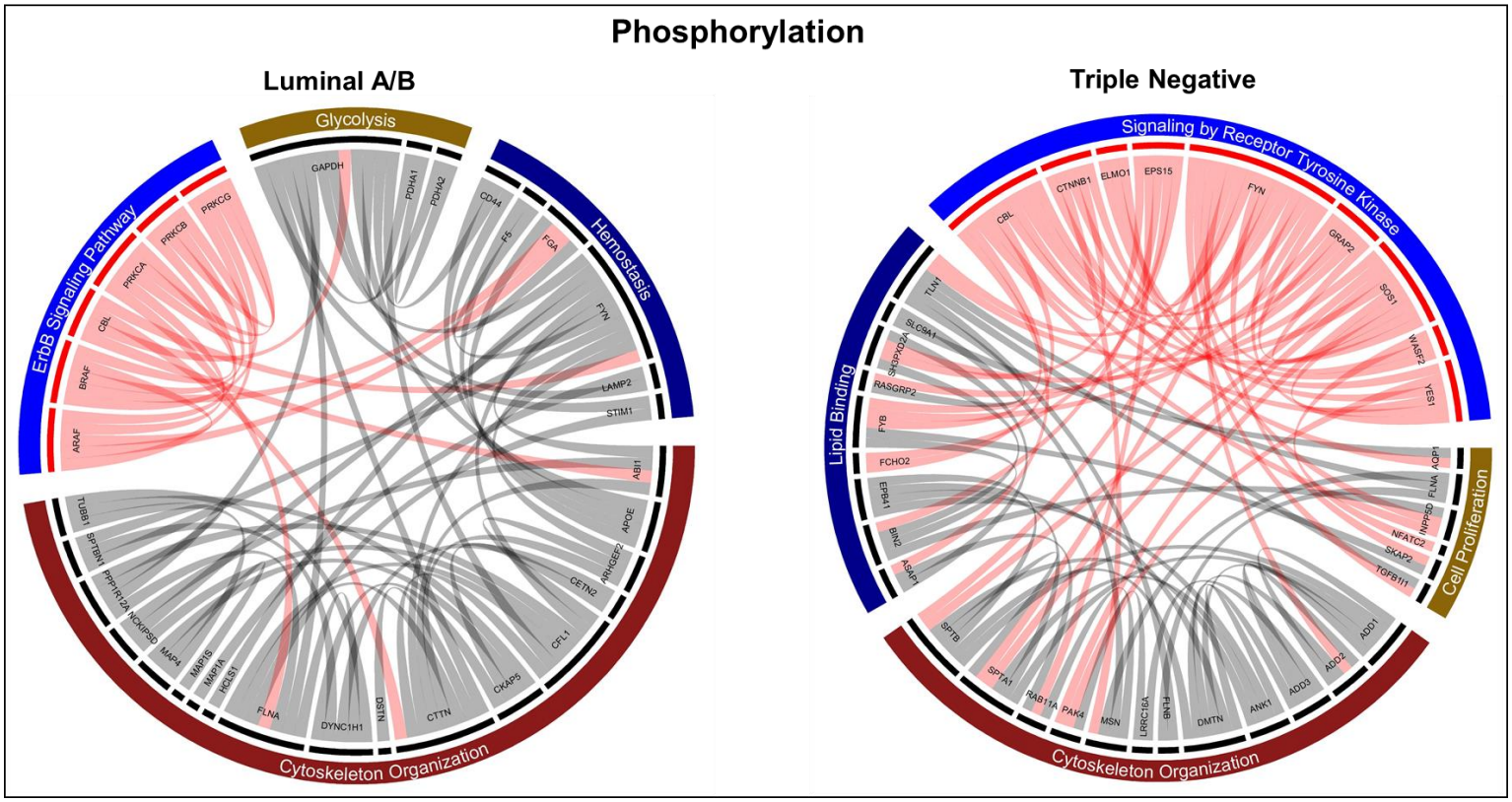

Figure 2.4. The gene ontology circos plot analysis of upregulated genes in the phosphoproteome data.
Circos plot depicting the upregulated genes and its connections with other upregulated targets in luminal A/B and triple negative breast cancer.

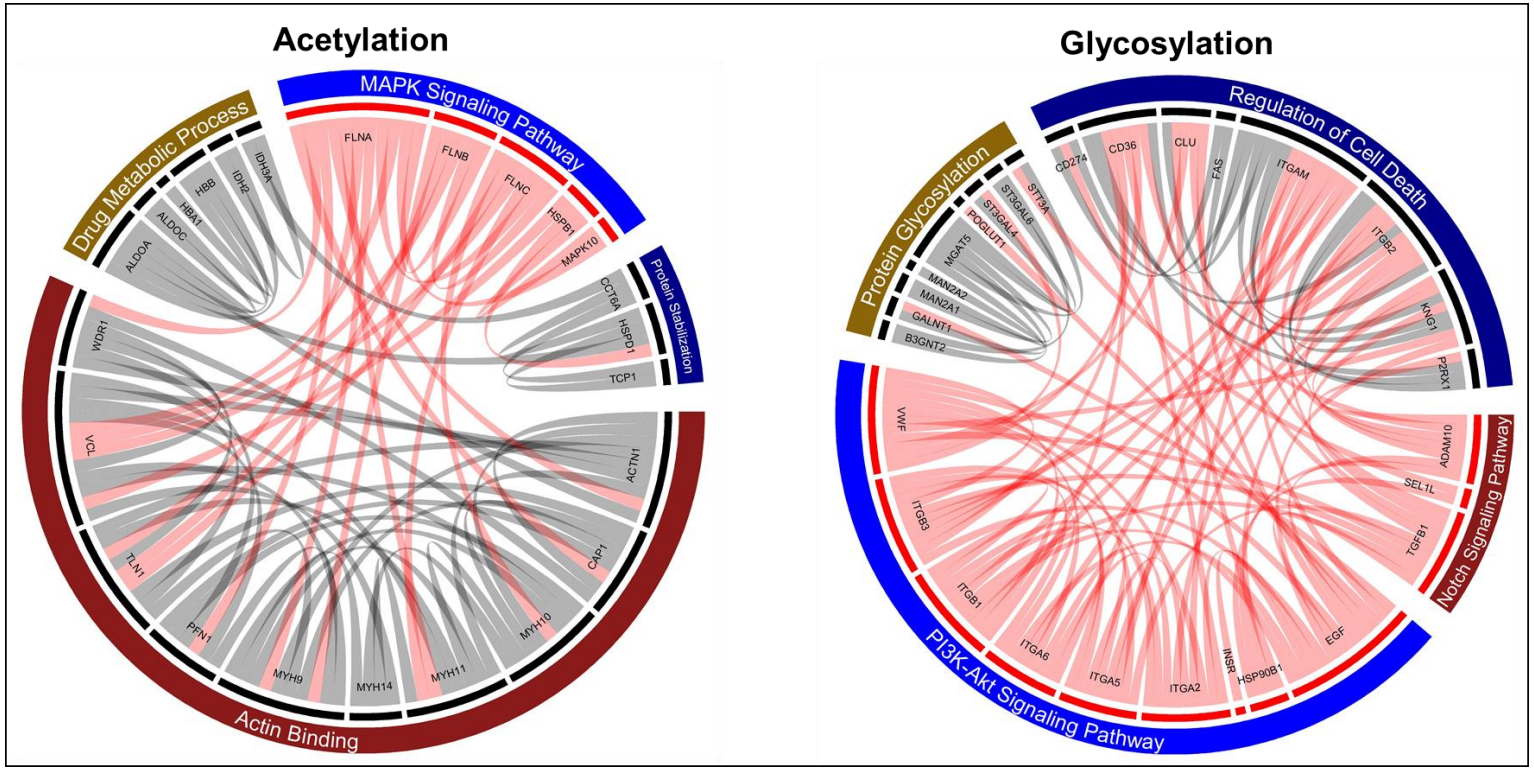

Figure 2.5. The gene ontology circos plot analysis of upregulated genes in the acetylproteome and glycoproteome data. Circos plot depicting the upregulated genes and its connections with other upregulated targets in breast cancer.

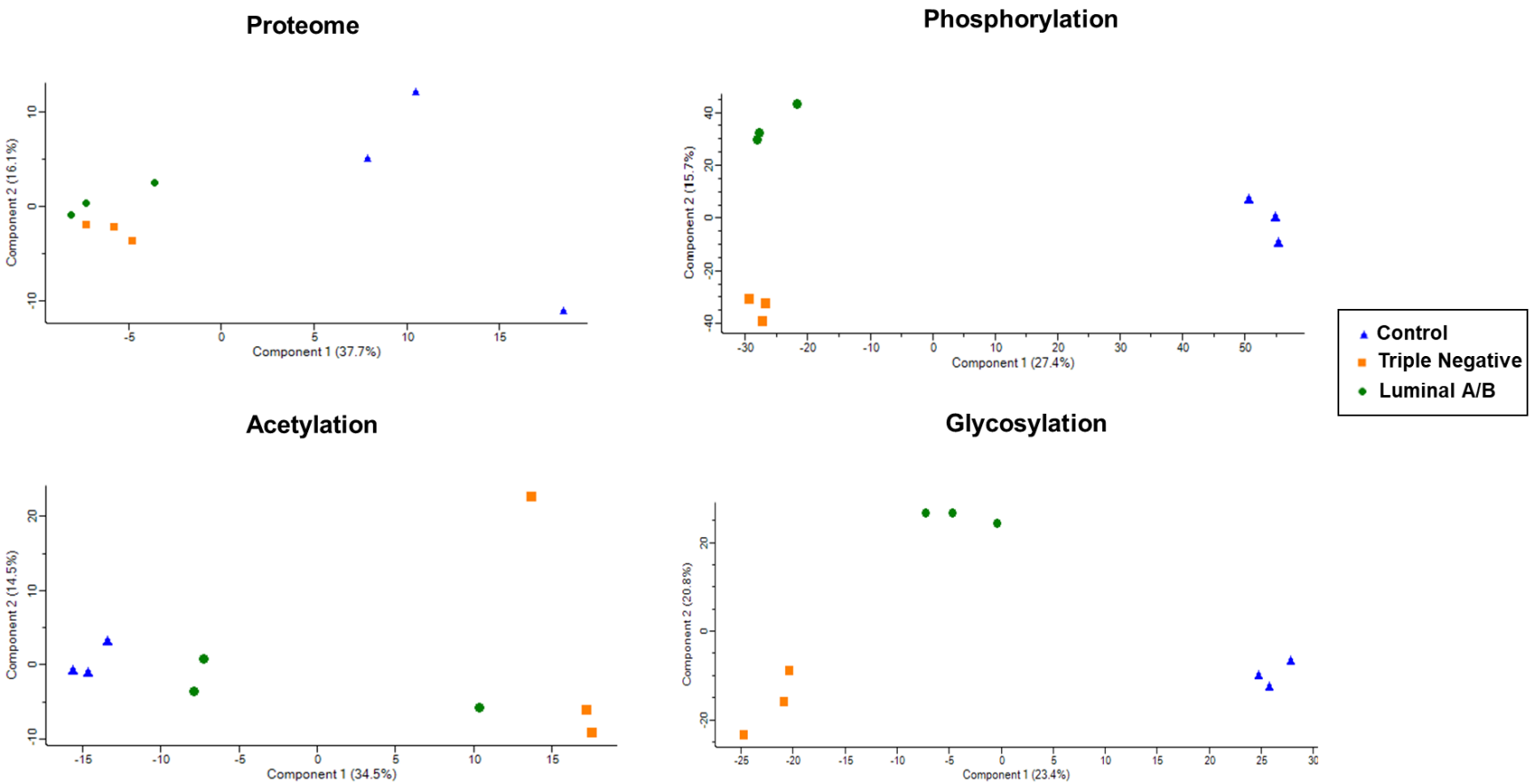

Figure 2.6. Principal component analysis (PCA) of proteome, phosphoproteome, acetylproteome, and glycoproteome data.

