analytical chemistry

Article

Cite This: Anal. Chem. XXXX, XXX, XXX−XXX

pubs.acs.org/ac

Analytical Pipeline for Discovery and Verification of Glycoproteins from Plasma-Derived Extracellular Vesicles as Breast Cancer Biomarkers

I-Hsuan Chen,[†] Hillary Andaluz Aguilar,[‡] J. Sebastian Paez Paez,[†] Xiaofeng Wu,[‡] Li Pan,[§] Michael K. Wendt,[§,#] Anton B. Iliuk,[∥] Ying Zhang,*[,⊥] and W. Andy Tao*[,†,‡,§,#]

[†]Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907, United States

[‡]Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States

[§]Department of Medicinal Chemistry & Molecular Pharmacology, Purdue University, West Lafayette, Indiana 47907, United States

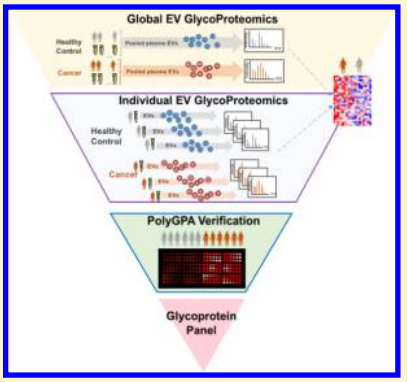
[∥]Tymora Analytical Operations, West Lafayette, Indiana 47906, United States

[⊥]Shanghai Cancer Center and Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China

[#]Purdue Center for Cancer Research, Purdue University, West Lafayette, Indiana 47907, United States

Supporting Information

ABSTRACT: Glycoproteins comprise more than half of current FDA-approved protein cancer markers, but the development of new glycoproteins as disease biomarkers has been stagnant. Here we present a pipeline to develop glycoproteins from extracellular vesicles (EVs) through integrating quantitative glycoproteomics with a novel reverse phase glycoprotein array and then apply it to identify novel biomarkers for breast cancer. EV glycoproteomics show promise in circumventing the problems plaguing current serum/plasma glycoproteomics and allowed us to identify hundreds of glycoproteins that have not been identified in blood. We identified 1,453 unique glycopeptides representing 556 glycoproteins in EVs, among which 20 were verified significantly higher in individual breast cancer patients. We further applied a novel glyco-specific reverse phase protein array to quantify a subset of the candidates. Together, this study demonstrates the great potential of this integrated pipeline for biomarker discovery.



The emerging liquid biopsy underscores our unyielding goal of achieving noninvasive disease diagnosis through blood tests.[1] With most proteins present in the blood being glycoproteins and aberrant glycosylation occurring in many diseases,[2] it is not surprising that most common FDA-approved biomarkers for cancer diagnosis and monitoring of malignant progression are glycoproteins.[3] However, plasma or serum proteomes contain a dynamic range of 12 orders of magnitude in protein concentration, thus analyzing glycoproteins in blood-derived plasma or serum to search for new biomarkers continues facing major challenges in terms of analytical sensitivity and depth.[4,5]

With increasing evidence about their important roles in cell−cell communication and relevance in the transmission of pathogenic and signaling molecules in diseases, extracellular vesicles (EVs) have been exploited as attractive sources for biomarker discovery and disease diagnosis.[6−8] Currently, most studies on EVs focus on mRNA and miRNA transfer and the role of proteins in EVs in particular their post-translational modifications (PTMs) has been rarely exploited.[9,10] PTMs increase the functional diversity of the proteome and influence almost all aspects of cell biology and pathogenesis. Thus, many PTMs are routinely tracked as disease markers, in particular glycoproteins mentioned above. Given that EVs are membrane-encapsulated packages, they are believed to carry a large assortment of resident cell-surface glycoproteins.[11] In theory, the glycoproteome of EVs should reflect their cellular origins and functions. Importantly, analyzing the glycoproteome in EVs instead of plasma or serum could eliminate the interference from highly abundant plasma components to a large extent, thus providing a wide dynamic range of detection and enabling the discovery of low-level glycoproteins at high sensitivity (as low as nanograms per milliliter).[12]

We present here an integrated pipeline that profiles glycoproteins from EVs through quantitative glycoproteomics using pooled and individual samples and then validated several targets using a novel reverse phase glycoprotein array termed polymer-based reverse phase glycoprotein array (polyGPA).[13] The lack of oligosaccharide-specific antibodies hinders the verification of the glycosylation changes on glycoproteins as biomarkers; as a result, developing glycoproteins as biomarkers in clinical settings has remained a huge challenge. Although mass spectrometry (MS) has been the driving force in profiling glycans and glycoproteomes for biomarker research,[14−16] MS-

