Improving Blood Plasma Glycoproteome Coverage by Coupling Ultracentrifugation Fractionation to Electrostatic Repulsion–Hydrophilic Interaction Chromatography Enrichment

Adav, Sunil S.; Hwa, Ho Hee; de Kleijn, Dominique; Sze, Siu Kwan

2015

Adav, S. S., Hwa, H. H., de Kleijn, D., & Sze, S. K. (2015). Improving Blood Plasma Glycoproteome Coverage by Coupling Ultracentrifugation Fractionation to Electrostatic Repulsion–Hydrophilic Interaction Chromatography Enrichment. *Journal of Proteome Research*, 14(7), 2828-2838.; Adav, S. S., Hwa, H. H., de Kleijn, D., & Sze, S. K. (2015). Improving Blood Plasma Glycoproteome Coverage by Coupling Ultracentrifugation Fractionation to Electrostatic Repulsion–Hydrophilic Interaction Chromatography Enrichment. *Journal of Proteome Research*, 14(7), 2828-2838.

<https://hdl.handle.net/10356/81887>

<https://doi.org/10.1021/acs.jproteome.5b00102>

© 2015 American Chemical Society. This is the author created version of a work that has been peer reviewed and accepted for publication by *Journal of Proteome Research*, American Chemical Society. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<http://dx.doi.org/10.1021/acs.jproteome.5b00102>].

**Improving Blood Plasma Glycoproteome Coverage by Coupling Ultracentrifugation
Fractionation to Electrostatic Repulsion-Hydrophilic Interaction Chromatography Enrichment**

Sunil S. Adav^{1,2}, Ho Hee Hwa³, Dominique de Kleijn^{4,5}, and Siu Kwan Sze^{1*}

¹School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551.

²KK Women's and Children's Hospital, Singapore 229899

³Department of Cardiology, Tan Tock Seng Hospital, 11 Jalan Tan Tock Seng, Singapore 308433.

⁴Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, & Cardiovascular Research Institute, Singapore 119228.

⁵Experimental Cardiology Laboratory, Cardiology, University Medical Center Utrecht, the Netherlands & Interuniversity Cardiovascular Institute of the Netherlands, Utrecht, the Netherlands.

Keywords: ERLIC, glycoproteins, plasma proteome, protein glycosylation, biomarkers.

Running Title: Enhanced plasma glycoproteome coverage by ultracentrifugation and ERLIC enrichment

***Correspondence:**

Siu Kwan SZE, PhD
School of Biological Sciences
Division of Structural Biology and Biochemistry
Nanyang Technological University,
60 Nanyang Drive, Singapore 637551
Tel: (+65) 6514-1006
Fax: (+65) 6791-3856
Email: sksze@ntu.edu.sg

Abstract

Blood plasma is considered as an excellent source of disease biomarkers because it contains proteins, lipids, metabolites, cell and cell-derived extracellular vesicles from different cellular origins including diseased tissues. Most secretory and membranous proteins that can be found in plasma are glycoproteins, therefore, the plasma glycoproteome is one of the major sub-proteome which is highly enriched with disease biomarkers. As a result, the glycoproteome has attracted much attention in clinical proteomic research. The modification of proteins with glycans regulates a wide range of functions in biology but profiling plasma glycoproteins on a global scale has been hampered by the presence of low stoichiometry of glycoproteins in a complex high abundance plasma proteome background and lack of effective analytical technique. This study aims to improve plasma glycoproteome coverage using pig plasma as a model sample with a two-step strategy. First step involves fractionation of the plasma proteins using ultracentrifugation into supernatant and pellet that is believed to contain low abundant glycoproteins. In second step, further enrichment of glycopeptides was achieved in both fractions by adopting electrostatic repulsion hydrophilic interaction chromatography (ERLIC) coupled with tandem mass spectrometry (LC-MS/MS) analysis. The coverage of enriched glycoproteins in supernatant, pellet and whole plasma sample as control were compared. Using this simple sample fractionation approach by ultracentrifugation and further ERLIC enrichment technique, sample complexity was reduced and glycoproteome coverage was significantly enhanced in supernatant and pellet fractions (by > 50%) compared with whole plasma sample. This study showed that when ultracentrifugation is coupled with ERLIC, glycopeptides enrichment and glycoproteome identification is significantly improved. This study demonstrates the combination of ultracentrifugation and ERLIC as a useful method for discovering plasma glycoprotein disease biomarkers.

1 **Introduction**

2 Modifications of amino acids on the side chain of a given protein can alter the
3 biological function of the protein and can involved in initiation and progression of the
4 molecular pathophysiology of a disease. Post-translational modifications (PTMs) of a protein
5 determine its localization and interactions with other proteins,¹ and also considered to be
6 significantly important in the study of cell biology, signaling, metabolic pathways,
7 biomarkers, disease treatments and many other processes. The study of protein glycosylation
8 is a vital research area in proteomics because as many as 50% of proteins in human body are
9 glycosylated.^{2, 3} It plays important roles in several biological processes including embryonic
10 development, cell-to-cell interactions, cell division, protein regulation and interaction, protein
11 stability, secretion, localization and many pathophysiological processes.⁴⁻⁸ Previous research
12 work on blood-group specificities revealed that the glycans attached to proteins and lipids
13 were the molecular determinants.⁹

14 Plasma glycoproteome has important clinical value, as most secreted biomarkers are
15 glycosylated.^{10, 11} For example, CA125 is an ovarian cancer antigen and its primary structure
16 suggests that CA125 is a giant mucin-like glycoprotein presents on the cell surface of tumor
17 cells.¹¹ Glycosylated HER2 is a member of the human epidermal growth factor receptor and
18 it over expressed in certain aggressive types of breast, ovarian and aggressive forms of
19 uterine cancer are well known.¹²⁻¹⁴ Prostate-specific antigen in prostate cancer and α -
20 fetoprotein in hepatocellular carcinoma are also glycosylated proteins.^{12, 13} Despite the great
21 importance of glycosylation in biological functions of proteins, glycoproteomic study has
22 been hampered by lacking of effective analytical methods for identifying low abundant
23 glycoproteins in complex biological samples.

24 Plasma is a major body fluid that contains proteins released from multiple organs and
25 its composition reflects the overall health status of the individual. Hence plasma is often used

1 to monitor health and disease status. Studies on plasma protein identification and their
2 characterization may provide valuable information to investigate the disease status and
3 underlying mechanism, to identify biomarkers associated with new drug targets, and be also
4 useful in early diagnostics of disease. Plasma is very complex due to the wide range in
5 plasma protein concentration and number of protein species. Moreover, plasma proteins can
6 be in soluble and aggregated forms, as well as attached to cell debris and enclosed in
7 extracellular vesicles. Most secretory and membranous proteins that can be found in plasma
8 are glycoproteins, therefore, the plasma glycoproteome is one of the major sub-proteome
9 which is highly enriched with disease biomarkers. A suitable method that can enrich
10 glycosylated peptides for identifying low abundant glycoproteins is therefore highly
11 demanded.