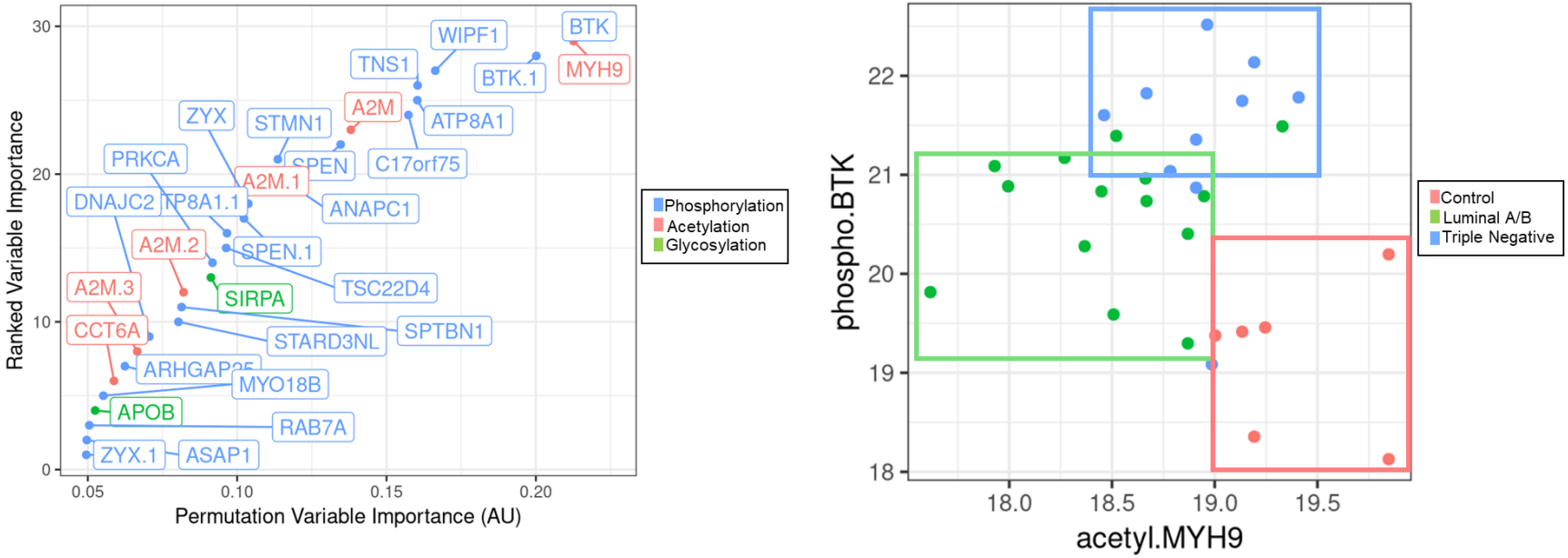

Figure 2.7. Variable importance ranking of the best 30 targets to distinguish breast cancer subtypes and control (left), and scatterplot depicting the log-2 intensities of the top two proteins to distinguish breast cancer subtypes and control (right).
Ranked permutation-based importance for prediction calculated by using a random forest. Scatterplot shows log-2 intensities of the top two variables ranked in importance by using random forest.

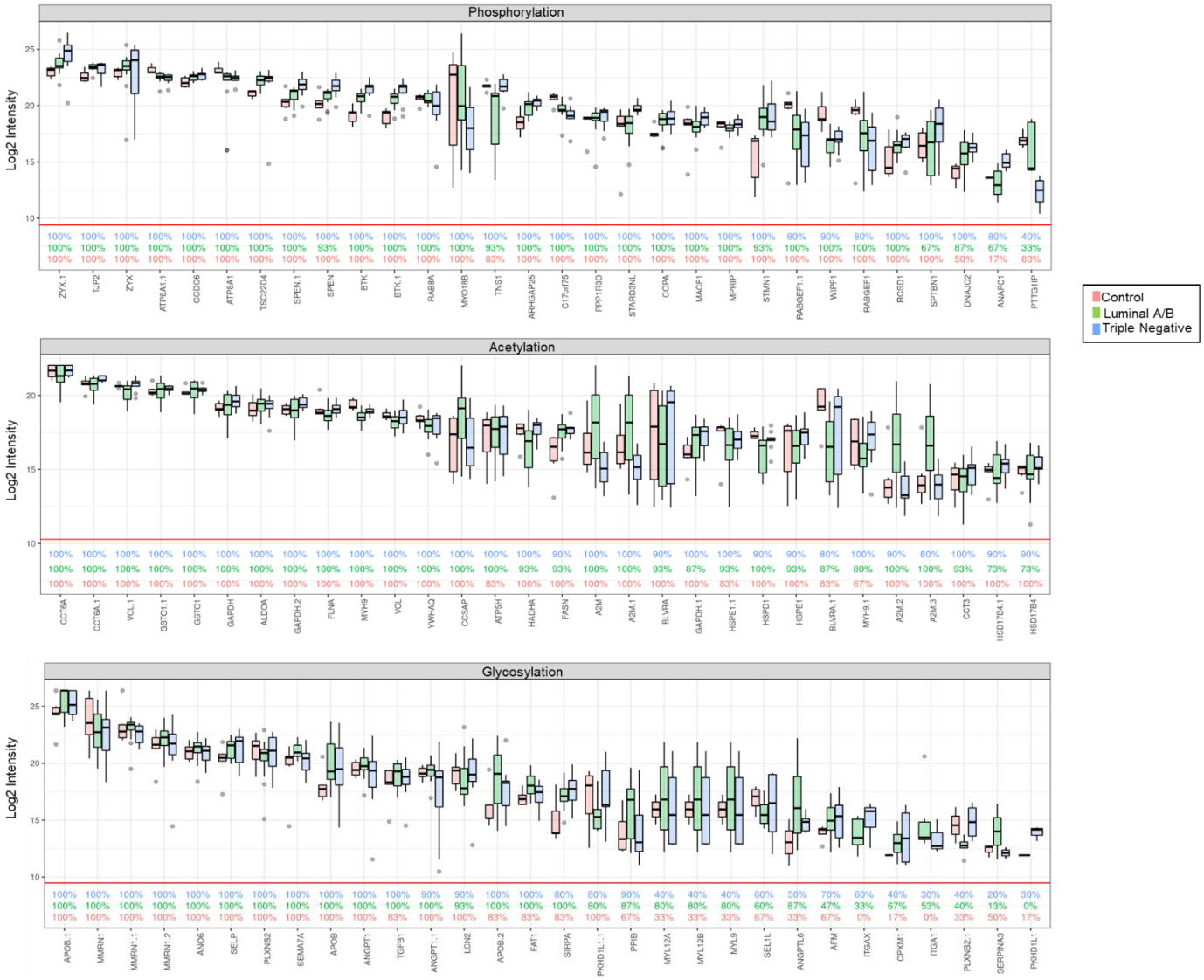

Figure 2.8. Boxplots of top 30 targets from PRM approach per modification showing percentages (%) of individuals with detectable levels of each specific target.
Targeted proteomics was performed to verify possible candidates using 44 individual samples

CHAPTER 3. PLASMA-DERIVED EXTRACELLULAR VESICLES ANALYSIS OF AUTOIMMUNE DISEASE PATIENTS BY TANDEM MASS TAG (TMT) QUANTITATIVE PROTEOMICS AND PHOSPHOPROTEOMICS

3.1 Summary

Extracellular vesicles (EVs) have emerged as promising sources for the discovery of disease-relevant biomarkers. Our recent studies have established efficient strategies to isolate and analyze EVs from different conditions like cancer in biofluids and verified targets as potential disease biomarkers. In autoimmune diseases, EVs play a crucial role in immune responses and as immunomodulatory agents as they can regulate the function of a wide variety of immune cells, especially in antigen-presenting cells (APCs). Several efforts have been made to study EVs and their cargo in numerous disease models, but very few in autoimmunity. Autoimmune diseases are chronic, have been underexplored especially in the omics area, and their diagnosis and treatment rely on traditional therapy. Therefore, there is a need for efficient methods to elucidate biomarkers that could provide additional layers of information for treatment, diagnosis, and prognosis.

In this study, we demonstrated the feasibility of a pipeline that incorporated plasma EV isolation, phosphopeptide enrichment, and TMT quantification for multiplexing to identify relevant disease-related targets in five autoimmune diseases. Our robust platform allowed for the identification of 12,440 peptides and 1,369 phosphopeptides from which we prioritized 28 and 13 autoimmune disease relevant targets from the proteome and phosphoproteome, respectively. Moreover, relevant signaling pathways associations were examined and gene ontology analysis unveiled the EV proteins profile and functions, which are in line with the findings of the autoimmune literature. Overall, this study offers valuable knowledge on the isolation and analysis of EVs through TMT quantification to detect and analyze candidate disease targets.

3.2 Introduction

Extracellular vesicles (EVs), comprised of microvesicles and exosomes, have emerged as promising sources for the discovery of pathology-related markers[123, 124]. Importantly, EVs are membrane-encapsulated nanoparticles, whose content is protected from external proteases and other enzymes[125]. EVs are released by virtually all cell types and contain proteins, lipids, and DNA

as their cargo from the original cell, and provide critical information about the affected tissue[82, 125]. The growing body of evidence shows that EVs are heavily involved in immune signaling and play an essential role as immunomodulatory agents[126, 127]. EVs can modulate the function of a wide variety of immune cells such as T and natural killer (NK) cells. Nevertheless, the most prominent role of EVs in immune activation is their binding or internalization by antigen-presenting cells (APCs)[126-128]. Therefore, EVs and APC interaction markedly amplifies the immune response. Generally, EVs can initiate immune responses via different pathways, including inflammatory cytokine release, cytokine activation of macrophages and neutrophils, self-antigen generation, and via direct binding or endocytosis by APCs[126, 128, 129]. In addition, EVs can activate toll-like receptors (TLRs) by mimicking damage-associated molecular patterns (DAMPs) and pathogen-associated molecular patterns (PAMPs) which are host biomolecules that can start and prolong inflammatory responses[126, 128]. Thus, EVs play a pivotal role in immune signaling propagation, which makes them great candidates to study autoimmune disorders.

Autoimmune diseases are a set of approximately one hundred immune disorders that affect nearly 7% of the population in the US[130, 131]. They are characterized by the overproduction of autoantibodies which lead to aberrant inflammatory responses causing the damage of the affected tissue[132]. B cell overstimulation and T cell hyperactivity are the main orchestrators for autoantibody generation[132]. The pathogenesis of autoimmune diseases remains largely unknown, although genetic and environmental factors have been linked to their development in certain disorders[131]. Rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), ulcerative colitis (UC), type 1 diabetes (T1D) and asthma are among the most widely studied autoimmune disorders. EVs have been associated with the pathogenesis of RA—characterized by damage and eventual destruction of joints—in several studies that compare RA to non-autoimmune conditions[127, 133-135]. One study investigated the role of platelet derived EVs in RA patients and showed that patients had a significant increase of EVs compared to healthy controls[135]. Moreover, this study reported that this increase in EVs was correlated with disease activity and progression[135]. Another study observed that platelet derived microvesicles were found in the synovial fluid of RA patients but not in the synovial fluid of patients with osteoarthritis[133]. Additionally, this study illustrates the amplifying role of platelet-derived microvesicles in the pathophysiology of RA, and states that they are the most abundant cellular component in the synovial fluid[133].

SLE has also been associated with several EV studies[134, 136, 137]. Some of them have shown increased levels of plasma-derived EVs from SLE patients compared to healthy controls[134, 136]. Another study showed plasma-derived microvesicles of SLE patients were associated with IgG as they contained particles with IgG binding[137]. Additionally, IgG levels were correlated with complement pathway activation and anti-DNA antibodies, suggesting that EVs are critical mediators of autoimmune responses[137].

The classification, diagnosis and prognosis of autoimmune diseases is based on traditional imaging studies, clinical examination, and laboratory testing[138, 139]. Thus, the development of proteomic technologies and platforms to enable the identification of novel biosignatures to classify, diagnose and help with therapeutics, would be highly advantageous as these diseases are chronic and need constant monitoring. Several efforts have shown the potential of proteomic strategies to elucidate relevant disease targets in autoimmune diseases, but this area of study is still underexplored[138]. Currently, there is a need for efficient EV proteomic strategies to better understand autoimmunity. Tandem Mass Tag (TMT) based proteomics offers great benefits to evaluate several samples simultaneously as it allows for multiplexing[140-142]. TMT is based on the premise of chemically labeling each sample of interest with specific isobaric tags to then combine all the samples and subsequently analyze them in a single run[141]. This method allows for high throughput and simultaneous analysis of various samples, therefore reducing technical variability and instrument time. To date, very few studies have evaluated the feasibility of a pipeline that incorporates biofluid EV isolation with TMT quantification[142, 143], especially in the study of autoimmune diseases.

Besides evaluating the proteome, overwhelming evidence from various non-autoimmune disease models suggests that protein post-translational modifications (PTMs), particularly phosphorylation, play crucial roles in disease progression[25, 144]. Nevertheless, protein phosphorylation has only been recently explored in autoimmune diseases. For example, current evidence from T1D has shown that phosphorylation plays an essential role in regulating glucose-mediated insulin secretion[145]. Additionally, key kinases involved in glucose-mediated insulin secretion in diabetic mouse models were detected through mass spectrometry (MS)-based proteomics and phosphoproteomics[145]. In addition, this study was able to reveal novel phosphorylation sites highly involved in insulin secretion control, overall demonstrating the advantages of MS-based proteomics and phosphoproteomics platforms[145].

In this study, we present a novel strategy to analyze the proteome and the phosphoproteome of plasma-derived EVs from autoimmune disease patients by TMT quantification. Previously, we have applied our pipeline to elucidate candidate disease targets on different types of cancer using label-free quantification in biofluids[25, 26, 37, 82, 83]. The current approach focuses on combining our plasma-derived EV isolation protocol with TMT quantification in patients with different autoimmune disorders. Altogether, this study explores the feasibility of multiplexing while elucidating relevant pathology-related targets in five predominant autoimmune diseases.

3.3 Experimental design

3.3.1 Plasma samples

The plasma samples used in this study were obtained from Indiana Biobank. Plasma samples were collected by standard protocol[82]. In brief, blood draw was performed in ethylenediaminetetraacetic acid (EDTA) containing tubes and plasma centrifugation was done within 30 min of blood draw. Samples were spun for 30 min at 3500 rpm to remove platelets and all cell debris. Samples were stored at -80°C until ready for use.