Received: March 11, 2018
Accepted: April 9, 2018
Published: April 9, 2018

ACS Publications © XXXX American Chemical Society

DOI: 10.1021/acs.analchem.8b01090
Anal. Chem. XXXX, XXX, XXX−XXX

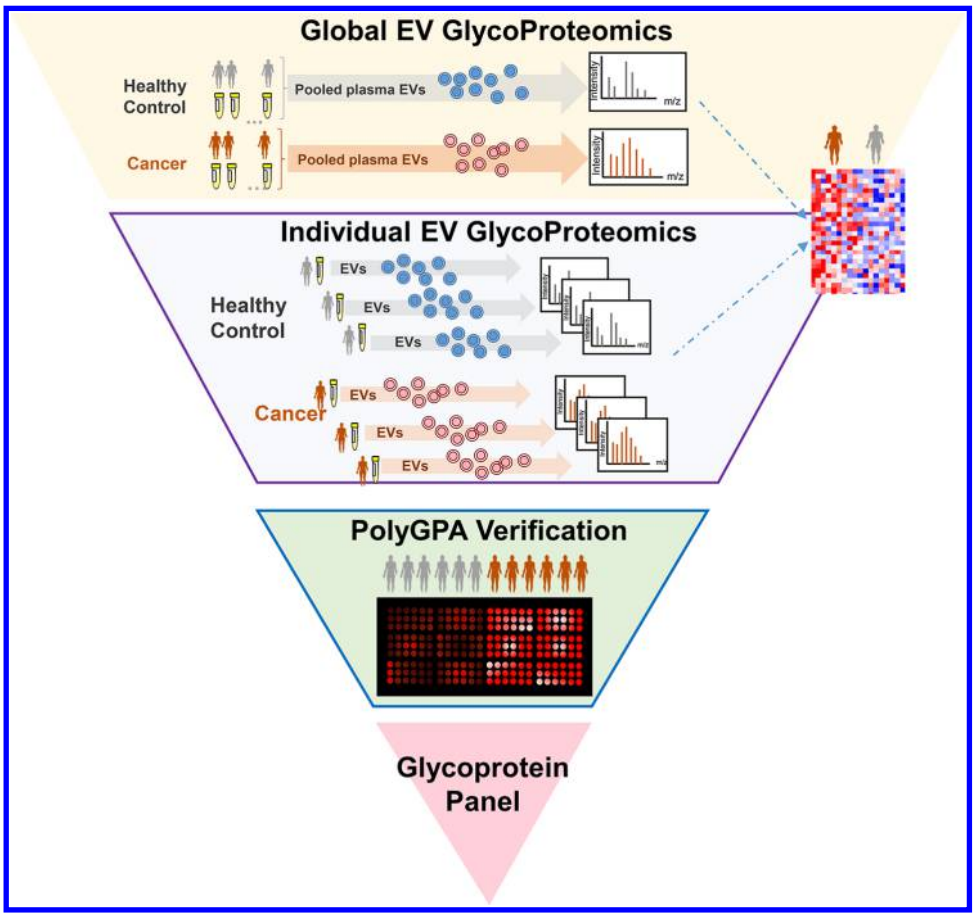


Figure 1. Workflow of the pipeline based on plasma EV glycoproteomics for biomarker discovery. See main text for details.

based glycoproteomics is typically used for in-depth profiling of glycoproteins during the discovery stage.[17,18] We applied polyGPA to validate several important glycoproteins with samples from patient and healthy individuals. We demonstrate here the universal performance of this pipeline and its value in discovering and validating glycoproteins in EVs as novel disease markers.

EXPERIMENTAL SECTION

Experimental details in materials, EV sample preparation, glycopeptide enrichment, LC-MS/MS analyses, and polyGPA analyses are included in the Supporting Information.

EV Sample Preparation. The Indiana University Institutional Review Board approved the use of human plasma samples. The EVs isolation and digestion were performed according to the reported protocol through high-speed and ultrahigh-speed centrifugation.[10] The digestion was performed with phase transfer surfactant aided (PTS) digestion,[19] and the resulting peptides were desalted using a 100 mg Sep-pak C18 column (Waters, Milford, MA, USA) for glycopeptides enrichment.