12 Recent glycoproteomics research focuses on enrichment of glycopeptides, site-
13 specific features of glycopeptides, glycosylation sites and glycan structures, because many of
14 these factors affect protein folding and functions.^{15, 16} To date high resolution liquid
15 chromatography coupled tandem mass spectrometry (LC-MS/MS) has been the preferred
16 technique for proteome, glycoproteome and phosphoproteome identification. To enhance the
17 dynamic proteome coverage, SCX chromatography is generally used for the fractionation of
18 tryptic peptides prior to LC-MS/MS analysis. However, the disadvantages of SCX
19 chromatography includes difficulty in separation of similarly charged clusters of peptides, use
20 high concentration of nonvolatile salts that is incompatible with LC-MS/MS and required an
21 extra peptide desalting prior to analysis. On the contrary, ERLIC, introduced by Alpert for
22 biomolecules and phosphopeptide enrichment¹⁷ and further developed by Sze et al. into a salt
23 free peptide fractionation,¹⁸ is a method simultaneously enriching for both phosphor- and
24 glyco-peptides.¹⁹ Next to this, it is used for proteome-wide deamidated peptides separation
25 and analysis.^{20, 21}

Although advances in mass spectrometry for subfemto-molar detection of biomolecules have enabled the global analysis of protein glycosylation to be addressed, this remains challenging because of the inherent low abundance of disease biomarkers, heterogeneous and diverse nature of glycans. Also due to different glycoforms of glycosylated peptides, analysis of glycopeptides by conventional MS-based techniques remains difficult. To solve this problem, several methods have been developed to target specific glycoproteomes, to enrich glycoproteins and glycopeptides from complex samples such as plasma, urine or cells.²² Lectin affinity-based enrichment, boronate, hydrophilic interaction liquid chromatography (HILIC), SCX and hydrazide covalent chromatography techniques are well-known for the enrichment of glycoproteins or glycopeptides.^{3, 23-27} The microheterogeneity at the glycosylation site, low concentration of glycopeptide, ion suppression effects and competition for ionization with co-eluting peptides during LC-MS/MS run, however, further limits the analysis of glycopeptides.²⁸ Considering the great importance and persistent interest of protein glycosylation in drug targeting and disease biomarkers, this study aimed to develop a method that can simultaneously analyze glycoproteome of plasma proteins existing in different forms such as soluble, low abundant, associated with protein aggregates or cell debris and extracellular vesicles. To achieve global plasma glycoproteome profile, ultracentrifugation fractionation was used as a first step to separate low abundant soluble glycoproteins into a pellet from the relatively high abundant soluble plasma proteins. The supernatant and pellet fractions after tryptic digestion were further subjected as a second step to salt free ERLIC based glycopeptide enrichment. In this study, we demonstrated in pig plasma that this method facilitates successful enrichment of both low and high abundant glycoproteins.

Materials and methods

Sample Preparation

Blood samples were collected from the ear pigs, plasma extracted and stored at -80 °C. Each of the samples described below was performed in two replicates and mean data with standard deviation were reported. We used 2 mL plasma and diluted to 10 mL. The plasma samples were divided into two groups. Group I was processed as it is (whole plasma) while group II was ultracentrifuged at 100000 g for 1 h at 4 °C. The resultant supernatant and pellet fractions were analyzed separately. The proteins from whole plasma sample and supernatant fractions were precipitated by adding ice cold acetone. The samples were kept at 4 °C overnight and precipitated proteins were collected by centrifugation at 20000 g for 10 min. Protein content was determined by the Bradford method and about 1 mg of sample protein from each group was dissolved in 8M urea in 100mM ammonium acetate buffer at pH6, reduced with 20 mM DTT at 60 °C for 1 h and alkylated with 80 mM iodoacetamide for 45 min in the dark. The proteins samples were diluted to 1M urea at pH6 and subjected to sequencing grade modified trypsin (Promega, Madison, WI) digestion at 37 °C for overnight. The obtained tryptic peptides were desalted using a Sep-Pak C18 cartridge (Waters, Milford, MA) and dried in a SpeedVac (Thermo Electron, Waltham, MA).

ERLIC Fractionation and LC-MS/MS analysis

The tryptic peptides of digested pig plasma samples were fractionated using a PolyWAX LP weak anion-exchange column (4.6× 200 mm, 5 µm, 300 Å, PolyLC, Columbia, MD) using Shimadzu Prominence UFLC system. Mobile phase A (85% ACN/0.5% FA) and mobile phase B (30% ACN/0.2% FA) were used to establish 60 min gradient consisting of 5–16% B over 35 min and 16–100% B over 15 min followed by 10 min at 100% B, at a flow rate of 0.9 mL/min and elution fractions were collected. The fractions were dried using vacuum centrifuge, resuspended in ammonium acetate buffer (100mM, pH 6.0) and treated

with PNGase F at 37° C for 6 h for deglycosylation of glycopeptides. The resultant peptides were dried using vacuum centrifuge and analyzed by LC-MS/MS.

The samples were separated and analyzed on a Dionex UHPLC system coupled to a LTQFT Ultra (Thermo Electron, Bremen, Germany). The samples were loaded to auto-sampler, and then 6µl sample (~1.7 µg peptides) was injected onto a Zorbax peptide trap column (Agilent, CA) for concentration and desalting. The peptides were separated in a capillary column (200 µm × 10 cm) packed with C18 AQ (5 µm, 300 Å; Bruker-Michrom, Auburn, CA) at flow rate 500 nL/min. Mobile phase A (0.1% FA in H₂O) and mobile phase B (0.1% FA in ACN) were utilized to establish the 60 min gradient comprised of 45 min of 5–35% B, 8 min of 35–50% B, and 2 min of 80% B followed by re-equilibration at 5% B for 5 min. Then, peptides were analyzed on the LTQFT Ultra with an ADVANCE CaptiveSpray Source (Bruker-Michrom) at an electrospray potential of 1.5 kV. Ion transfer tube temperature of 180 °C, and collision gas pressure of 0.85 mTorr were other set parameters. The LTQFT Ultra was set to perform data dependent acquisition in the positive ion mode as previously described.²⁹ Briefly, a full MS scan (350–1600 m/z range) was acquired in the FT-ICR cell at a resolution of 100000 and a maximum ion accumulation time of 1000 ms. The default AGC setting was used (full MS target at 3.0e+04, MSn1e+04) in linear ion trap. The 10 most intense ions above a 500 count threshold were selected for fragmentation in CAD, which was concurrently performed with a maximum ion accumulation time of 200 ms. Dynamic exclusion was activated for the process, with a repeat count of 1 and exclusion duration of 60 s. Single-charged ion was excluded from MS/MS. For CAD, normalized collision energy was set to 35%, activation Q was set to 0.25, and activation time was 30 ms.