3.3.2 EV purification

Before EV isolation, in collaboration with Dr. Majid Kazemian, plasma samples were subjected to peripheral blood mononuclear cells (PBMCs) isolation which requires to dilute the plasma samples 2-fold with a density gradient medium called Ficoll, which is a polysaccharide. This was done by Dr. Majid Kazemian's lab to isolate T cells and B cells for further analysis. Diluted plasma was given to us for EV analysis. EV isolation was performed according to the reported protocol[82]. In brief, 5 mL of pooled diluted plasma were collected from both healthy individuals (n=5) and patients diagnosed with an autoimmune disorder: asthma (n=2), RA (n=5), T1D (n=5), UC (n=1), and SLE (n=2). Plasma samples were centrifuged at 20,000 xg at 4 °C for 1hr. Pellets were washed with cold PBS and centrifuged again at 20,000 xg at 4 °C for 1 hr, the pellets were microvesicles. Supernatant of the first centrifugation was further centrifuged at 100,000 xg at 4 °C for 1hr. Pellets were washed with cold PBS and centrifuged at 100,000 xg for 1hr again. The pellets from ultra-high-speed centrifugations were exosomes. Two separate isolated EVs were combined during sample lysis.

3.3.3 EV lysis and enzymatic digestion

The enzymatic digestion was performed according to the phase transfer surfactant aided (PTS) digestion protocol with minor modifications[67, 82]. In brief, EVs were solubilized in lysis buffer containing 12mM sodium deoxycholate (SDC), 12mM sodium lauroyl sarcosinate (SLS) and phosphatase inhibitor cocktail in 100mM triethylammonium bicarbonate (TEAB). Proteins were reduced and alkylated with 10 mM tris-(2-carboxyethyl)phosphine (TCEP) and 40 mM chloroacetamide (CAA) at 95 ℃ for 5 min. Alkylated proteins were diluted to 5 fold by 50mM TEAB and digested with Lys-C (Wako) in a 1:100 (w/w) enzyme-to-protein ratio for 3 hr at 37 ℃. Trypsin was added to a final 1:50 (w/w) enzyme-to-protein ratio for overnight digestion. The digested peptides were acidified with trifluoroacetic acid (TFA) to final concentration of 0.5% TFA, and 250 µL of ethyl acetate was added to 250 µL digested solution. The mixture was shaken for 2 min, then centrifuged at 13,200 rpm for 2 min to obtain aqueous and organic phases. The aqueous phase was collected and desalted using Top-Tip C18 tips (Glygen) according to the manufacturer's instructions. Bicinchoninic acid (BCA) analysis was then performed according to manufacturer's instructions (Pierce) to determine peptide concentration, and then samples were dried in a vacuum centrifuge.

3.3.4 TMT labeling and fractionation

TMT protocol was performed according to manufacturer's instructions (Thermo). TMT labeling reagents were reconstituted in 100% anhydrous acetonitrile and digested peptides were dissolved in 75 µL 50mM HEPES pH:8.5. For labeling, 0.6 mg of each tag was used to label 40 µg of digested peptides. Samples were incubated with the respective reconstituted TMT labeling reagent for 1 hour: 127C for healthy control, 128C for asthma, 129C for RA, 130N for T1D, 130C for SLE and 131 for UC. The reaction was quenched with 6 µL of 5% hydroxylamine for 15 min, pooled together and dried under vacuum centrifuge. Peptides were then subjected to basic pH reverse phase peptide fractionation using C18 stage tips and then dried under vacuum centrifuge.

3.3.5 PolyMAC phosphopeptide enrichment

Peptides were enriched using PolyMAC phosphopeptides enrichment kit (Tymora Analytical) according to manufacturer's instructions. In brief, peptides were resuspended in 200

μL of loading buffer and incubated with PolyMAC-Ti silica beads for 20 min. The beads were loaded into the tip with frit to remove the flow-through. The beads were then washed with 200 μL of loading buffer, then washing buffer 1 and once with washing buffer 2. The phosphopeptides were then eluted twice from the beads with 50 μL of the elution buffer. The eluates were collected and dried under vacuum centrifuged.

3.3.6 LC-MS/MS

Dried peptides and phosphopeptides were dissolved in 10.5 μL of 0.05% trifluoroacetic acid with 3% (vol/vol) acetonitrile and 10 μL of each sample was injected into an Ultimate 3000 nano UHPLC system (Thermo Fisher Scientific). Peptides were captured on a 2-cm Acclaim PepMap trap column and separated on a heated 50-cm Acclaim PepMap column (Thermo Fisher Scientific) containing C18 resin. The mobile phase buffer consisted of 0.1% formic acid in ultrapure water (buffer A) with an eluting buffer of 0.1% formic acid in 80% (vol/vol) acetonitrile (buffer B) run with a linear 60-min gradient of 6–30% buffer B at flow rate of 300 nL/min. The UHPLC was coupled online with a Q-Exactive HF-X mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in the data-dependent mode, in which a full-scan MS, from m/z 375 to 1,400 with the resolution of 60,000, was followed by MS/MS of the 15 most intense ions. For MS2 parameters resolving power was set at 60,000, AGC target 1e5, a maximum injection time of 100 ms and a normalized collision energy (NCE) of 32.

3.3.7 Data processing

The raw files were searched directly against the human Swiss-Prot database updated on July 16, 2019 with no redundant entries using Sequest search engines loaded into Proteome Discoverer 2.3 software (Thermo Fisher Scientific). MS1 precursor mass tolerance was set at 10 ppm, and MS2 tolerance was set at 20ppm. Normalization was done using total peptide amount and quantification using reporter ions (node in Proteome Discoverer 2.3 software). Search criteria included a static carbamidomethylation of cysteines (+57.0214 Da) and a static TMT in lysine residues (+229.163 Da). Variable modifications included TMT (+229.163 Da) at the peptide N terminus, oxidation (+15.9949 Da) on methionine residues, acetylation (+42.011 Da) at N terminus of proteins, and phosphorylation (+79.996 Da) on serine, threonine, or tyrosine residues. Search

was performed with full trypsin/P digestion and allowed a maximum of two missed cleavages on the peptides analyzed from the sequence database. The false-discovery rates of proteins and peptides were set at 0.01. All protein and peptide identifications were grouped, and any redundant entries were removed. Only unique peptides and unique master proteins were reported.

3.3.8 Quantitative data analysis

Data was analyzed using the Perseus software (version 1.6.14)[94, 146]. For both proteomic and phosphoproteomic data, the intensities of proteins and phosphoproteins were extracted from Proteome Discoverer search results, and the missing values of intensities were replaced by normal distribution with a downshift of 1.8 SDs and a width of 0.3 SDs. For heatmap, the imputed data set was normalized by z-score within each dataset. For gene ontology analysis and visualization, PANTHER classification system (v.14.0)[147] and Prism were used. Functional enrichment analysis of the different autoimmune disorders was done using ToppCluster[148] tool and Cytoscape for visualization.

3.4 Results

3.4.1 Identification of 12,440 and 1,369 unique peptides and phosphopeptides from plasma EVs

To identify novel biosignatures we combined our plasma-derived EV isolation protocol, previously reported[82], with TMT quantification in patients with different autoimmune disorders (Figure 3.1). Plasma samples were collected from healthy controls (n=5), RA (n=5), T1D (n=5), asthma (n=2), SLE (n=2) and UC (n=1). Each pool consisted of 5 mL of diluted plasma. Human plasma was sequentially centrifuged at high speed (20,000 xg) and ultra-high speed (100,000 xg) to isolate microvesicles and exosomes, respectively. EVs were lysed, proteins extracted, denatured, alkylated and enzymatically digested using Lys-C and trypsin through a phase transfer surfactant (PTS) digestion protocol which allows for fewer missed tryptic sites as reported previously[67, 82].

TMT labeling was performed according to manufacturer's instructions (Thermo) in which digested peptides were resuspended in 50mM HEPES pH:8.5 and 0.6 mg of the TMT labeling reagent was used to label 40 µg of digested peptides. TMT reaction was quenched with 5% hydroxylamine and samples were pooled together. Fractionation was done after TMT labeling as several studies have shown how the fractionation of EV samples using a C18 stage tip could

significantly increase the identification of proteins[149-151]. In our platform, we performed basic pH reverse phase peptide fractionation using ACN and ammounium formate as the solvents. We collected 8 fractions from which we took 5% for proteome analysis and enriched the rest for phosphorylated peptides using PolyMAC (Tymora Analytical). Both set of fractions were analyzed by LC-MS/MS on a high-resolution and high-speed mass spectrometer to determine differential expression between the different autoimmune disorders and healthy controls. This pipeline allowed for the identification of 12,440 peptides corresponding to 2,752 proteins in the proteome, and 1,369 phosphopeptides corresponding to 775 phosphoproteins in the phosphoproteome. In addition, TMT labeling efficiency for both the proteome and phosphoproteome was 99% with 12,364 and 2,579 labeled peptides in the proteome and phosphoproteome, respectively. The selectivity for the phosphorylation method—ratio of enriched phosphopeptides over total identified peptides—was 54% which compares to studies isolating and analyzing the phosphoproteome in EVs isolated from biofluids such as plasma or urine[37, 82, 83].

3.4.2 EV purity, GO analysis and TMT fractionation evaluation

To assess EV purity and isolation, we listed 10 common exosome markers detected in our experiment. Our proteins included common exosome markers such as CD9, CD81, CD63 and TSG101 which have been used in several studies to assess the EV isolation method purity[82, 83, 152-154]. Additionally, we also displayed the amount of plasma contaminant proteins in our study (Figure 3.2). The contamination level and markers correlate with previous studies and occurs to some extent due to nonspecific binding to exosome surface proteins[37, 154]. Thus, we expect that a portion of free plasma proteins were captured together with EVs. Moreover, we compared our data against publicly available manually curated EV databases, Vesiclepedia and Exocarta[155, 156]. Results showed than more than 90% of our identified proteins overlap with either EV database or both (Figure 3.2). Only 204 proteins did not overlap, which represents 8% of our total identified proteins.

We also examined the gene ontology of our identified proteins to better understand the role and profiles of our identified EV proteins (Figure 3.3). Most of our EV proteins were identified as part of the cellular subcomponent, and for the biological process, most were identified involved in cellular processes. Moreover, a good portion of these proteins were also categorized as being involved in signaling pathways, response to stimulus and metabolic processes, which is in

congruence with findings in autoimmune diseases[157-159]. EV proteins were also categorized based on their class. Most of these EV proteins were identified to be metabolite interconversion enzymes such as transferases, or hydrolases. Another significant portion (nearly 100 proteins) were identified as defense/immunity proteins, which would be expected to be present in patients of with these conditions[132, 160]. To evaluate our fractionation method after TMT labeling, we identified the number of peptides eluted in each fraction (Figures 3.4 and 3.5). For the proteome, fractions 4-8 were the ones that contributed the most to the identified EV proteins. This can be explained as greater numbers of peptides are eluted with increased hydrophobicity of the solvent. In this case we started with 4% ACN and 96% 200mM ammonium formate and then increased the ACN percentage (7, 10, 13, 16, 19, 22, 40) while decreasing the salt percentage (93, 90, 87, 84, 81, 78, 60). In the context of the phosphoproteome, we did not see a clear trend as in the proteome. We observed fractions 1, 4, 5 and 8 as being the major contributors but no clear correlation was observed. This can be explained as phosphorylated peptides vary in terms of hydrophobicity or hydrophilicity. Phosphorylation introduces a hydrophilic group in the side of amino acids, usually changing a protein's structure by modifying adjacent amino acids. Some known phosphorylated proteins such as p53 have multiple phosphorylation sites, changing its confirmation and overall structure in terms of hydrophobicity and hydrophilicity[161].

3.4.3 Autoimmune disease specific targets in EVs

Quantification of plasma-derived EVs by TMT in five autoimmune diseases and healthy controls allowed us to identify differentially expressed targets in both the proteome and phosphoproteome (Figure 3.6). In the proteome, 123 targets significantly changed in asthma, 89 in RA, 127 in T1D, 136 in SLE and 92 in UC. At the phosphoproteome level, 48 targets were significantly changed in asthma, 46 in RA, 65 in T1D, 57 in SLE and 89 in UC. The targets from each group were subjected to clustering according to their functional pathway's association using ToppCluster[148] (Figures 3.7 and 3.8). In the proteome, asthma, SLE and UC had the most hits in terms of pathway clustering association. Among autoimmune disease literature relevant pathways, we can observe that our pipeline was able to elucidate pathology-relevant associations. Some of the most relevant pathways in our clustering were MHC class II antigen presentation, innate immune system, neutrophil degranulation, MAPK signaling, signaling by the B cell receptor, antigen presentation, adaptive immune system, and cytokine signaling in immune system. All of

these pathways have been associated in several studies in autoimmune disorders, especially related to the diseases in this study[126, 128, 132].

At the phosphoproteome level, in contrast with the proteome, T1D and RA showed significant pathway clustering associations. Some of the most relevant pathways, also relevant in the autoimmune disease literature, were insulin responsive facilitative sugar transporter mediated glucose transport, innate immune system, and neutrophil degranulation[126, 128, 145]. Particularly, the insulin responsive facilitative sugar transporter mediated glucose transport, has been investigated through phosphoproteomics[145]. More specifically, a study showed the crucial role of phosphorylation in regulating glucose-mediated insulin secretion by detecting key kinases using plasma from T1D mouse models through MS based phosphoproteomics[145]. It is important to note that this pathway association was not detected in the proteome but only in the phosphoproteome, further confirming our pipeline's capability on elucidating relevant pathology-related targets.