Glycopeptide Enrichment and LC-MS/MS Analysis. Peptides were oxidized with sodium periodate and captured by hydrazide magnetic beads according to a previous protocol.[13] After washing away the nonspecifically adsorbed peptides, PNGase F (NEB) was added to release the formerly *N*-glycosylated peptides to then be analyzed by liquid chromatography−tandem mass spectrometric analysis (LC-MS/MS). The Easy-nLC 1000 equipped with an in-house packed C18 column was coupled online with a LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific) for the LC-MS/MS analysis. All MS proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) with project accession no. PDX007572 via the PRIDE partner repository.[20]

RESULTS AND DISCUSSION

Identification of 1,453 Unique *N*-Glycopeptides from Plasma EV. An overview of the EV glycoprotein biomarker pipeline and its application to the identification of potential breast cancer biomarkers is illustrated in Figure 1. We first applied global quantitative *N*-glycoproteomic analyses with EVs, including microvesicles (MVs) and exosomes, using pooled samples from healthy and patient plasma, to generate a candidate biomarker list. Plasma samples were collected and pooled from healthy individuals (n = 6) and from patients diagnosed with breast cancer (n = 18). MVs and exosomes were isolated from human plasma through high-speed and ultrahigh-speed centrifugation, respectively. Characterization of EV isolation was evaluated using dynamic light scattering (DLS) (Figure 2A), MS, and immunoassay with multiple EV marker antibodies (Figure S1). The DLS data indicated that most MVs isolated after 20*K* centrifugation are in the range of 100−1000 nm while exosomes isolated by 100*K* centrifugation are in the range of 30−100 nm. MS and Western Blotting analyses identified several protein markers only in microvesicles or exosomes, but at the same time a few surface markers were identified in both microvesicles and exosomes, indicating there

DOI: 10.1021/acs.analchem.8b01090
Anal. Chem. XXXX, XXX, XXX−XXX

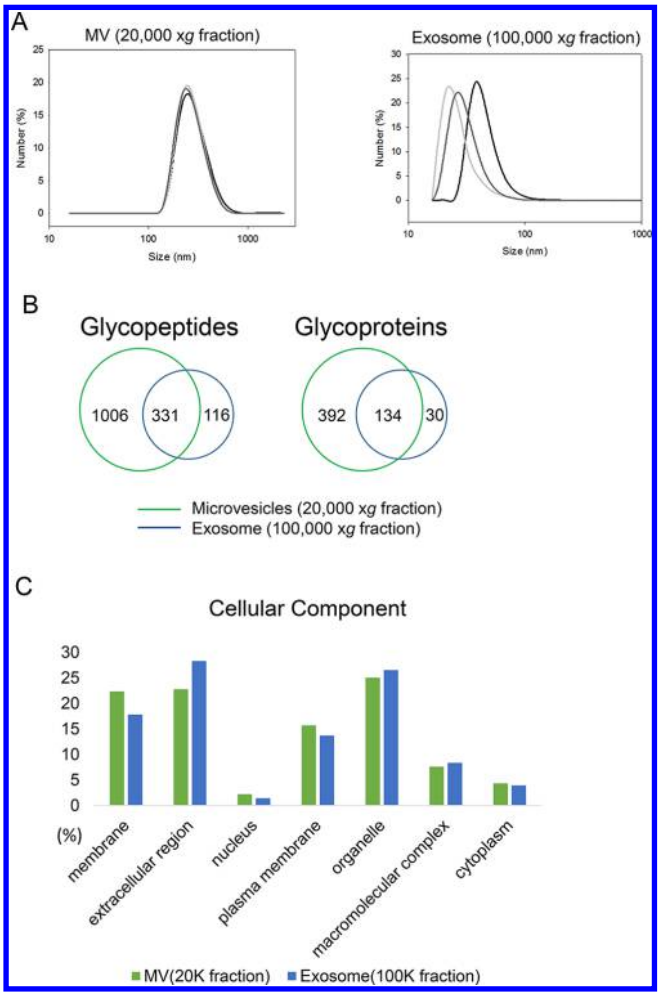


Figure 2. Characteristic analysis of glycoproteins in plasma-derived EVs. (A) Size distribution of EVs isolated from two high-speed centrifugations measured by DLS. Each line corresponds to one acquired result from a single sample. (B) Venn diagram showing the glycopeptides and glycoproteins identification overlap between microvesicles and exosome. (C) Classification of the identified glycoproteins in EVs based on their cellular component.

is no sharply defined definition of the plasma EV populations isolated by high- and ultrahigh-speed centrifugation.[21,22] After isolation, EVs were lysed and proteins were extracted and enzymatically digested with LysC and trypsin, followed by the hydrazide chemistry to enrich preoxidized glycopeptides. Formerly *N*-glycosylated peptides were recovered using PNGase F and analyzed by nanoflow LC-MS/MS. Three technical replicates were performed on the pooled samples, and label free quantitation was performed to measure glycopeptides in EV samples in the plasma of control and breast cancer patient samples.