Data Analysis

The raw data files were converted into Mascot generic file format using Proteome Discoverer version 1.4 (Thermo Electron, Bremen, Germany), Uniprot pig protein database (downloaded on 7 Jan, 2015 containing 11,822,469 residues and 26,149 sequences as well as the reverse sequences for decoy search) was used for database searches. The database search was performed using an in-house Mascot server (version 2.4.1, Matrix Science, Boston, MA) with MS tolerance of 5.1 ppm and MS/MS tolerance of 0.8 Da³⁰. Two missed cleavage sites of trypsin were allowed. Carbamidomethylation (C) was set as a fixed modification, while oxidation (M), and deamidation (N and Q) were variable modifications. The obtained peptide/protein list for each fraction was exported to Microsoft Excel or processed using an in-house script for further analysis. The FDR of peptide identification was set to be less than 1% ($\text{FDR} = 2.0 \times \text{decoy_hits}/\text{total_hits}$).

Results and Discussion

Over the years it has become apparent that changes in plasma glycoproteins play an important role in the pathogenesis and progression of various diseases. Plasma glycoproteins are considered as excellent source of disease biomarkers in cancer, liver and many other diseases.^{31, 32} However, progress in this field of glycoproteomics has been hampered by various technical challenges. Our main goal in this study was to develop protocol for the reduction of the complexity of glycoproteins in plasma by ultracentrifugation fractionation, and subsequent enrichment of glycosylated peptides using ERLIC to maximize glycoprotein identification coverage using pig plasma as a model sample.

To develop a method for glycoprotein enrichment, we diluted pig plasma by 5-fold using PBS and separate it into supernatant and pellet fractions by ultracentrifugation. Whole plasma sample (without dilution and ultracentrifugation) was parallel analyzed and served as

control for comparison with the fractionated plasma for glycopeptide coverage. The experimental conditions like alkaline pH buffer that promote non-enzymatic deamidation^{19, 30, 33} and may introduce false positive deamidation sites were avoided in our protocol. Mild acidic buffer conditions that keep trypsin activity and favor more accurate identification of glycosylation sites^{34, 35} were used in tryptic digestion and **deglycosylation** procedures. We used acidic conditions throughout the whole experiment and evaluated accurately assigned glycosylation sites after enrichment by ERLIC. Carbohydrate moiety in N-linked glycosylation is attached to an asparagine residue (N) followed by a non-proline amino acid (X) and then a serine (S)/threonine (T)/Cysteine (C), resulting in a consensus N-X-S/T/C sequence motif. Plasma glycopeptides were detected in the N-X-S/T/C sequence motif. Our study used more specific PNGase F that catalyzes the cleavage of N-linked oligosaccharides between the innermost GlcNAc and asparagine residues of high mannose, hybrid and complex oligosaccharides from N-linked glycoproteins. Thus, glycopeptides were identified as deamidated peptides with N-X-S/T/C motif. PNGase F cannot remove oligosaccharides containing alpha-(1,3)-linked core fucose³⁶, therefore, fucosylated sites and fucosylated proteins might not be covered in this study. When analyzing glycoproteins that known to contain fucosylation, glycosidase with a broader specificity need to be used.

Glycoproteome enrichment using ERLIC

Adopting ERLIC based glycopeptides enrichment method followed with RPLC-MS/MS analysis we identified 14196±887, 18798±2795 and 21262±983 peptides that corresponds to 565, 608, 585 proteins in whole plasma, supernatant and pellet fractions. Although pellet showed higher number of peptides than whole plasma but protein number was not proportionally high. This indicates that pellet **contains** relatively high molecular

weight proteins. Protein glycosylation typically results in aberrantly high molecular weights. However, comparison of peptide and protein data suggests that the high molecular weight proteins of pellet fraction may have low tryptic sites or poor proteolytic cleavage. Poor proteolytic cleavage could stem from the structural features of the protein and its biological role. For instance, kininogen-1 has multiple posttranslation modifications (four phosphorylations, 12 glycosylations) and disulfide bonds³⁷, which, together with its native fold, could interfere with trypsin's ability to hydrolyse specific amide bonds³⁸. Again, phosphorylated residues within two amino acids of the point of cleavage can hinder proteolysis.^{39, 40} Using shotgun liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) Zhang et al.⁴¹ identified 392 proteins during proteome profiling of healthy pig plasma, while 183 proteins were identified from pig plasma by MALDI-TOF/TOF MS.⁴² Further analysis by overlap (Fig.1) of identified proteins revealed 211 proteins common in whole plasma, pellet and supernatant samples; 114 proteins common in whole and supernatant sample, 49 proteins common in whole and pellet, 49 proteins common in supernatant and pellet. Interestingly, 191, 234 and 276 proteins were unique in whole, supernatant and pellet fraction suggesting simplifying analytical sample by ultracentrifugation effectively separate proteins from different origins in complex plasma and enhances protein coverage (Fig. 1). As expected, higher overlap of protein population (374 proteins) from whole plasma and, pellet and supernatant were noted but yet 191 unique proteins were identified in whole plasma. These proteins (191 proteins) could not overlap either with supernatant or pellet proteins and reason [remains](#) still unknown. However, majority of these proteins were lipid associated proteins that formed the lipid bilayer at the top of ultracentrifuged and not pelleted down. Although, ultracentrifugation is method of choice for separation of lipoproteins but single cycle of ultracentrifugation might not be sufficient to pellet down lipoproteins.⁴³ The incorporation of more ultracentrifugation steps

would improve the results. Reducing the sample complexity by dividing sample into supernatant and pellet enhanced protein identification and detection. Other researchers⁴⁴ also suggested to decrease sample complexity prior to LC-MS analysis to get optimizing quantitative proteomics output. Plasma protein composition reveals the overall health status of individual since it contains proteins released from multiple organs and tissue compartments. Identified protein localization revealed 16.8% cellular components, 10.9% cell, 10.1% organelle, 9.5% intracellular, 7.6% extracellular, 7.4% cytoplasmic and many more as shown in Fig. 2A. Plasma has been considered as a potential source of disease biomarkers and hence proteomic profiling of plasma and protein functions may highlight clinical useful protein candidates. The functional classification of the proteome showed majority of proteins associated with enzyme activities and regulation (32.5%), ion binding (18.9%), DNA binding (4.7%) and RNA binding (3.7%) (Fig. 2B). The classification based on the biological processes was shown in Fig. 2C.