Based on our previous analysis, we prioritized a panel of 28 and 13 targets in the proteome and phosphoproteome, respectively, based on their specific relevance in the autoimmune disease literature (Figures 3.9 and 3.10). In the proteome, we prioritized targets like GSTO1, CD40, GAPDH, and TRIM21 (Figure 3.9). Glutathione transferase omega 1 (GSTO1) has been found to play a pro-inflammatory role in mouse models of colitis[162]. This study also reveals the therapeutic potential of GSTO1 inhibitors in the modulation of inflammation, more specifically, in UC[162]. Another relevant target is CD40 which is a tumor necrosis factor receptor family member expressed by non-immune and immune cells[163]. CD40 mediates T cell priming and T-dependent B cell responses, therefore, it has been studied as a candidate target in autoimmune diseases as it plays crucial roles in B cell overstimulation and T cell hyperactivity[163]. Autoimmune diseases associated with high levels of CD40 are T1D, UC, RA, SLE, psoriasis and multiple sclerosis[163]. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) has also been reported as a biomarker in autoimmune diseases, more specifically, SLE[164]. A study showed anti-GAPDH autoantibody was increased in the serum of 130 SLE patients compared to 55 healthy controls, and that concentrations of these autoantibodies correlated with disease severity[164]. This was measured using ELISA in patients with SLE with and without neuropsychiatric symptoms[164]. Evidence has also shown the role of TRIM21 in autoimmunity[165]. More specifically, tripartite motif containing 21 (TRIM21), has been found relevant in autoimmune diseases like SLE and Sjögren's syndrome

as its autoantigens react with anti-SS-A antibody (Ab) present in patients with SLE and Sjögren's syndrome[165].

In the phosphoproteome we prioritized targets like CYBA, TOM1 and MAPK4 (Figure 3.10). Cytochrome B-245 alpha chain (CYBA), has been specifically found to be associated with T1D and diabetic peripheral neuropathy in adolescents and children[166]. More specifically, this study found polymorphisms of CYBA in a cohort of 90 T1D vs a healthy control cohort[166]. TOM1, target of Myb1 membrane trafficking protein, has also been associated with early onset of autoimmunity, as it participates in immune receptor recycling and inhibits TLR signaling[167]. MAPK4, although involved in nearly every disease model, in autoimmune disorders activates the FOS gene (Fos proto-oncogene) which is a hallmark in Crohn's disease, UC, multiple sclerosis, and T1D[158]. Overall, these prioritized targets in both the proteome and phosphoproteome, have been associated in different ways to autoimmunity and autoimmune disorders.

3.5 Discussion

Recently, EVs have gained attention as they are capable of transferring disease-related signaling molecules and can be isolated from nearly any biofluid[89, 124, 125]. Several efforts have been made to study EVs and their cargo in numerous disease models, but very few in autoimmunity. Given that EVs play a crucial role in immune responses and as immunomodulatory agents, it is critical to explore their potential for treating or inhibiting inflammatory responses[127, 129]. Moreover, autoimmune diseases are chronic, diagnosis and treatment rely on traditional therapy, and they have been underexplored, especially in the omics area. Therefore, there is a need for efficient methods to elucidate biomarkers that could provide additional layers of information for treatment, diagnosis, and prognosis. Existing data have proposed EVs as a drug delivery system as they can cross the blood brain barrier and the synovial membrane[126, 128, 133]. Other studies have reinforced the idea of therapeutic actions to reduce the load of EVs in autoimmune diseases as they amplify the immune responses[126, 128]. Altogether, these requires platforms for efficiently examining EV cargo for generating the most efficient therapeutic strategies.

In this study, we demonstrated the feasibility of a pipeline that incorporated biofluid EV isolation with TMT quantification for multiplexing to elucidate relevant disease-related targets in five autoimmune diseases. For both the proteome and phosphoproteome, the TMT labeling efficiency was 99%. Additionally, more than 90% of our identified EV proteins overlapped with

the EV proteins from known, curated EV databases and we were able to detect common and established exosome marker proteins in our data. Our robust platform allowed for the identification of 12,440 peptides and 1,369 phosphopeptides from which we prioritized 28 and 13 targets from the proteome and phosphoproteome, respectively. Moreover, relevant signaling pathways associations were examined and gene ontology analysis revealed the EV proteins profile and functions. Overall, this study offers pivotal information on the isolation and analysis of EVs through TMT quantification to detect and analyze candidate disease targets.

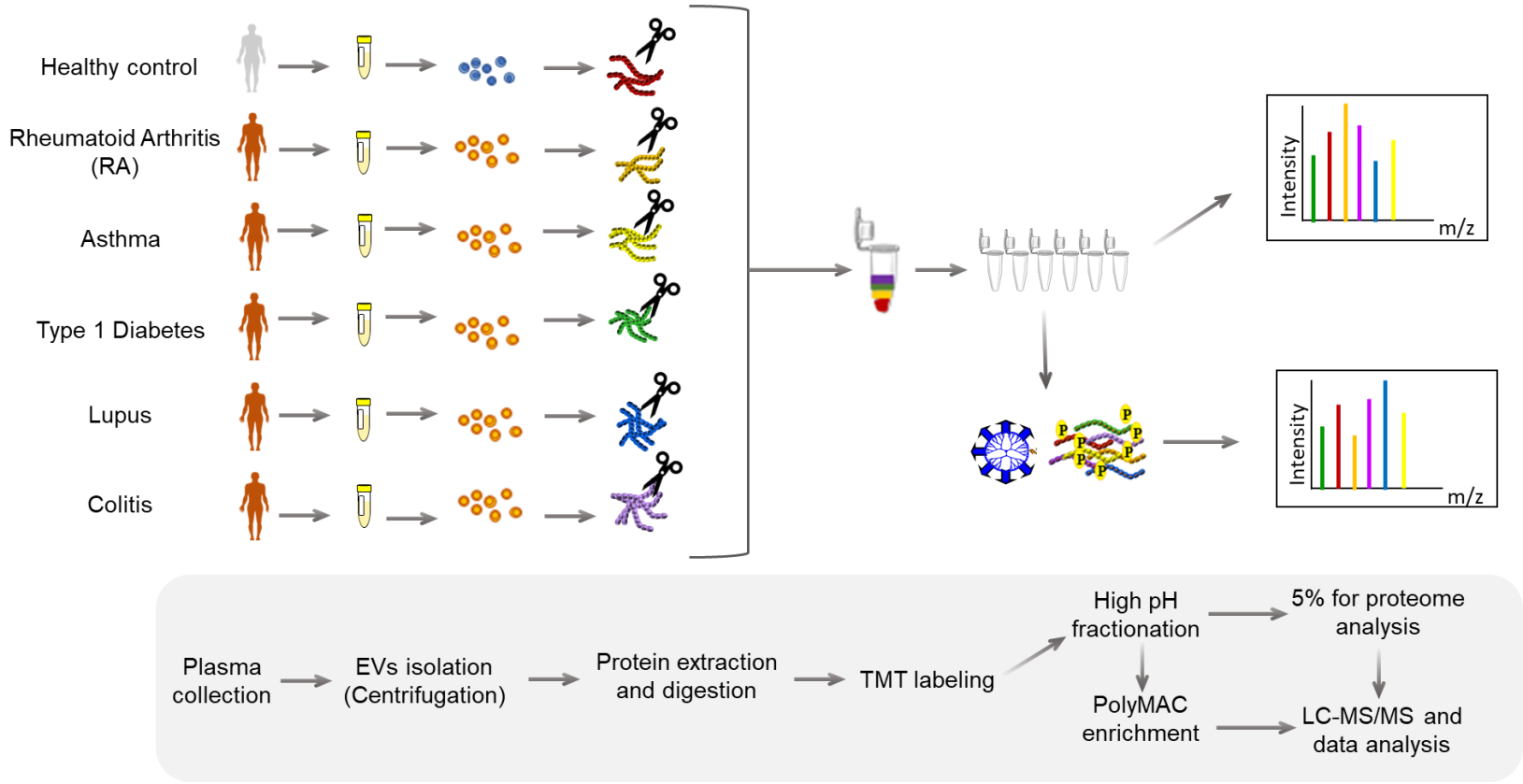

Figure 3.1. Workflow of the EV-TMT pipeline to isolate and analyze EVs in autoimmune disease patients.

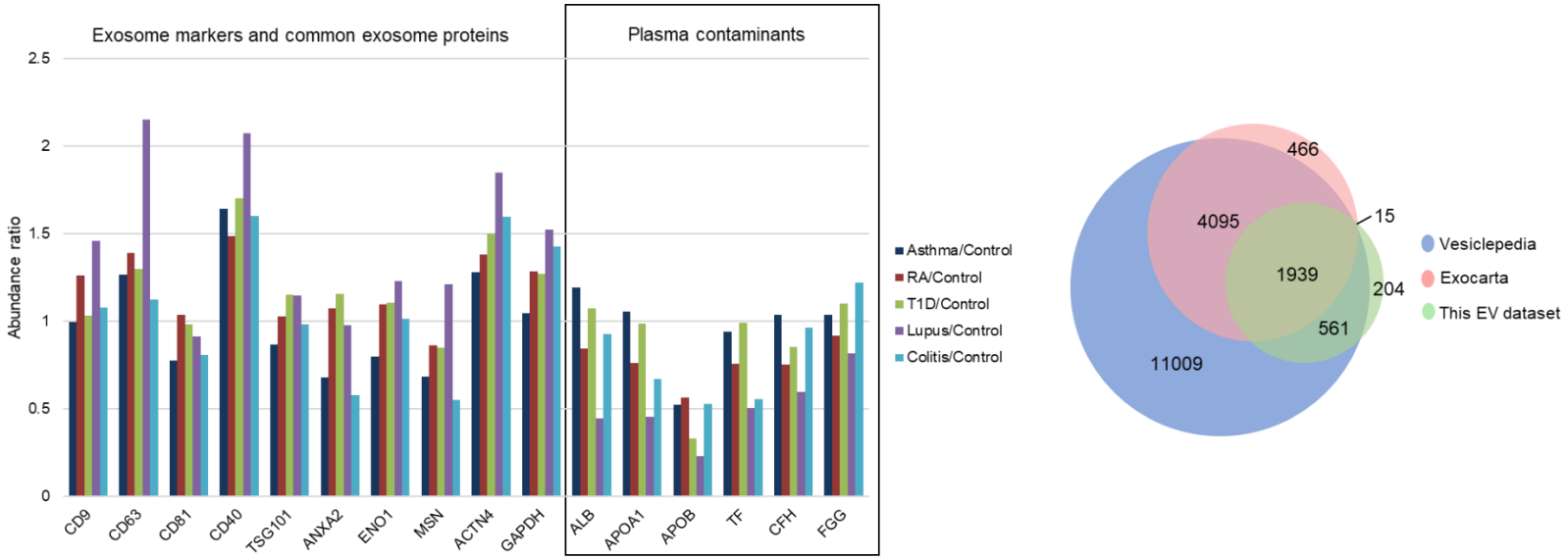

Figure 3.2. Quantitation of 10 common exosome proteins and 6 common plasma protein contaminants (left) and Venn diagram showing the overlap between our EV dataset, Exocarta and Vesiclepedia, both EV databases (right).

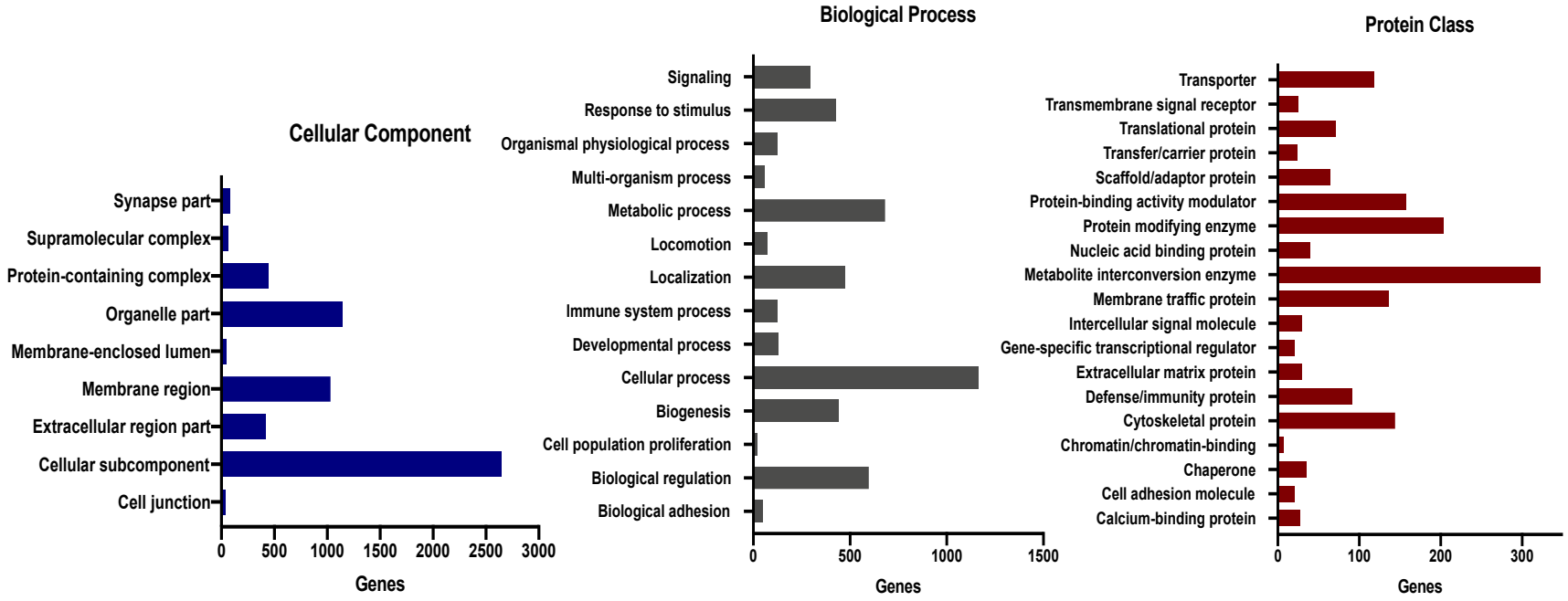

Figure 3.3. Gene ontology analysis of the EV proteins: cellular component, biological process, and protein class.