We identified 1,453 unique glycopeptides, including 1,337 from microvesicles and 447 from exosomes, representing 526 and 164 glycoproteins in MV and exosomes, respectively (Figure 2B). Gene ontology analysis of the glycoproteins indicated a significant portion of the identified glycoproteins are from membrane, extracellular region, and organelles (Figure 2C). Overall, similar cellular components were observed for MV and exosomes. There is also significant overlap of identified glycopeptides and glycoproteins in MV and exosomes. With only 30 glycoproteins being unique in exosomes, we reasoned that it is not critical to differentiate glycoproteins in MV from those in exosomes for disease biomarker discovery and therefore all following data collected in MVs and exosomes in this study were combined and analyzed as EV *N*-glycoproteomes.

The current data reported here represent one of the largest *N*-glycoproteomic data sets using serum or plasma as the source. For direct comparison, we carried out a conventional *N*-glycoproteomic study using the breast cancer plasma samples. The conventional workflow with plasma samples resulted in a larger portion of high abundant plasma glycoproteins while EV glycoproteomics identified more glycoproteins in low abundance (Supporting Information Figure.S2A). We further examined the identified EV *N*-glycoproteins against previous reported serum/plasma glycoproteins. Strikingly, about one-quarter (126) of glycoproteins have not been previously reported as serum/plasma glycoproteins (Supporting Information Figure S2B). The data support our hypothesis that EVs are an ideal source to identify novel glycoproteins as potential disease biomarkers.

Cancer-Specific Glycoproteins in EV. Label-free quantitation of glycopeptides was performed to identify a list of glycoproteins changing in breast cancer. Quantitative *N*-glycoproteomics identified 77 glycopeptides that showed significant difference in abundance in breast cancer patients vs healthy controls (Figure 3A). The difference represents the abundance changes in glycoproteins or changes in glycosylation. To distinguish these factors, we also measured the abundance of the non-glycopeptides before enrichment by MS, as non-glycopeptides represent the abundance of the total protein expression. In comparison, there is a larger and wider difference in glycopeptides than in non-glycopeptides, indicating that some glycosylation differences between cancer patients and healthy individuals are not due to changes in protein expression, and thus reflect true cancer patient-specific glycosylation (glycosylation occupancy differences or total glycoprotein amount changes) (Figure 3B).

To validate the quantitation data with pooled samples, we then carried out label-free quantitative EV *N*-glycoproteomics with individual plasma samples using another cohort of sample, including 18 patients with breast cancer and 10 healthy controls. Glycoproteins with significantly increased glycosylation in patient samples were identified by the *p*-value from a two sample *t* test with a permutation-based FDR cutoff 0.05 with S0 set on 0.2. The imputed data set was further normalized by *z*-score for the heat map analysis, and together, we identified a total of 20 glycoproteins specific in patients with 21 unique glycosylation sites (p-value < 0.05) (Figure 3C).

Verification of Specific Glycoprotein Changes in Cancer Patients via polyGPA. Validation of biomarkers has typically been carried out using antibody-based sandwich assays such as ELISA or targeted quantitative MS methods such as selected reaction monitoring (SRM) and multiple reaction monitoring (MRM). However, there are virtually no glyco-specific antibodies available commercially. On the other hand, the development of SRM/MRM assays requires a great deal of effort including the high cost of synthetic stable isotope labeled peptides, in particular, formerly *N*-glycosylated peptides in this case.

We have recently developed a three-dimensionally functionalized reverse phase protein array, polyGPA, to validate glycoproteins in high throughput with high specificity, high sensitivity, and good quantitative capabilities.[13] PolyGPA uses