ERLIC is a mixed-mode chromatographic technique that separates peptides based on charge and polarity.⁴⁵ In ERLIC separations, a weak cationic column surface chemistry are used to repel a cationic polar group on an analyte such as amino group in peptides or within a set of analytes, but retains analytes by hydrophilic interaction such as carbohydrate groups, and by electrostatic interaction with anionic groups such as carboxylic or sialic acid groups to facilitate separation by these polar groups.¹⁷ Since this study focused on the identification of glycopeptides, the bulk of peptides eluted in the first 5 min was not collected. A total of 55 fractions from 5 to 60 min were collected (chromatogram shown in Fig. S1), 3 consecutive fractions were combined and thus 18 fractions were treated with PGNase-F and analyzed by LC-MS/MS. In ERLIC, unmodified charged peptides are repelled electrostatically by the weak anion-exchange material (WAX) but yet glycopeptides were retained in column due to

organic mobile phase through ionic and hydrophilic interactions and get distributed into multiple fractions.

Glycoproteomic analysis by LC-MS/MS

The glycoproteomic analysis based on hydrazide chemistry has been standardized and widely used²⁷ whereas ERLIC is a relatively new method. Adopting ERLIC method for glycopeptide enrichment as described above and analyzing the samples using shotgun LC-MS/MS based proteomics method, we identified 1570.5 ± 80.5 glycopeptides and 1602 ± 83 glycosylation sites in whole pig plasma sample. While total glycopeptides and glycosylation sites in supernatant and pellet fraction were 1582 ± 14 , 1609 ± 19 and 1706 ± 93 , 1758 ± 88 respectively. After removed the redundant identifications, the unique glycopeptides in whole plasma, supernatant and pellet fraction were 161 ± 15 (Supplementary data 1), 167 ± 3 (Supplementary data 2), and 145 ± 6 (Supplementary data 3) respectively. In total, 246 unique glycopeptides were identified in all fractions (Supplementary data 5), this represents $[(246 - 161)/161] * 100\%$ or $>50\%$ increase in unique glycopeptides over the unfractionated whole plasma. Some unique glycopeptides were identified only in whole plasma which could be due to non-pelleted lipid associated proteins. The comparison of unique glycopeptides in whole plasma, supernatant and pellet fraction revealed glycoproteome coverage enhancement in supernatant and pellet fractions. Percent glycopeptides in total number of collected peptides increased with shallow increase in gradient (Fig. 3). The proteins name, peptide sequence and sites of modification are shown in Table S1 (Supplementary information). The spectra showing MS/MS fragmentation profile of the peptides, identified Mascot score and peptide expect value (< 0.05), their sequence and sites of modification were provided as [annotated tandem mass spectral format](#) in supplementary information (Supplementary data 1: [annotated tandem mass spectra](#) of glycopeptides detected in whole plasma; Supplementary data 2:

1 [annotated tandem mass spectra](#) of glycopeptides detected in supernatant fraction of pig
2 plasma; Supplementary data 3: [annotated tandem mass spectra](#) of glycopeptides detected in
3 pellet fraction of pig plasma; Supplementary data 4: [annotated tandem mass spectra](#) of
4 glycopeptides detected in supernatant and pellet fractions of pig plasma; Supplementary data
5 5, [annotated tandem mass spectra](#) of glycopeptides unfractionated whole and S&P fractions
6 of pig plasma). The advantage of ERLIC is that it simultaneously retains and separates
7 different glycopeptides according to the peptides hydrophilicity and charges from salic acid
8 and carboxylic groups, thus a very few glycopeptides get eluted in flow through (fraction 1-5).
9 SCX or other techniques exploit salts or non-volatile salts in the gradient buffers and hence
10 require subsequent desalting of each fraction which may results into loss of hydrophilic
11 peptides.⁴⁵ On the contrary, ERLIC technique uses salt free volatile solvents for sample
12 loading and peptide elution, and thus loss of peptide during desalting can be eliminated.

13 The elution profile of glycopeptides and unique glycopeptides is shown in Fig. 4. The
14 percentage of glycopeptides in total glycopeptide increased with elution time in whole plasma,
15 supernatant and pellet fraction. We further analyzed unique glycosylation sites (Fig. 5A) and
16 unique glycoproteins (Fig. 5B). The glycosylation site analysis revealed 83 glycosylation
17 sites common in all analyzed samples i.e. whole plasma, supernatant and pellet. However, 23,
18 37 and 42 glycosylation sites were unique in whole plasma, supernatant and pellet fraction
19 indicating that sample preparation by single ultracentrifugation step enhances unique
20 glycoprotein coverage. We mapped these unique glycosylation sites in the proteins and
21 further analyzed it by overlap diagram (Fig. 5B). Thus, 48 proteins were common in all three
22 samples i.e. whole plasma, supernatant and pellet. But 12 and 13 unique glycoproteins were
23 identified in supernatant and pellet sample. Again, 8 unique glycoproteins were identified in
24 the whole plasma sample. Although these unique glycoproteins are named as uncharacterized
25 protein but based on the gene symbol these proteins were tissue factor pathway inhibitor,

coagulation factor XII, contactin-3 (CNTN3), plasma protease C1 inhibitor (SERPING1), protein HEG homolog 1 (HEG1), hematopoietic progenitor cell antigen CD34, intercellular adhesion molecule 2 (ICAM2) and lumican (LUM). Tissue factor pathway inhibitor (TFPI), a plasma Kunitz-type serine protease inhibitor, which modulates initiations of coagulation induced by TF, remain complexed with plasma lipoproteins.⁴⁶ The most popular methods adopted for lipoprotein separation are ultracentrifugation which separates lipoproteins according to hydrated densities and floatation rates. However, proper separation of lipoprotein by single cycle of ultracentrifugation might not be achieved efficiently.⁴³ Coagulation factor XII that participates in the initiation of blood coagulation, fibrinolysis, and the generation of bradykinin and angiotensin through serine-type endopeptidase activity.⁴⁷ According to published literature,³⁸ coagulation factor XIII A was omitted in MRM study due to poor proteolytic cleavage contributed to the reduced digestion efficiency. Glycoproteins like alpha-mannosidase and complement component C8G was common between whole plasma and pellet samples. Based on the gene symbol, glycosylated proteins like complement component C9, C8B protein, VNN3 protein, alpha-2-macroglobulin, intercellular adhesion molecule (ICAM-1), protein HEG homolog 1, leucine-rich repeat and coiled-coil domain-containing protein 1 were unique in supernatant sample

Apart from free proteins, lipids and metabolites; blood plasma also contains cell-derived extracellular vesicles, including exosomes and microparticles from several different cellular origins. The proteins like alpha-1-acid glycoprotein, coagulation factor XII, apolipoprotein D (APOD) and many more enriched in this study were recently shown to be associated with circulating extracellular vesicles (EVs).⁴⁸ Alpha-1-acid glycoprotein that has been treated as a potential cancer biomarker proteins in human plasma³⁸, was enriched with protein score of 8372 and identified with 304 peptides in supernatant fraction. Other EVs such as apolipoprotein D, coagulation factor XII, hemopexin, haptoglobin, alpha-1B-

glycoprotein, alpha-1-antichymotrypsin 2, clusterin, vitronectin, serotransferrin and many more were identified with high protein score in this study. Earlier studies^{38, 48, 49}, that have reported these proteins associated with EVs followed extensive ultracentrifugation coupled sucrose gradient or immunoprecipitation technique for their isolation and mostly focused on identification and their concentrations in plasma samples. Interestingly, our study found these circulating EVs are glycosylated and this information could be very useful to interpret their role with new insights as well as their role as potential biomarkers for a variety of human cancers, from laryngeal to ovarian, with breast cancer having the highest correlation.^{38, 48, 49}