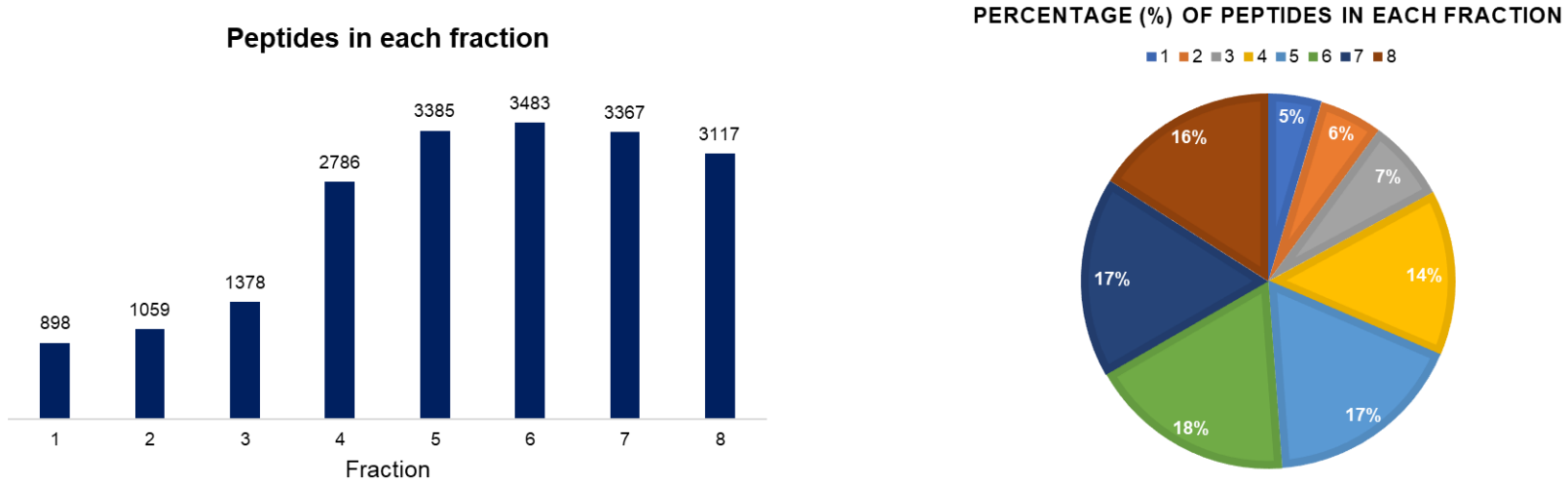

Figure 3.4. Number and percentage (%) of peptides isolated from each fraction.

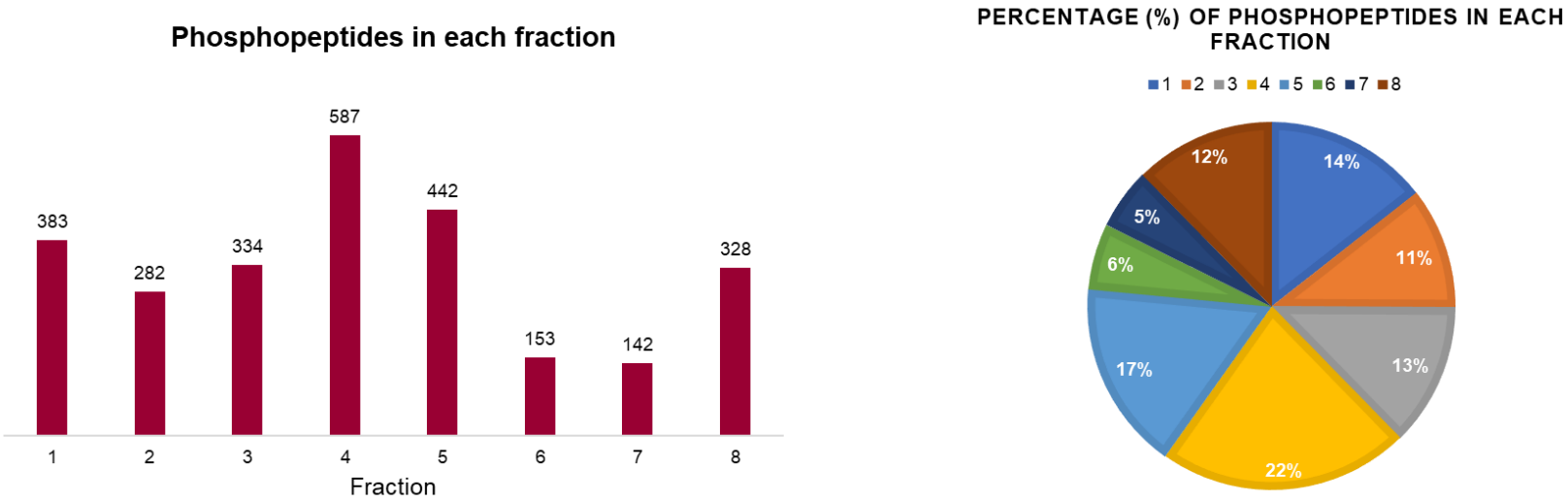

Figure 3.5. Number and percentage (%) of phosphopeptides isolated from each fraction.

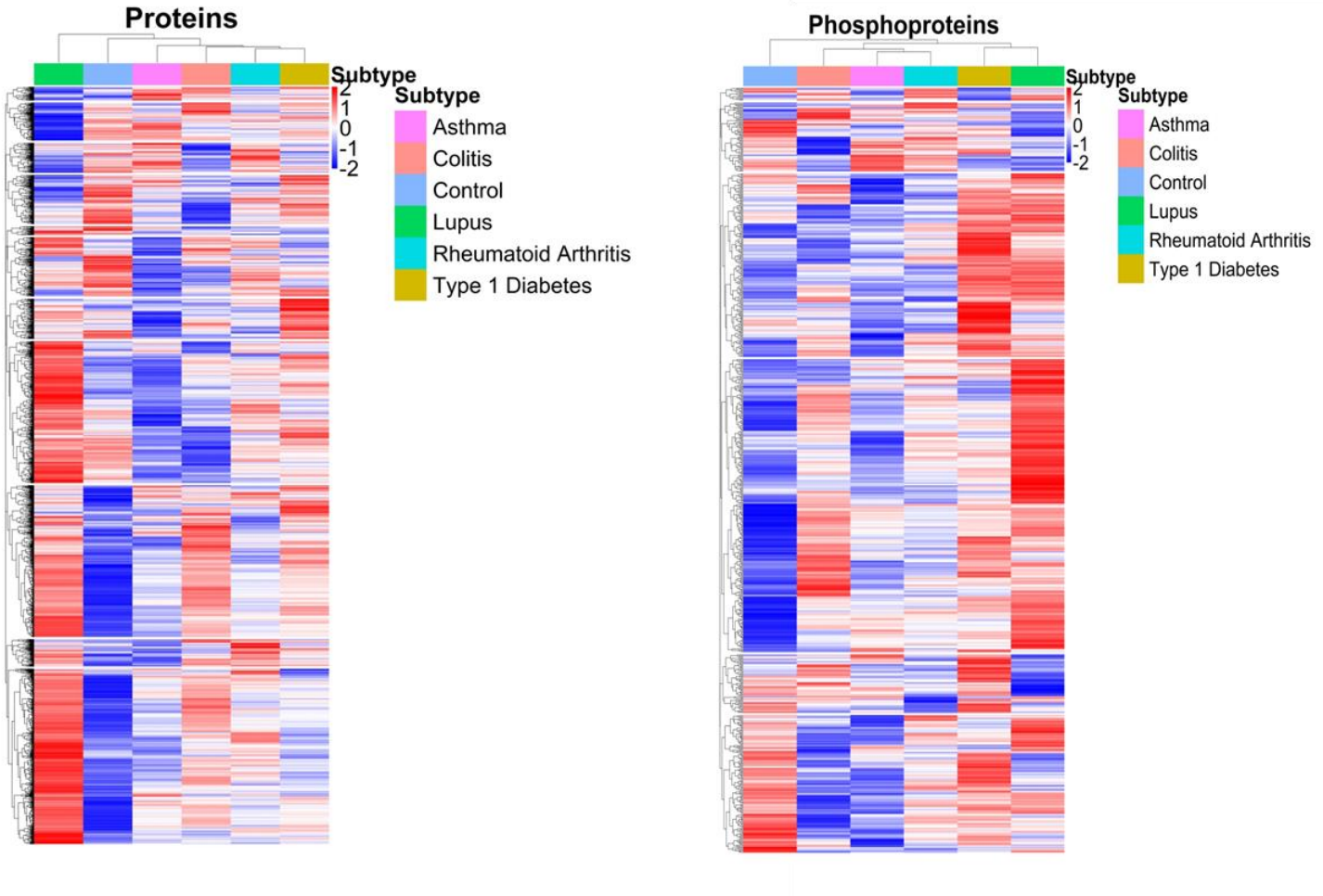

Figure 3.6. Heatmaps showing quantitative proteomics and phosphoproteomics by TMT between healthy controls and five autoimmune diseases.

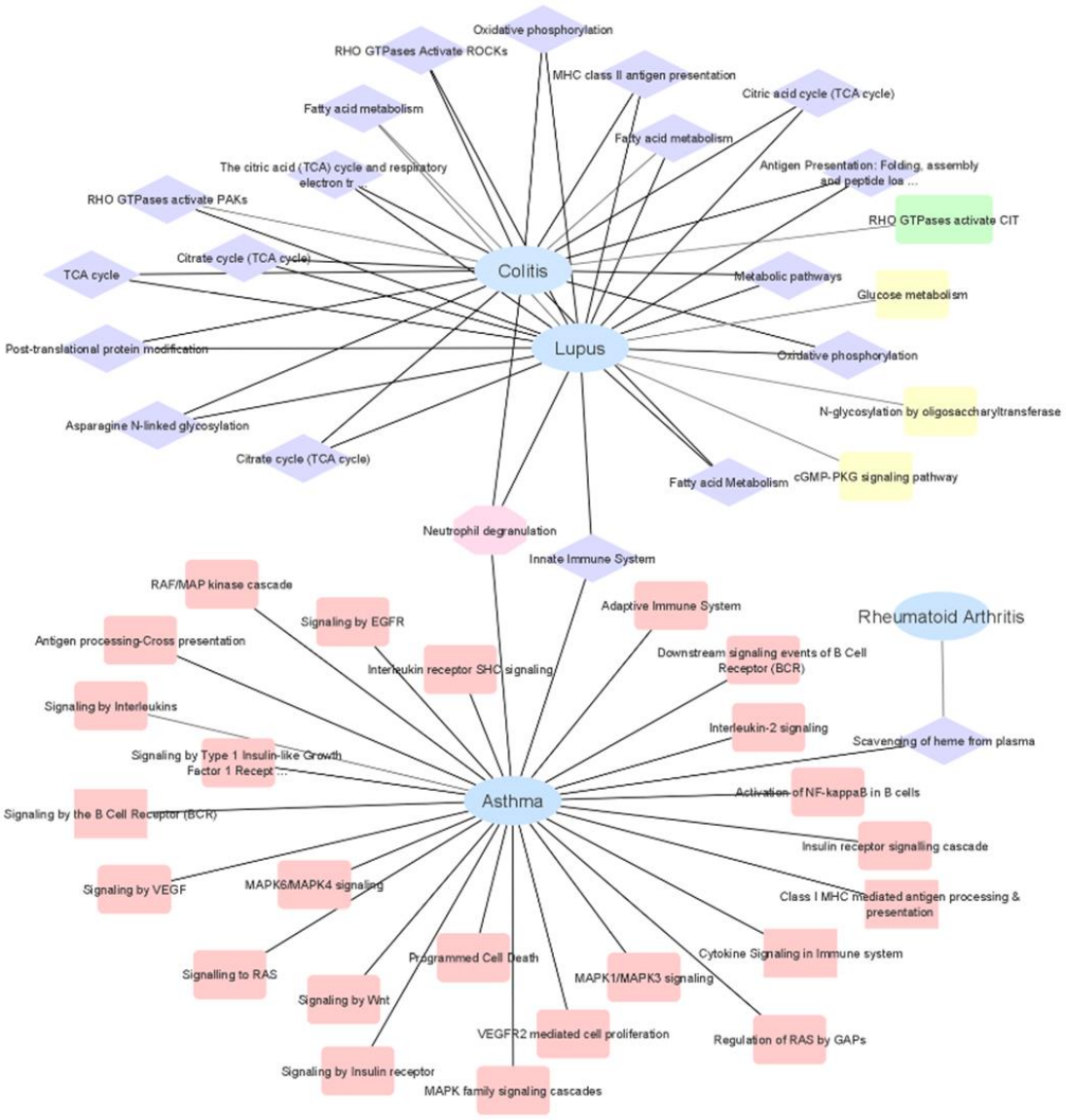

Figure 3.7. Clustering analysis of the EV proteome according to their functional pathway's association using ToppCluster.