Proteins like angiopoietin-related protein 6 (ANGPTL6, tr|F1S3I1|), laminin subunit beta-1 (LAMB1, tr|F1SAE9|), alpha-1-antichymotrypsin 2 (SERPINA3-2, tr|Q9GMA6|), alpha-mannosidase (MAN2A1, tr|F1RM55|), vasorin (VASN, tr|I3LS87|), alpha-amylase (tr|F1S573|), corticosteroid-binding globulin (Serpina6, sp|Q9GK37|), amine oxidase (AOC3, tr|F1S1G8|), alpha-2-macroglobulin (A2M, tr|F1SLX2|) and many more (Supplementary table S1) were found to be glycosylated. In addition to these, large numbers of uncharacterized proteins were found as glycosylated. However, their biological function and impact of glycosylation need to be established. The low abundant proteins such as integrin beta (0.000011 mg/mL),⁴⁹ complement C1q subcomponent subunit A (0.0012 mg/mL),⁴⁹ C4b-binding protein beta chain (0.00053 mg/mL)⁴⁹ and many more were identified only in pellet fraction. Glycoproteins like thyroxine-binding globulin (0.0013 mg/mL),⁴⁹ C4b-binding protein beta chain (0.00053 mg/mL),⁴⁹ clusterin (0.0914 mg/mL)⁵⁰ etc. were detected only in pellet. Thus, sample preparation by single step ultracentrifugation and adopting ERLIC-glycopeptide enrichment coupled to RPLC-MS/MS analysis can potentially enhance identification of low abundant glycoproteins that have been considered as a potential disease biomarker^{38, 48-55} (Table S2).

Functional analysis of unique glycoprotein

To evaluate importance of these unique glycoproteome, one should know their function, cellular location and biological processes in which they are involved. To assess the molecular functions, biological processes and their cellular location, we followed Gene Ontology (GO). The unique glycoprotein data was analyzed using web tools available at <http://www.agbase.msstate.edu>.^{56, 57} GO molecular function categorization of unique glycoprotein (Fig. 6A) categorized them into molecular functions (38%), ion binding (17.5%), peptidase activity (9.0%), enzyme regulator activity (7.2%), enzyme binding (4.8%), signal transducer activity (3.6%), lipid binding (3.6%), RNA binding (2.4%), oxidoreductase activity (2.4%), transmembrane transporter activity (2.4%) etc. This indicates that glycoproteins play role in wide variety of protein function. The principle biological processes in which these unique glycoproteins involved were biological process (20.9%), response to stress (6.8%), anatomical structure development (6.1%), immune system process (5.9%), transport (4.7%), signal transduction (4.7%), cell differentiation (4.3%), cell adhesion (4.1%), protein maturation (3.4%), cell proliferation (2.5%), biosynthetic process (2.3), vesicle-mediated transport (2.0%), cellular nitrogen compound metabolic process (2.0%), growth (2.0%), homeostatic process (2.0%), cell motility (2.0%) and many more (Fig. 6B).

Protein glycosylation has multiple functions in the cell e.g. in the ER, glycosylation is used to monitor the status of protein folding, acting as a quality control mechanism to ensure proper protein folding. Again, sugar moieties on soluble proteins help to bind proteins to specific receptors in the *trans* Golgi network to facilitate their delivery to the correct destination. Thus, protein glycosylation study is important in understanding underlying molecular mechanism of the disease. Alpha-1-acid glycoprotein (AGP) is a 41-43 kDa glycoprotein with a pI of 2.8-3.8 and one of the major acute phase proteins in humans, rats, mice and other species. Its biological function has been correlated with number of activities

1 including various immunomodulating effects, transport of many drugs and drug bindings.^{58, 59}

2 Thus, protein glycosylation could be important in drug transport and drug binding. Altered
3 glycosylation of AGP in patients with inflammation and diabetes mellitus has been
4 documented⁶⁰. EVs containing proteins like AGP, apolipoprotein I, coagulation factor XII,
5 haptoglobin, alpha-1B-glycoprotein are potential candidate cancer biomarker and their
6 concentrations in plasma has been estimated by multiplexed MRM-based quantitation.³⁸

7 Glycosyltransferases are located in the endoplasmic reticulum (ER) and golgi
8 apparatus and N-glycosylation is typically restricted to secreted proteins and extracellular
9 membrane proteins.⁶¹ Cellular localization of unique glycoproteins revealed their association
10 with cellular component (18.0%), extracellular region (15.8%), extracellular space (12.5%),
11 organelle (12.3%), cell (9.3%), plasma membrane (6.3%) and protein complex (3.5%) (Fig.
12 6C). However, some of glycoproteins exhibited atypical localization such as (cytoplasm,
13 5.3%; intracellular, 4.8%). Variations in the concentration of glycosyltransferase changes the
14 degree of glycosylation of multiple proteins.⁶² Techniques used for the enrichment of
15 glycopeptides and glycoproteins have varied with time and among researchers. Most of the
16 studies depleted high abundant proteins to identify low abundant glycoproteins. However, our
17 study could able to detect low abundant glycoprotein proteins, without adopting deletion of
18 high abundant proteins and thus represents reasonable strategy for enriching and identifying
19 low-abundance glycoproteins. Further, glycosylated proteins in circulating extracellular
20 vesicles (EVs) in plasma were identified. Lectin-based affinity enrichment is one of the
21 earliest and most widely used approaches and this technique need use of two or more types of
22 lectins to recover glycoproteins. Combining the use of two types of lectin (Con A and WGA),
23 a total of 77 human serum glycoproteins were identified by LC MS/MS⁶³. Approximately 130
24 sialylated glycoproteins were identified in human cancer serum by adopting lectin affinity
25 selection, a liquid separation and μ LC-MS/MS.⁶⁴ Using lectin affinity capture, isotope-coded

1 tagging and mass spectrometry, Kaji et al.⁶⁵ identified 250 glycoproteins with 400 unique N-
2 glycosylation sites from an extract of *Caenorhabditis elegans* proteins. However, this study
3 identified significantly higher number of glycoprotein and glycosylation sites. The glycans
4 are not static and vary temporally with disease, and may be responsible for modulating as
5 opposed to initiating a biological function as in a direct protein–protein interaction.⁶⁶ Protein
6 glycosylation study by mass spectrometry, however, has its own challenges including highly
7 labile nature of glycosylations, especially when the glycans are bound to threonine or serine
8 residues through an acetal bond.² Thus, such liability of glycan complicates analyses and
9 challenges the reproducibility. Another important issue with glycosylation study is poor
10 ionization efficiencies of glycosylated peptides. Yet again, glycan modifications larger than a
11 single sugar residue can be very heterogeneous and are not particularly amenable to current
12 bioinformatics platforms. If single peptide is modified by several forms of glycans, it
13 complicates identification and sites of glycosylation. Coupling glycoprotein fractionation,
14 glycopeptide enrichment, and PNGase-F treatment that simplifies glycoprotein analysis and
15 allows identification of protein glycosylation sites by mass spectrometry could be useful
16 approach for identification of glycosylation sites.

19 **Conclusions**

20 Plasma contains proteins and cell-derived extracellular vesicles from several different cellular
21 origins and hence considered as an excellent source of disease biomarkers. A focus on global
22 glycoprotein analysis has recently resulted in technical advances in the field with a major
23 focus on biomarkers. This study demonstrated the novelty of coupling ultracentrifugation for
24 glycoprotein fractionation and ERLIC for glycopeptide enrichment, followed with RPLC-
25 MS/MS analysis that improves glycoprotein coverage and promotes identification of variety

of glycoproteins in blood plasma. The simple reduction of test sample complexity via supernatant and pellet separation by ultracentrifugation enhances coverage of proteome and glycoproteome. Adopting ERLIC enrichment for further RPLC-MS/MS analysis revealed 161, 167 and 145 unique glycopeptides respectively in whole plasma, supernatant and pellet samples. In total, 246 unique glycopeptides were identified in all fractions, this represents >50% increase in unique glycopeptides over the unfractionated whole plasma. The novel method reported in this study, could be used to identify potential disease glycoprotein biomarkers in human plasma.

Conflict of interest

None for this study

Supporting Information:

Figure S1: experimental design and HPLC fractionation chromatogram of tryptic peptides; Table S1: glycosylated proteins along with their peptide sequence and site of glycosylation sites; Supplementary data 1: [Annotated tandem mass spectra](#) of glycopeptides detected in whole plasma; Supplementary data 2: [Annotated tandem mass spectra](#) of glycopeptides detected in supernatant fraction of pig plasma; Supplementary data 3: [Annotated tandem mass spectra](#) of glycopeptides detected in pellet fraction of pig plasma; Supplementary data 4: [Annotated tandem mass spectra](#) of glycopeptides detected in supernatant and pellet fraction of pig plasma Supplementary data 5: [Annotated tandem mass spectra](#) of glycopeptides detected in whole plasma, supernatant and pellet fractions of pig plasma. This material is provided free of charge via the Internet at <http://pubs.acs.org>.

Acknowledgements

This work is in part supported by grants from the Singapore Ministry of Education (Tier

1: RGT15/13), NTU-NHG Ageing Research Grant (ARG/14017) and NTU iFood (grant #: S006).

References

- (1) Mann, M.; Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **2003**, *21*, (3), 255-261.
- (2) Budnik, B. A.; Lee, R. S.; Steen, J. A. J. Global methods for protein glycosylation analysis by mass spectrometry. *Biochimica et Biophysica Acta - Proteins and Proteomics* **2006**, *1764*, (12), 1870-1880.
- (3) Rebecchi, K. R.; Wenke, J. L.; Go, E. P.; Desaire, H. Label-Free Quantitation: A New Glycoproteomics Approach. *J. Am. Soc. Mass. Spectrom.* **2009**, *20*, (6), 1048-1059.
- (4) Morelle, W.; Canis, K.; Chirat, F.; Faid, V.; Michalski, J. C. The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics* **2006**, *6*, (14), 3993-4015.
- (5) Lee, J.; Park, J. S.; Moon, J. Y.; Kim, K. Y.; Moon, H. M. The influence of glycosylation on secretion, stability, and immunogenicity of recombinant HBV pre-S antigen synthesized in *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* **2003**, *303*, (2), 427-432.
- (6) Morris, H. R.; Chalabi, S.; Panico, M.; Sutton-Smith, M.; Clark, G. F.; Goldberg, D.; Dell, A. Glycoproteomics: Past, present and future. *Int. J. Mass spectrom.* **2007**, *259*, (1-3), 16-31.
- (7) Tissot, B.; North, S. J.; Ceroni, A.; Pang, P. C.; Panico, M.; Rosati, F.; Capone, A.; Haslam, S. M.; Dell, A.; Morris, H. R. Glycoproteomics: Past, present and future. *FEBS Lett.* **2009**, *583*, (11), 1728-1735.
- (8) Geyer, H.; Geyer, R. Strategies for analysis of glycoprotein glycosylation. *Biochimica et Biophysica Acta - Proteins and Proteomics* **2006**, *1764*, (12), 1853-1869.
- (9) Watkins, W. M. Blood-group substances. *Science* **1966**, *152*, (3719), 172-181.
- (10) Kraus, M. H.; Fedi, P.; Starks, V.; Muraro, R.; Aaronson, S. A. Demonstration of ligand-dependent signaling by the erbB-3 tyrosine kinase and its constitutive activation in human breast tumor cells. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, (7), 2900-2904.
- (11) Seelenmeyer, C.; Wegehingel, S.; Lechner, J.; Nickel, W. The cancer antigen CA125 represents a novel counter receptor for galectin-1. *J. Cell Sci.* **2003**, *116*, (7), 1305-1318.
- (12) Ahn, Y. H.; Kim, K. H.; Shin, P. M.; Ji, E. S.; Kim, H.; Yoo, J. S. Identification of low-abundance cancer biomarker candidate TIMP1 from serum with lectin fractionation and peptide affinity enrichment by ultrahigh-resolution mass spectrometry. *Anal. Chem.* **2012**, *84*, (3), 1425-1431.
- (13) Berven, F. S.; Ahmad, R.; Clauser, K. R.; Carr, S. A. Optimizing performance of glycopeptide capture for plasma proteomics. *J. Proteome Res.* **2010**, *9*, (4), 1706-1715.
- (14) Mitri, Z.; Constantine, T.; O'Regan, R. The HER2 Receptor in Breast Cancer: Pathophysiology, Clinical Use, and New Advances in Therapy. *Chemotherapy Research and Practice* **2012**, *2012*, 743193.
- (15) Rudd, P. M.; Dwek, R. A. Glycosylation: Heterogeneity and the 3D structure of proteins. *Crit. Rev. Biochem. Mol. Biol.* **1997**, *32*, (1), 1-100.
- (16) Küster, B.; Krogh, T. N.; Mørtz, E.; Harvey, D. J. Glycosylation analysis of gel-separated proteins. *Proteomics* **2001**, *1*, (2), 350-361.