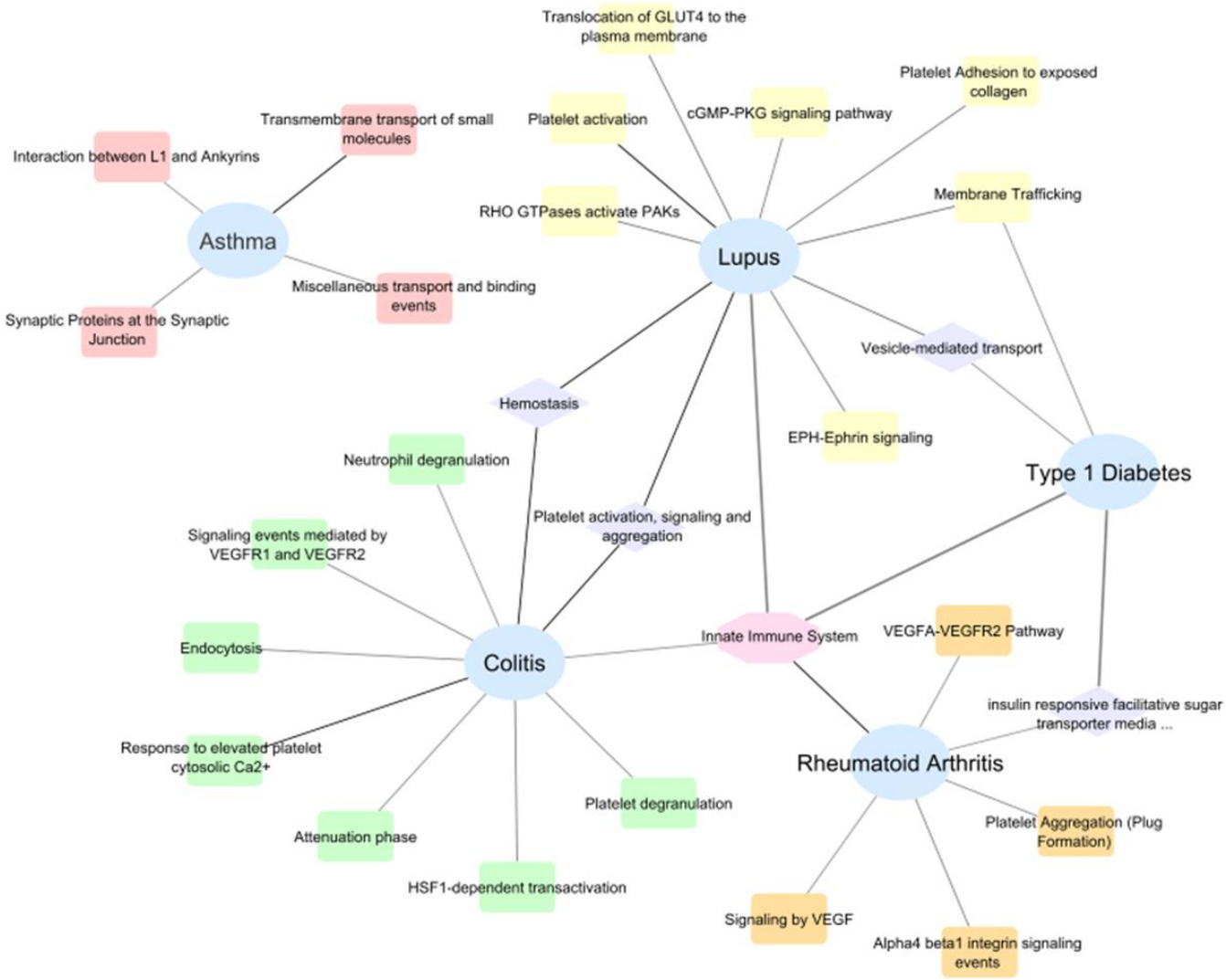

Figure 3.8. Clustering analysis of the EV phosphoproteome according to their functional pathway's association using ToppCluster.

.

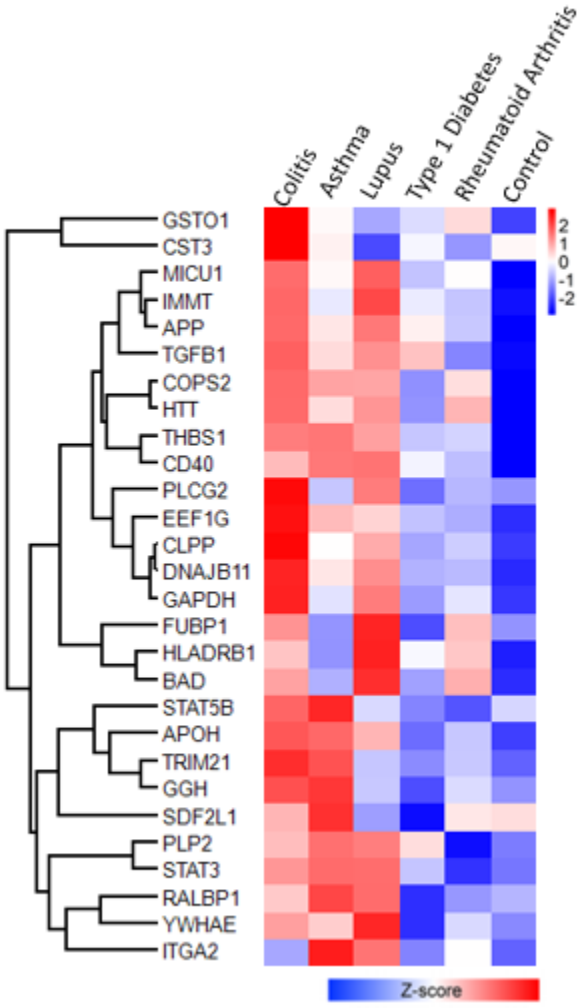

Figure 3.9. Prioritized panel of 28 EV proteins relevant in autoimmune diseases.

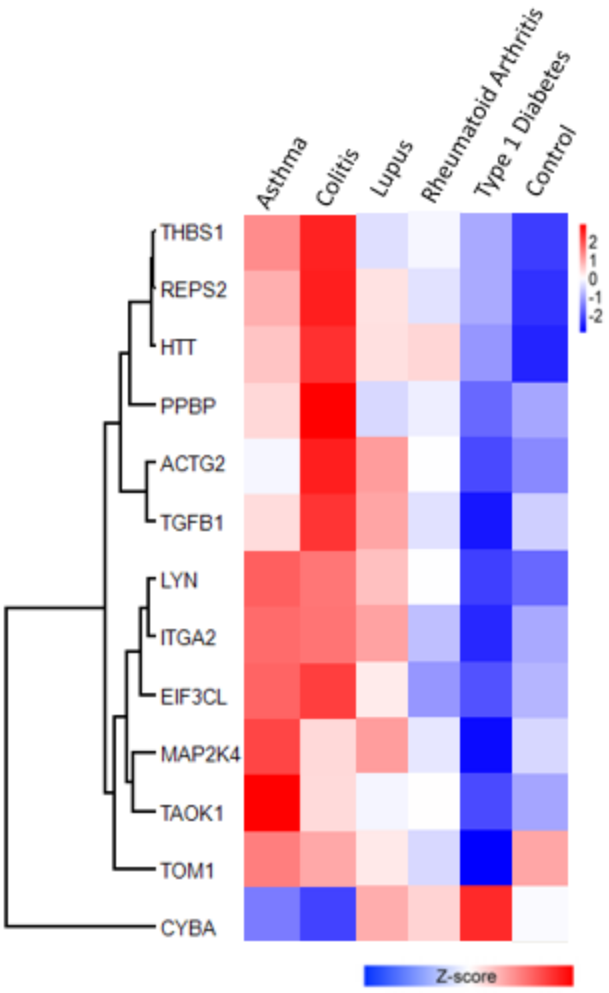

Figure 3.10. Prioritized panel of 13 EV phosphoproteins relevant in autoimmune diseases.

REFERENCES

1. Harel, M., Oren-Giladi, P., Kaidar-Person, O., Shaked, Y. & Geiger, T. Proteomics of microparticles with SILAC Quantification (PROMIS-Quan): a novel proteomic method for plasma biomarker quantification. *Mol Cell Proteomics* **14**, 1127-1136 (2015).

2. Milane, L., Singh, A., Mattheolabakis, G., Suresh, M. & Amiji, M.M. Exosome mediated communication within the tumor microenvironment. *J Control Release* **219**, 278-294 (2015).

3. Cocucci, E. & Meldolesi, J. Ectosomes and exosomes: shedding the confusion between extracellular vesicles. *Trends in cell biology* **25**, 364-372 (2015).

4. An, T. et al. Exosomes serve as tumour markers for personalized diagnostics owing to their important role in cancer metastasis. *Journal of extracellular vesicles* **4**, 27522 (2015).

5. Dobrowolski, R. & De Robertis, E.M. Endocytic control of growth factor signalling: multivesicular bodies as signalling organelles. *Nat Rev Mol Cell Biol* **13**, 53-60 (2011).

6. Melo, S.A. et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature* **523**, 177-182 (2015).

7. Gonzales, P.A. et al. Large-scale proteomics and phosphoproteomics of urinary exosomes. *Journal of the American Society of Nephrology : JASN* **20**, 363-379 (2009).

8. Boukouris, S. & Mathivanan, S. Exosomes in bodily fluids are a highly stable resource of disease biomarkers. *Proteomics Clin Appl* **9**, 358-367 (2015).

9. Wu, A.Y., Ueda, K. & Lai, C.P. Proteomic Analysis of Extracellular Vesicles for Cancer Diagnostics. *Proteomics* **19**, e1800162 (2019).

10. Xu, R., Greening, D.W., Zhu, H.J., Takahashi, N. & Simpson, R.J. Extracellular vesicle isolation and characterization: toward clinical application. *J Clin Invest* **126**, 1152-1162 (2016).

11. Bandu, R., Oh, J.W. & Kim, K.P. Mass spectrometry-based proteome profiling of extracellular vesicles and their roles in cancer biology. *Exp Mol Med* **51**, 30 (2019).

12. Emmanouilidi, A., Paladin, D., Greening, D.W. & Falasca, M. Oncogenic and Non-Malignant Pancreatic Exosome Cargo Reveal Distinct Expression of Oncogenic and Prognostic Factors Involved in Tumor Invasion and Metastasis. *Proteomics* **19**, e1800158 (2019).

13. Hurwitz, S.N. & Meckes, D.G., Jr. Extracellular Vesicle Integrins Distinguish Unique Cancers. *Proteomes* **7** (2019).

14. Bae, S., Brumbaugh, J. & Bonavida, B. Exosomes derived from cancerous and non-cancerous cells regulate the anti-tumor response in the tumor microenvironment. *Genes Cancer* **9**, 87-100 (2018).

15. Ghosh, A. et al. Rapid isolation of extracellular vesicles from cell culture and biological fluids using a synthetic peptide with specific affinity for heat shock proteins. *PLoS One* **9**, e110443 (2014).

16. Lobb, R.J. et al. Optimized exosome isolation protocol for cell culture supernatant and human plasma. *Journal of extracellular vesicles* **4**, 27031 (2015).

17. Gallart-Palau, X. et al. Extracellular vesicles are rapidly purified from human plasma by PRotein Organic Solvent PRecipitation (PROSPR). *Sci Rep* **5**, 14664 (2015).

18. Moreno-Gonzalo, O., Villarroya-Beltri, C. & Sanchez-Madrid, F. Post-translational modifications of exosomal proteins. *Front Immunol* **5**, 383 (2014).

19. Zhang, Y., Wu, X. & Andy Tao, W. Characterization and Applications of Extracellular Vesicle Proteome with Post-Translational Modifications. *Trends Analyt Chem* **107**, 21-30 (2018).

20. Gerlach, J.Q. & Griffin, M.D. Getting to know the extracellular vesicle glycome. *Molecular bioSystems* **12**, 1071-1081 (2016).

21. Oeyen, E. et al. Bladder Cancer Diagnosis and Follow-Up: The Current Status and Possible Role of Extracellular Vesicles. *Int J Mol Sci* **20** (2019).

22. Mann, M. & Jensen, O.N. Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21**, 255-261 (2003).

23. Aebersold, R. & Goodlett, D.R. Mass spectrometry in proteomics. *Chem Rev* **101**, 269-295 (2001).

24. Kettenbach, A.N., Rush, J. & Gerber, S.A. Absolute quantification of protein and post-translational modification abundance with stable isotope-labeled synthetic peptides. *Nat Protoc* **6**, 175-186 (2011).

25. Chen, I.H. et al. Phosphoproteins in extracellular vesicles as candidate markers for breast cancer. *Proc Natl Acad Sci U S A* (2017).

26. Chen, I.H. et al. Analytical Pipeline for Discovery and Verification of Glycoproteins from Plasma-Derived Extracellular Vesicles as Breast Cancer Biomarkers. *Anal Chem* **90**, 6307-6313 (2018).

27. Jaros, J.A. et al. Clinical use of phosphorylated proteins in blood serum analysed by immobilised metal ion affinity chromatography and mass spectrometry. *J Proteomics* **76 Spec No.**, 36-42 (2012).