- (17) Alpert, A. J. Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal. Chem.* **2008**, *80*, (1), 62-76.
- (18) Hao, P.; Zhang, H.; Sze, S. K., Application of electrostatic repulsion hydrophilic interaction chromatography to the characterization of proteome, glycoproteome, and phosphoproteome using nano LC-MS/MS. In *Methods in Molecular Biology*, Toms, S. A.; Weil, R. J., Eds. 2011; Vol. 790, pp 305-318.
- (19) Hao, P.; Guo, T.; Sze, S. K. Simultaneous analysis of proteome, phospho- and glycoproteome of rat kidney tissue with electrostatic repulsion hydrophilic interaction chromatography. *PLoS One* **2011**, *6*, (2).
- (20) Hao, P.; Qian, J.; Dutta, B.; Cheow, E. S. H.; Sim, K. H.; Meng, W.; Adav, S. S.; Alpert, A.; Sze, S. K. Enhanced separation and characterization of deamidated peptides with RP-ERLIC-based multidimensional chromatography coupled with tandem mass spectrometry. *J. Proteome Res.* **2012**, *11*, (3), 1804-1811.
- (21) Hao, P.; Sze, S. K. Proteomic analysis of protein deamidation. *Curr Protoc Protein Sci* **2014**, *78*, (24), 1-24.
- (22) Wuhler, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* **2007**, *849*, (1-2), 115-128.
- (23) Alvarez-Manilla, G.; Atwood Iii, J.; Guo, Y.; Warren, N. L.; Orlando, R.; Pierce, M. Tools for glycoproteomic analysis: Size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites. *J. Proteome Res.* **2006**, *5*, (3), 701-708.
- (24) Alvarez-Manilla, G.; Warren, N. L.; Abney, T.; Atwood Iii, J.; Azadi, P.; York, W. S.; Pierce, M.; Orlando, R. Tools for glycomics: Relative quantitation of glycans by isotopic permethylation using ¹³CH₃I. *Glycobiology* **2007**, *17*, (7), 677-687.
- (25) Jensen, O. N. Modification-specific proteomics: Characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* **2004**, *8*, (1), 33-41.
- (26) Xiong, L.; Andrews, D.; Regnier, F. Comparative Proteomics of Glycoproteins Based on Lectin Selection and Isotope Coding. *J. Proteome Res.* **2003**, *2*, (6), 618-625.
- (27) 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.* **2003**, *21*, (6), 660-666.
- (28) Wohlgemuth, J.; Karas, M.; Eichhorn, T.; Hendriks, R.; Andrecht, S. Quantitative site-specific analysis of protein glycosylation by LC-MS using different glycopeptide-enrichment strategies. *Anal. Biochem.* **2009**, *395*, (2), 178-188.
- (29) Gan, C. S.; Guo, T.; Zhang, H.; Lim, S. K.; Sze, S. K. A comparative study of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) versus SCX-IMAC-based methods for phosphopeptide isolation/enrichment. *J. Proteome Res.* **2008**, *7*, (11), 4869-4877.
- (30) Hao, P.; Ren, Y.; Alpert, A. J.; Siu, K. S. Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. *Mol. Cell. Proteomics* **2011**, *10*, (10).
- (31) Blomme, B.; Van Steenkiste, C.; Callewaert, N.; Van Vlierberghe, H. Alteration of protein glycosylation in liver diseases. *J. Hepatol.* **2009**, *50*, (3), 592-603.
- (32) Chatterjee, B. P.; Mondal, G.; Chatterjee, U. Glycosylation of Acute Phase Proteins: A Promising Disease Biomarker. *Proceedings of the National Academy of Sciences India Section B - Biological Sciences* **2014**, *84*, (4), 865-874.

- (33) Adav, S. S.; Qian, J.; Ang, Y. L.; Kalaria, R. N.; Lai, M. K. P.; Chen, C. P.; Sze, S. K. ITRAQ quantitative clinical proteomics revealed role of Na⁺K⁺-ATPase and its correlation with deamidation in vascular dementia. *J. Proteome Res.* **2014**, *13*, (11), 4635-4646.
- (34) Beck, F.; Lewandrowski, U.; Wiltfang, M.; Feldmann, I.; Geiger, J.; Sickmann, A.; Zahedi, R. P. The good, the bad, the ugly: Validating the mass spectrometric analysis of modified peptides. *Proteomics* **2011**, *11*, (6), 1099-1109.
- (35) Adav, S. S.; Ravindran, A.; Sze, S. K. Study of *Phanerochaete chrysosporium* secretome revealed protein glycosylation as a substrate-dependent post-translational modification. *J. Proteome Res.* **2014**, *13*, (10), 4272-4280.
- (36) Tretter, V.; Altmann, F.; Marz, L. Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1----3 to the asparagine-linked N-acetylglucosamine residue. *Eur. J. Biochem.* **1991**, *199*, (3), 647-52.
- (37) Polaskova, V.; Kapur, A.; Khan, A.; Molloy, M. P.; Baker, M. S. High-abundance protein depletion: comparison of methods for human plasma biomarker discovery. *Electrophoresis* **2010**, *31*, (3), 471-82.
- (38) Percy, A. J.; Chambers, A. G.; Yang, J.; Borchers, C. H. Multiplexed MRM-based quantitation of candidate cancer biomarker proteins in undepleted and non-enriched human plasma. *Proteomics* **2013**, *13*, (14), 2202-2215.
- (39) Schlosser, A.; Pipkorn, R.; Bossemeyer, D.; Lehmann, W. D. Analysis of protein phosphorylation by a combination of elastase digestion and neutral loss tandem mass spectrometry. *Anal. Chem.* **2001**, *73*, (2), 170-6.
- (40) Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, (7), 2199-204.
- (41) Zhang, K.; Liu, Y.; Shang, Y.; Zheng, H.; Guo, J.; Tian, H.; Jin, Y.; He, J.; Liu, x. Analysis of pig serum proteins based on shotgun liquid chromatography-tandem mass spectrometry. *African Journal of Biotechnology* **2012**, *11*, (57), 12118-12127.
- (42) Małgorzata, O.; Adam, L.; Agnieszka, H. Two-dimensional gel-based serum protein profile of growing piglets. *Turkish Journal of Biology* **2015**, doi:10.3906/biy-1408-45.
- (43) Havel, R. J.; Eder, H. A.; Bragdon, J. H. The distribution and chemical composition of ultracentrifugally separated lipoproteins in human serum. *J. Clin. Invest.* **1955**, *34*, (9), 1345-53.
- (44) Hoener, M. C.; Brodbeck, U. Phosphatidylinositol-glycan-specific phospholipase D is an amphiphilic glycoprotein that in serum is associated with high-density lipoproteins. *Eur. J. Biochem.* **1992**, *206*, (3), 747-57.
- (45) Zhang, H.; Guo, T.; Li, X.; Datta, A.; Park, J. E.; Yang, J.; Lim, S. K.; Tam, J. P.; Sze, S. K. Simultaneous characterization of glyco- and phosphoproteomes of mouse brain membrane proteome with electrostatic repulsion hydrophilic interaction chromatography. *Mol. Cell. Proteomics* **2010**, *9*, (4), 635-647.
- (46) Lwaleed, B. A.; Bass, P. S. Tissue factor pathway inhibitor: structure, biology and involvement in disease. *J. Pathol.* **2006**, *208*, (3), 327-39.
- (47) Kwon, M. J.; Kim, H. J.; Lee, K. O.; Jung, C. W.; Kim, S. H. Molecular genetic analysis of Korean patients with coagulation factor XII deficiency. *Blood Coagul. Fibrinolysis* **2010**, *21*, (4), 308-12.
- (48) Bastos-Amador, P.; Royo, F.; Gonzalez, E.; Conde-Vancells, J.; Palomo-Diez, L.; Borrás, F. E.; Falcon-Perez, J. M. Proteomic analysis of microvesicles from plasma of healthy donors reveals high individual variability. *J. Proteomics* **2012**, *75*, (12), 3574-3584.
- (49) Farrah, T.; Deutsch, E. W.; Omenn, G. S.; Campbell, D. S.; Sun, Z.; Bletz, J. A.; Mallick, P.; Katz, J. E.; Malmström, J.; Ossola, R.; Watts, J. D.; Lin, B.; Zhang, H.; Moritz, R.