28. Hu, L. et al. Profiling of endogenous serum phosphorylated peptides by titanium (IV) immobilized mesoporous silica particles enrichment and MALDI-TOFMS detection. *Anal Chem* **81**, 94-104 (2009).

29. Sokolova, V. et al. Characterisation of exosomes derived from human cells by nanoparticle tracking analysis and scanning electron microscopy. *Colloids and surfaces. B, Biointerfaces* **87**, 146-150 (2011).

30. Palmisano, G. et al. Characterization of membrane-shed microvesicles from cytokine-stimulated beta-cells using proteomics strategies. *Mol Cell Proteomics* **11**, 230-243 (2012).

31. Royo, F. et al. Different EV enrichment methods suitable for clinical settings yield different subpopulations of urinary extracellular vesicles from human samples. *Journal of extracellular vesicles* **5**, 29497 (2016).

32. Cvjetkovic, A., Lotvall, J. & Lasser, C. The influence of rotor type and centrifugation time on the yield and purity of extracellular vesicles. *Journal of extracellular vesicles* **3** (2014).

33. Enderle, D. et al. Characterization of RNA from Exosomes and Other Extracellular Vesicles Isolated by a Novel Spin Column-Based Method. *PLoS One* **10**, e0136133 (2015).

34. Niu, Z. et al. Polymer-based precipitation preserves biological activities of extracellular vesicles from an endometrial cell line. *PLoS One* **12**, e0186534 (2017).

35. Yu, L.L. et al. A Comparison of Traditional and Novel Methods for the Separation of Exosomes from Human Samples. *Biomed Res Int* **2018**, 3634563 (2018).

36. Tauro, B.J. et al. Two distinct populations of exosomes are released from LIM1863 colon carcinoma cell-derived organoids. *Mol Cell Proteomics* **12**, 587-598 (2013).

37. Wu, X., Li, L., Iliuk, A. & Tao, W.A. Highly Efficient Phosphoproteome Capture and Analysis from Urinary Extracellular Vesicles. *J Proteome Res* **17**, 3308-3316 (2018).

38. Ramirez, M.I. et al. Technical challenges of working with extracellular vesicles. *Nanoscale* **10**, 881-906 (2018).

39. Contreras-Naranjo, J.C., Wu, H.J. & Ugaz, V.M. Microfluidics for exosome isolation and analysis: enabling liquid biopsy for personalized medicine. *Lab on a chip* **17**, 3558-3577 (2017).

40. Rosa-Fernandes, L., Rocha, V.B., Carregari, V.C., Urbani, A. & Palmisano, G. A Perspective on Extracellular Vesicles Proteomics. *Front Chem* **5**, 102 (2017).

41. Wisniewski, J.R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009).

42. Heath, N. et al. Rapid isolation and enrichment of extracellular vesicle preparations using anion exchange chromatography. *Sci Rep* **8**, 5730 (2018).

43. Feist, P. & Hummon, A.B. Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int J Mol Sci* **16**, 3537-3563 (2015).

44. Gundry, R.L. et al. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Curr Protoc Mol Biol* **Chapter 10**, Unit10 25 (2009).

45. Bereman, M.S., Egertson, J.D. & MacCoss, M.J. Comparison between procedures using SDS for shotgun proteomic analyses of complex samples. *Proteomics* **11**, 2931-2935 (2011).

46. Hsu, C.C. et al. Universal Plant Phosphoproteomics Workflow and Its Application to Tomato Signaling in Response to Cold Stress. *Mol Cell Proteomics* **17**, 2068-2080 (2018).

47. Bayramoglu, G., Celikbicak, O., Arica, M.Y. & Salih, B. Trypsin Immobilized on Magnetic Beads via Click Chemistry: Fast Proteolysis of Proteins in a Microbioreactor for MALDI-ToF-MS Peptide Analysis. *Industrial & Engineering Chemistry Research* **53**, 4554-4564 (2014).

48. Sielaff, M. et al. Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *J Proteome Res* **16**, 4060-4072 (2017).

49. Ludwig, K.R., Schroll, M.M. & Hummon, A.B. Comparison of In-Solution, FASP, and S-Trap Based Digestion Methods for Bottom-Up Proteomic Studies. *J Proteome Res* **17**, 2480-2490 (2018).

50. Swaney, D.L. & Villen, J. Enrichment of Phosphopeptides via Immobilized Metal Affinity Chromatography. *Cold Spring Harb Protoc* **2016**, pdb prot088005 (2016).

51. Fila, J. & Honys, D. Enrichment techniques employed in phosphoproteomics. *Amino Acids* **43**, 1025-1047 (2012).

52. Yue, X., Schunter, A. & Hummon, A.B. Comparing multistep immobilized metal affinity chromatography and multistep TiO2 methods for phosphopeptide enrichment. *Anal Chem* **87**, 8837-8844 (2015).

53. Montoya, A., Beltran, L., Casado, P., Rodriguez-Prados, J.C. & Cutillas, P.R. Characterization of a TiO(2) enrichment method for label-free quantitative phosphoproteomics. *Methods* **54**, 370-378 (2011).

54. Huang, J. et al. Highly Efficient Release of Glycopeptides from Hydrazide Beads by Hydroxylamine Assisted PNGase F Deglycosylation for N-Glycoproteome Analysis. *Anal Chem* **87**, 10199-10204 (2015).

55. Zhang, H., Li, X.J., Martin, D.B. & Aebersold, R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **21**, 660-666 (2003).

56. Cao, W., Huang, J., Jiang, B., Gao, X. & Yang, P. Highly Selective Enrichment of Glycopeptides Based on Zwitterionically Functionalized Soluble Nanopolymers. *Sci Rep* **6**, 29776 (2016).

57. Zhu, R., Zacharias, L., Wooding, K.M., Peng, W. & Mechref, Y. Glycoprotein Enrichment Analytical Techniques: Advantages and Disadvantages. *Methods Enzymol* **585**, 397-429 (2017).

58. Zhang, C. et al. Evaluation of Different N-Glycopeptide Enrichment Methods for N-Glycosylation Sites Mapping in Mouse Brain. *J Proteome Res* **15**, 2960-2968 (2016).

59. Nilsson, J. et al. Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat Methods* **6**, 809-811 (2009).

60. Yang, W. et al. Comparison of Enrichment Methods for Intact N- and O-Linked Glycopeptides Using Strong Anion Exchange and Hydrophilic Interaction Liquid Chromatography. *Anal Chem* **89**, 11193-11197 (2017).

61. Liebler, D.C. & Zimmerman, L.J. Targeted quantitation of proteins by mass spectrometry. *Biochemistry* **52**, 3797-3806 (2013).

62. Osinalde, N., Aloria, K., Omaetxebarria, M.J. & Kratchmarova, I. Targeted mass spectrometry: An emerging powerful approach to unblock the bottleneck in phosphoproteomics. *J Chromatogr B* **1055-1056**, 29-38 (2017).

63. Thery, C. 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. *Journal of extracellular vesicles* **7**, 1535750 (2018).

64. Shao, H. et al. New Technologies for Analysis of Extracellular Vesicles. *Chem Rev* **118**, 1917-1950 (2018).

65. Szatanek, R. et al. The Methods of Choice for Extracellular Vesicles (EVs) Characterization. *Int J Mol Sci* **18** (2017).

66. Lotvall, J. et al. Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. *Journal of extracellular vesicles* **3**, 26913 (2014).

67. Masuda, T., Saito, N., Tomita, M. & Ishihama, Y. Unbiased quantitation of Escherichia coli membrane proteome using phase transfer surfactants. *Mol Cell Proteomics* **8**, 2770-2777 (2009).

68. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896-1906 (2007).

69. Humphrey, S.J., Karayel, O., James, D.E. & Mann, M. High-throughput and high-sensitivity phosphoproteomics with the EasyPhos platform. *Nat Protoc* **13**, 1897-1916 (2018).

70. Kowal, J. et al. Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. *Proc Natl Acad Sci U S A* **113**, E968-977 (2016).

71. Coumans, F.A. et al. Reproducible extracellular vesicle size and concentration determination with tunable resistive pulse sensing. *Journal of extracellular vesicles* **3**, 25922 (2014).

72. Yuana, Y. et al. Cryo-electron microscopy of extracellular vesicles in fresh plasma. *Journal of extracellular vesicles* **2** (2013).

73. Iliuk, A.B., Martin, V.A., Alicie, B.M., Geahlen, R.L. & Tao, W.A. In-depth analyses of kinase-dependent tyrosine phosphoproteomes based on metal ion-functionalized soluble nanopolymers. *Molecular & cellular proteomics : MCP* **9**, 2162-2172 (2010).

74. Ye, J. et al. Optimized IMAC-IMAC protocol for phosphopeptide recovery from complex biological samples. *J Proteome Res* **9**, 3561-3573.

75. Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P. & Jorgensen, T.J. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics* **4**, 873-886 (2005).

76. Rauniyar, N. Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry. *Int J Mol Sci* **16**, 28566-28581 (2015).

77. Cordwell, S.J. & White, M.Y. Targeted proteomics for determining phosphorylation site-specific associations in cardiovascular disease. *Circulation* **126**, 1803-1807 (2012).

78. Zauber, H., Kirchner, M. & Selbach, M. Picky: a simple online PRM and SRM method designer for targeted proteomics. *Nat Methods* **15**, 156-157 (2018).

79. Milane, L., Singh, A., Mattheolabakis, G., Suresh, M. & Amiji, M.M. Exosome mediated communication within the tumor microenvironment. *J Control Release* (2015).

80. Dobrowolski, R. & De Robertis, E.M. Endocytic control of growth factor signalling: multivesicular bodies as signalling organelles. *Nat Rev Mol Cell Biol* **13**, 53-60 (2012).

81. Walsh, G. & Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nat Biotechnol* **24**, 1241-1252 (2006).

82. Andaluz Aguilar, H., Iliuk, A.B., Chen, I.H. & Tao, W.A. Sequential phosphoproteomics and N-glycoproteomics of plasma-derived extracellular vesicles. *Nat Protoc* **15**, 161-180 (2020).

83. Iliuk, A. et al. Plasma-Derived Extracellular Vesicle Phosphoproteomics through Chemical Affinity Purification. *J Proteome Res* **19**, 2563-2574 (2020).

84. Fragomeni, S.M., Sciallis, A. & Jeruss, J.S. Molecular Subtypes and Local-Regional Control of Breast Cancer. *Surg Oncol Clin N Am* **27**, 95-120 (2018).

85. Ballinger, T.J., Meier, J.B. & Jansen, V.M. Current Landscape of Targeted Therapies for Hormone-Receptor Positive, HER2 Negative Metastatic Breast Cancer. *Front Oncol* **8**, 308 (2018).

86. Gajria, D. & Chandarlapaty, S. HER2-amplified breast cancer: mechanisms of trastuzumab resistance and novel targeted therapies. *Expert Rev Anticancer Ther* **11**, 263-275 (2011).

87. Bodai, B.I. & Tuso, P. Breast cancer survivorship: a comprehensive review of long-term medical issues and lifestyle recommendations. *Perm J* **19**, 48-79 (2015).

88. Feng, Y. et al. Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes Dis* **5**, 77-106 (2018).

89. Rontogianni, S. et al. Proteomic profiling of extracellular vesicles allows for human breast cancer subtyping. *Commun Biol* **2**, 325 (2019).

90. Grimes, M. et al. Integration of protein phosphorylation, acetylation, and methylation data sets to outline lung cancer signaling networks. *Sci Signal* **11**, 1-16 (2018).

91. Tyanova, S. et al. Proteomic maps of breast cancer subtypes. *Nature communications* **7**, 10259 (2016).

92. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).

93. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301-2319 (2016).

94. Tyanova, S. & Cox, J. Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. *Methods Mol Biol* **1711**, 133-148 (2018).

95. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968 (2010).

96. Wright, M.N. & Ziegler, A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software* **77** (2017).

97. Nembrini, S., König, I.R., Wright, M.N. & Valencia, A. The revival of the Gini importance? *Bioinformatics* **34**, 3711-3718 (2018).

98. Sabatier, R. et al. Prognostic and predictive value of PDL1 expression in breast cancer. *Oncotarget* **6**, 5449-5464 (2014).

99. Muenst, S. et al. Expression of programmed death ligand 1 (PD-L1) is associated with poor prognosis in human breast cancer. *Breast Cancer Research and Treatment* **146**, 15-24 (2014).

100. Mittendorf, E.A. et al. PD-L1 Expression in Triple-Negative Breast Cancer. *Cancer Immunology Research* **2**, 361-370 (2014).

101. Mazel, M. et al. Frequent expression of PD-L1 on circulating breast cancer cells. *Molecular Oncology* **9**, 1773-1782 (2015).

102. Li, C.-W. et al. Eradication of Triple-Negative Breast Cancer Cells by Targeting Glycosylated PD-L1. *Cancer cell* **33**, 187-201.e110 (2018).

103. Ferreira, I.G. et al. Glycosylation as a Main Regulator of Growth and Death Factor Receptors Signaling. *International Journal of Molecular Sciences* **19** (2018).

104. Yang, L. et al. Targeting cancer stem cell pathways for cancer therapy. *Signal Transduction and Targeted Therapy* **5** (2020).

105. Wu, N. et al. Precision medicine based on tumorigenic signaling pathways for triple-negative breast cancer (Review). *Oncology Letters* (2018).

106. Hardy, K.M., Booth, B.W., Hendrix, M.J.C., Salomon, D.S. & Strizzi, L. ErbB/EGF Signaling and EMT in Mammary Development and Breast Cancer. *Journal of Mammary Gland Biology and Neoplasia* **15**, 191-199 (2010).

107. Giuli, M.V., Giuliani, E., Screpanti, I., Bellavia, D. & Checquolo, S. Notch Signaling Activation as a Hallmark for Triple-Negative Breast Cancer Subtype. *Journal of Oncology* **2019**, 1-15 (2019).

108. Sosa, M.S. et al. Identification of the Rac-GEF P-Rex1 as an Essential Mediator of ErbB Signaling in Breast Cancer. *Molecular cell* **40**, 877-892 (2010).

109. Butti, R. et al. Receptor tyrosine kinases (RTKs) in breast cancer: signaling, therapeutic implications and challenges. *Molecular Cancer* **17** (2018).

110. Akinleye, A., Chen, Y., Mukhi, N., Song, Y. & Liu, D. Ibrutinib and novel BTK inhibitors in clinical development. *Journal of Hematology & Oncology* **6**, 1-9 (2013).

111. Eifert, C. et al. A novel isoform of the B cell tyrosine kinase BTK protects breast cancer cells from apoptosis. *Genes Chromosomes Cancer* **52**, 961-975 (2013).

112. Wang, X. et al. Bruton's Tyrosine Kinase Inhibitors Prevent Therapeutic Escape in Breast Cancer Cells. *Molecular Cancer Therapeutics* **15**, 2198-2208 (2016).

113. Liu, L. et al. MYH9 overexpression correlates with clinicopathological parameters and poor prognosis of epithelial ovarian cancer. *Oncology Letters* (2019).

114. Kas, S.M. et al. Insertional mutagenesis identifies drivers of a novel oncogenic pathway in invasive lobular breast carcinoma. *Nature genetics* **49**, 1219-1230 (2017).

115. Wang, B. et al. MYH9 Promotes Growth and Metastasis via Activation of MAPK/AKT Signaling in Colorectal Cancer. *Journal of Cancer* **10**, 874-884 (2019).

116. Zhou, W. et al. The expression of MYH9 in osteosarcoma and its effect on the migration and invasion abilities of tumor cell. *Asian Pacific Journal of Tropical Medicine* **9**, 597-600 (2016).

117. Curmi, P. et al. Overexpression of stathmin in breast carcinomas points out to highly proliferative tumours. *British Journal of Cancer* **82**, 142-1150 (1999).

118. Zhou, H. et al. Elevated transgelin/TNS1 expression is a potential biomarker in human colorectal cancer. *Oncotarget* **9**, 1107-1113 (2017).

119. Lu, Y.-M., Cheng, F. & Teng, L.-S. The association between phosphatase and tensin homolog hypermethylation and patients with breast cancer, a meta-analysis and literature review. *Sci Rep* **6** (2016).

120. Askeland, C. et al. Stathmin expression associates with vascular and immune responses in aggressive breast cancer subgroups. *Sci Rep* **10** (2020).

121. Obayashi, S. et al. Stathmin1 expression is associated with aggressive phenotypes and cancer stem cell marker expression in breast cancer patients. *International Journal of Oncology* **51**, 781-790 (2017).

122. Legare, S. et al. The Estrogen Receptor Cofactor SPEN Functions as a Tumor Suppressor and Candidate Biomarker of Drug Responsiveness in Hormone-Dependent Breast Cancers. *Cancer research* **75**, 4351-4363 (2015).

123. Merchant, M.L., Rood, I.M., Deegens, J.K.J. & Klein, J.B. Isolation and characterization of urinary extracellular vesicles: implications for biomarker discovery. *Nat Rev Nephrol* **13**, 731-749 (2017).

124. Ramis, J.M. Extracellular Vesicles in Cell Biology and Medicine. *Sci Rep* **10**, 8667 (2020).

125. van Niel, G., D'Angelo, G. & Raposo, G. Shedding light on the cell biology of extracellular vesicles. *Nat Rev Mol Cell Biol* **19**, 213-228 (2018).

126. Turpin, D. et al. Role of extracellular vesicles in autoimmune diseases. *Autoimmun Rev* **15**, 174-183 (2016).

127. Liu, H. et al. Immunomodulatory Effects of Mesenchymal Stem Cells and Mesenchymal Stem Cell-Derived Extracellular Vesicles in Rheumatoid Arthritis. *Front Immunol* **11**, 1912 (2020).

128. Katsiougiannis, S. Extracellular Vesicles: Evolving Contributors in Autoimmunity. *For Immunopathol Dis Therap* **6**, 163-170 (2015).

129. Tian, J., Casella, G., Zhang, Y., Rostami, A. & Li, X. Potential roles of extracellular vesicles in the pathophysiology, diagnosis, and treatment of autoimmune diseases. *Int J Biol Sci* **16**, 620-632 (2020).

130. Ramos, P., Shedlock, A. & Langefeld, C. Genetics of autoimmune diseases: insights from population genetics. *Journal of Human Genetics* **60**, 657-664 (2015).

131. Miller, F.W. et al. Criteria for environmentally associated autoimmune diseases. *Journal of Autoimmunity* **39**, 253-258 (2012).

132. Liu, Y., Yin, H., Zhao, M. & Lu, Q. TLR2 and TLR4 in autoimmune diseases: a comprehensive review. *Clin Rev Allergy Immunol* **47**, 136-147 (2014).

133. Boilard, E. et al. Platelets Amplify Inflammation in Arthritis via Collagen Dependent Microparticle Production. *Science* **327**, 580-583 (2010).

134. Sellam, J. et al. Increased levels of circulating microparticles in primary Sjogren's syndrome, systemic lupus erythematosus and rheumatoid arthritis and relation with disease activity. *Arthritis Res Ther* **11**, R156 (2009).

135. Knijff-Dutmer, E.A., Koerts, J., Nieuwland, R., Kalsbeek-Batenburg, E.M. & van de Laar, M.A. Elevated levels of platelet microparticles are associated with disease activity in rheumatoid arthritis. *Arthritis Rheum* **46**, 1498-1503 (2002).

136. Pereira, J. et al. Circulating platelet-derived microparticles in systemic lupus. *Arthritis Reumatology* **2**, 94-99 (2005).

137. Ullal, A.J. et al. Microparticles as antigenic targets of antibodies to DNA and nucleosomes in systemic lupus erythematosus. *J Autoimmun* **36**, 173-180 (2011).

138. Hueber, W. & Robinson, W.H. Proteomic biomarkers for autoimmune disease. *Proteomics* **6**, 4100-4105 (2006).

139. Hershko, A.Y. & Naparstek, Y. Autoimmunity in the era of genomics and proteomics. *Autoimmun Rev* **5**, 230-233 (2006).

140. Navarrete-Perea, J., Yu, Q., Gygi, S.P. & Paulo, J.A. Streamlined Tandem Mass Tag (SL-TMT) Protocol: An Efficient Strategy for Quantitative (Phospho)proteome Profiling Using Tandem Mass Tag-Synchronous Precursor Selection-MS3. *J Proteome Res* **17**, 2226-2236 (2018).

141. Dayon, L. & Sanchez, J.C. Relative protein quantification by MS/MS using the tandem mass tag technology. *Methods Mol Biol* **893**, 115-127 (2012).

142. Clark, D.J. et al. Redefining the Breast Cancer Exosome Proteome by Tandem Mass Tag Quantitative Proteomics and Multivariate Cluster Analysis. *Anal Chem* **87**, 10462-10469 (2015).

143. Zheng, X. et al. A circulating extracellular vesicles-based novel screening tool for colorectal cancer revealed by shotgun and data-independent acquisition mass spectrometry. *Journal of extracellular vesicles* **9**, 1750202 (2020).

144. Ardito, F., Giuliani, M., Perrone, D., Troiano, G. & Lo Muzio, L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int J Mol Med* **40**, 271-280 (2017).

145. Sacco, F. et al. Glucose-regulated and drug-perturbed phosphoproteome reveals molecular mechanisms controlling insulin secretion. *Nature communications* **7**, 13250 (2016).

146. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731-740 (2016).

147. Mi, H. et al. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat Protoc* **8**, 1551–1566 (2013).

148. Kaimal, V., Bardes, E.E., Tabar, S.C., Jegga, A.G. & Aronow, B.J. ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. *Nucleic Acids Res* **38**, W96-102 (2010).

149. Shiromizu, T. et al. Quantitation of putative colorectal cancer biomarker candidates in serum extracellular vesicles by targeted proteomics. *Sci Rep* **7**, 12782 (2017).

150. Lin, Y. et al. Proteomic analysis of seminal extracellular vesicle proteins involved in asthenozoospermia by iTRAQ. *Mol Reprod Dev* **86**, 1094-1105 (2019).

151. Li, J., He, X., Deng, Y. & Yang, C. An Update on Isolation Methods for Proteomic Studies of Extracellular Vesicles in Biofluids. *Molecules* **24** (2019).

152. Brennan, K. et al. A comparison of methods for the isolation and separation of extracellular vesicles from protein and lipid particles in human serum. *Sci Rep* **10**, 1039 (2020).

153. Kirbas, O.K. et al. Optimized Isolation of Extracellular Vesicles From Various Organic Sources Using Aqueous Two-Phase System. *Sci Rep* **9**, 19159 (2019).

154. Takov, K., Yellon, D.M. & Davidson, S.M. Comparison of small extracellular vesicles isolated from plasma by ultracentrifugation or size-exclusion chromatography: yield, purity and functional potential. *Journal of extracellular vesicles* **8**, 1560809 (2019).

155. Simpson, R.J., Kalra, H. & Mathivanan, S. ExoCarta as a resource for exosomal research. *Journal of extracellular vesicles* **1** (2012).

156. Keerthikumar, S. et al. ExoCarta: A Web-Based Compendium of Exosomal Cargo. *J Mol Biol* **428**, 688-692 (2016).

157. Piranavan, P., Bhamra, M. & Perl, A. Metabolic Targets for Treatment of Autoimmune Diseases. *Immunometabolism* **2** (2020).

158. Tuller, T., Atar, S., Ruppin, E., Gurevich, M. & Achiron, A. Common and specific signatures of gene expression and protein-protein interactions in autoimmune diseases. *Genes Immun* **14**, 67-82 (2013).

159. Luan, M. et al. The shared and specific mechanism of four autoimmune diseases. *Oncotarget* **8**, 108355-108374 (2017).

160. Ruiz-Romero, C. et al. Mining the Proteome Associated with Rheumatic and Autoimmune Diseases. *J Proteome Res* **18**, 4231-4239 (2019).

161. Johnson, L.N. The effects of phosphorylation on the structure and function of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 199-232 (1993).

162. Menon, D. et al. GSTO1-1 plays a pro-inflammatory role in models of inflammation, colitis and obesity. *Sci Rep* **7**, 17832 (2017).

163. Peters, A.L., Stunz, L.L. & Bishop, G.A. CD40 and autoimmunity: the dark side of a great activator. *Semin Immunol* **21**, 293-300 (2009).

164. Sun, J. et al. Anti-GAPDH Autoantibody Is Associated with Increased Disease Activity and Intracranial Pressure in Systemic Lupus Erythematosus. *J Immunol Res* **2019**, 7430780 (2019).

165. Kunishita, Y. et al. TRIM21 Dysfunction Enhances Aberrant B-Cell Differentiation in Autoimmune Pathogenesis. *Front Immunol* **11**, 98 (2020).

166. Patente, T.A. et al. Allelic variations in the CYBA gene of NADPH oxidase and risk of kidney complications in patients with type 1 diabetes. *Free Radic Biol Med* **86**, 16-24 (2015).

167. Keskitalo, S. et al. Dominant TOM1 mutation associated with combined immunodeficiency and autoimmune disease. *NPJ Genom Med* **4**, 14 (2019).

PUBLICATIONS

1. **Hillary Andaluz Aguilar**, Anton B. Iliuk, I-Hsuan Chen, W. Andy Tao (2020). Sequential phosphoproteomics and N-glycoproteomics of plasma-derived extracellular vesicles. ***Nature Protocols*** 15, 61-180. doi.org/10.1038/s41596-019-0260-5.

2. I-Hsuan Chen, **Hillary Andaluz Aguilar**, J. Sebastian Paez, Xiaofeng Wu, Michael K. Wendt, Li Pan, Anton B. Iliuk, Ying Zhang, W. Andy Tao (2018). Analytical pipeline for discovery and verification of glycoproteins from plasma-derived extracellular vesicles as breast cancer biomarkers. ***Analytical Chemistry*** 90, 6307-6313. doi: 10.1021/acs.analchem.8b01090.

3. Li Pan, **Hillary Andaluz Aguilar**, Linna Wang, Anton Iliuk, and W. Andy Tao (2016). Three-dimensionally functionalized reverse phase glycoprotein array for cancer biomarker discovery and validation, ***Journal of the American Chemical Society (JACS)*** 138, 15311-15314. doi.org/10.1021/jacs.6b10239.

4. I-Hsuan Chen, Liang Xue, Chuan-Chih Hsu, Juan Sebastian Paez Paez, Li Pan, **Hillary Andaluz**, Michael K. Wendt, Anton B. Iliuk, Jian-Kang Zhu, and W. Andy Tao (2017). Phosphoproteins in extracellular vesicles as candidate markers for breast cancer. ***Proceedings of the National Academy of Sciences (PNAS)*** 114, 3175-3180. doi.org/10.1073/pnas.1618088114.