L.; Aebersold, R. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell. Proteomics* **2011**, *10*, (9).

(50) Looze, C.; Yui, D.; Leung, L.; Ingham, M.; Kaler, M.; Yao, X.; Wu, W. W.; Shen, R. F.; Daniels, M. P.; Levine, S. J. Proteomic profiling of human plasma exosomes identifies PPAR γ as an exosome-associated protein. *Biochem. Biophys. Res. Commun.* **2009**, *378*, (3), 433-438.

(51) Nielsen, M. J.; Petersen, S. V.; Jacobsen, C.; Oxvig, C.; Rees, D.; Møller, H. J.; Moestrup, S. K. Haptoglobin-related protein is a high-affinity hemoglobin-binding plasma protein. *Blood* **2006**, *108*, (8), 2846-2849.

(52) Nielsen, M. J.; Petersen, S. V.; Jacobsen, C.; Thirup, S.; Enghild, J. J.; Graversen, J. H.; Moestrup, S. K. A unique loop extension in the serine protease domain of haptoglobin is essential for CD163 recognition of the haptoglobin-hemoglobin complex. *J. Biol. Chem.* **2007**, *282*, (2), 1072-1079.

(53) Udby, L.; Cowland, J. B.; Johnsen, A. H.; Sørensen, O. E.; Borregaard, N.; Kjeldsen, L. An ELISA for SGP28/CRISP-3, a cysteine-rich secretory protein in human neutrophils, plasma, and exocrine secretions. *J. Immunol. Methods* **2002**, *263*, (1-2), 43-55.

(54) Percy, A. J.; Chambers, A. G.; Smith, D. S.; Borchers, C. H. Standardized protocols for quality control of MRM-based plasma proteomic workflows. *J. Proteome Res.* **2013**, *12*, (1), 222-233.

(55) Percy, A. J.; Chambers, A. G.; Yang, J.; Domanski, D.; Borchers, C. H. Comparison of standard- and nano-flow liquid chromatography platforms for MRM-based quantitation of putative plasma biomarker proteins. *Anal. Bioanal. Chem.* **2012**, *404*, (4), 1089-1101.

(56) McCarthy, F. M.; Gresham, C. R.; Buza, T. J.; Chouvarine, P.; Pillai, L. R.; Kumar, R.; Ozkan, S.; Wang, H.; Manda, P.; Arick, T.; Bridges, S. M.; Burgess, S. C. AgBase: Supporting functional modeling in agricultural organisms. *Nucleic Acids Res.* **2011**, *39*, (SUPPL. 1), D497-D506.

(57) McCarthy, F. M.; Wang, N.; Magee, G. B.; Nanduri, B.; Lawrence, M. L.; Camon, E. B.; Barrell, D. G.; Hill, D. P.; Dolan, M. E.; Williams, W. P.; Luthe, D. S.; Bridges, S. M.; Burgess, S. C. AgBase: A functional genomics resource for agriculture. *BMC Genomics* **2006**, *7*.

(58) Fournier, T.; Medjoubi, N. N.; Porquet, D. Alpha-1-acid glycoprotein. *Biochim. Biophys. Acta* **2000**, *1482*, (1-2), 157-71.

(59) Huang, Z.; Ung, T. Effect of alpha-1-acid glycoprotein binding on pharmacokinetics and pharmacodynamics. *Curr. Drug Metab.* **2013**, *14*, (2), 226-38.

(60) Higai, K.; Azuma, Y.; Aoki, Y.; Matsumoto, K. Altered glycosylation of alpha1-acid glycoprotein in patients with inflammation and diabetes mellitus. *Clin. Chim. Acta* **2003**, *329*, (1-2), 117-25.

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

(62) Hashii, N.; Kawasaki, N.; Itoh, S.; Hyuga, M.; Kawanishi, T.; Hayakawa, T. Glycomic/glycoproteomic analysis by liquid chromatography/mass spectrometry: Analysis of glycan structural alteration in cells. *Proteomics* **2005**, *5*, (18), 4665-4672.

(63) Bunkenborg, J.; Pilch, B. J.; Podtelejnikov, A. V.; Wiśniewski, J. R. Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* **2004**, *4*, (2), 454-465.

(64) Zhao, J.; Simeone, D. M.; Heidt, D.; Anderson, M. A.; Lubman, D. M. Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass

spectrometric analysis: Application to pancreatic cancer serum. *J. Proteome Res.* **2006**, *5*, (7), 1792-1802.

(65) Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K. I.; Takahashi, N.; Isobe, T. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol.* **2003**, *21*, (6), 667-672.

(66) Raman, R.; Raguram, S.; Venkataraman, G.; Paulson, J. C.; Sasisekharan, R. Glycomics: An integrated systems approach to structure-function relationships of glycans. *Nat. Methods* **2005**, *2*, (11), 817-824.

Figure legends

Figure 1. Venn diagram showing the overlap of the proteins identified in different whole plasma, supernatant and pellet sample.

Figure 2. Cellular location and classification of identified proteome of pig plasma based on molecular function. The number in parentheses indicates percentage of proteins.

Figure 3. Percent glycopeptides in total modified peptides in whole plasma, supernatant and pellet sample.

Figure 4. Glycopeptides and unique glycopeptides elution profile.

Figure 5. Unique glycoproteins in whole plasma, pellet and supernatant fraction.

Figure 6. Cellular location and classification of unique glycoproteome of pig plasma based on molecular function.

1 Figure 1

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

1 Figure 2

2
3

Figure 3

1 Figure 4

2
3
4
5
6
7
8
9
10
11

Figure 5

(A) Unique glycosylation sites

(B) Unique glycoproteins

1 Figure 6