DOI: 10.1021/acs.analchem.8b01090
Anal. Chem. XXXX, XXX, XXX−XXX

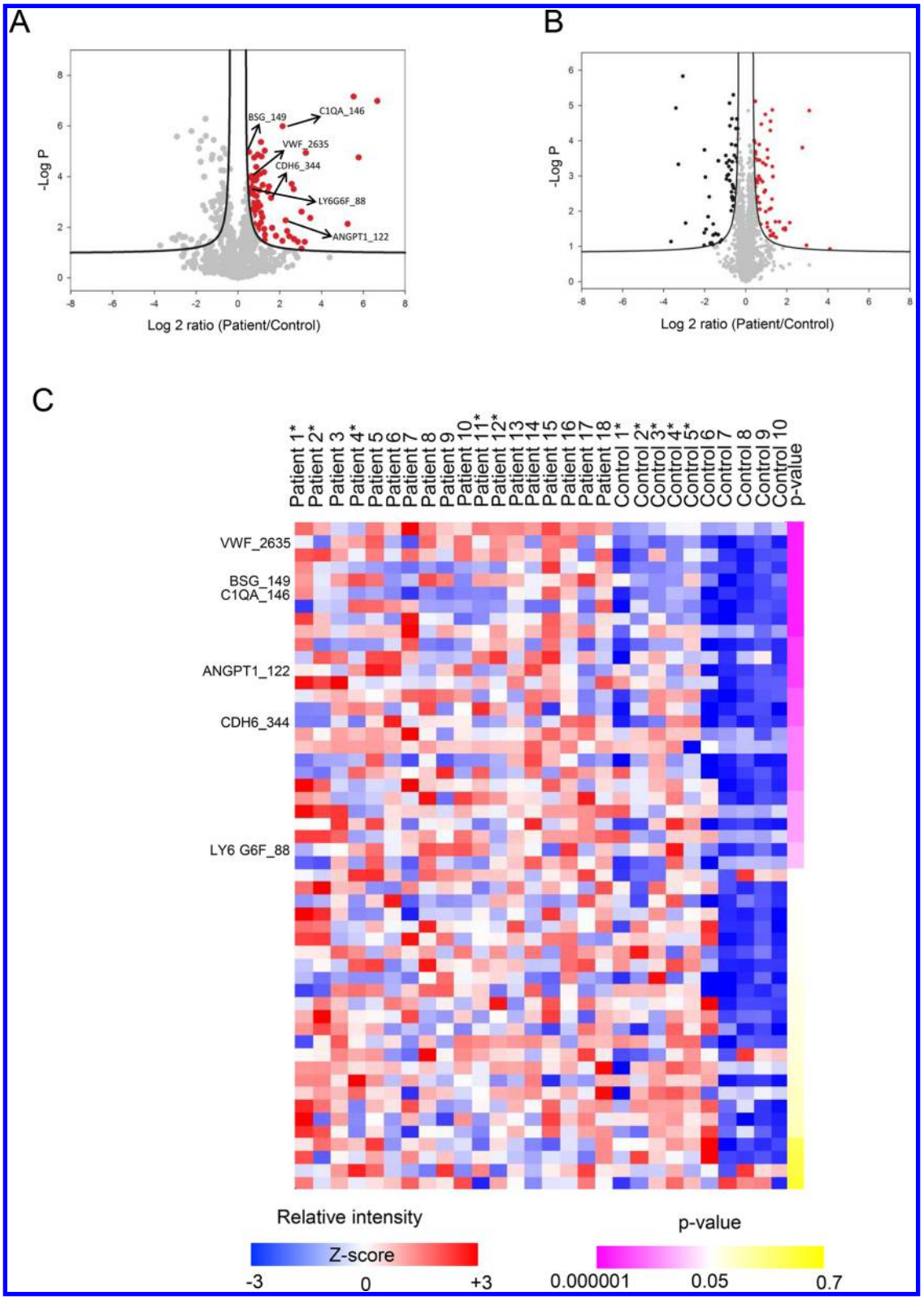


Figure 3. (A) Quantitative analysis of EV *N*-glycoproteomics between breast cancer and healthy controls. For global glycoproteomics, 18 breast cancer and 6 healthy controls were pooled to create a preliminary list of statistically increased glycosylated proteins. Volcano plot representing the quantitative analysis of the glycoproteomes of microvesicles in breast cancer patients vs healthy controls. Significant changes in proteins and glycosites in breast cancer were identified through a permutation-based FDR test (FDR = 0.05; S0 = 0.2) based on three technical replicates. The significant up-regulated proteins and glycosites are colored in red, and down-regulated are colored in gray on the left part of the volcano plot. (B) Quantitative analysis of EV proteins between breast cancer samples and healthy controls. (C) Quantitative glycoproteomics were performed on individuals to verify the preliminary list found in global glycoproteomics, and *p*-value represents the significance of comparing individual patients and controls. In total, 18 patients and 10 healthy controls were examined in MS-based verification experiment; 5 out of 18 patients and 5 out of 10 healthy controls were used in both global first individual verification glycoproteomics experiment (asterisk marked).

DOI: 10.1021/acs.analchem.8b01090
Anal. Chem. XXXX, XXX, XXX−XXX

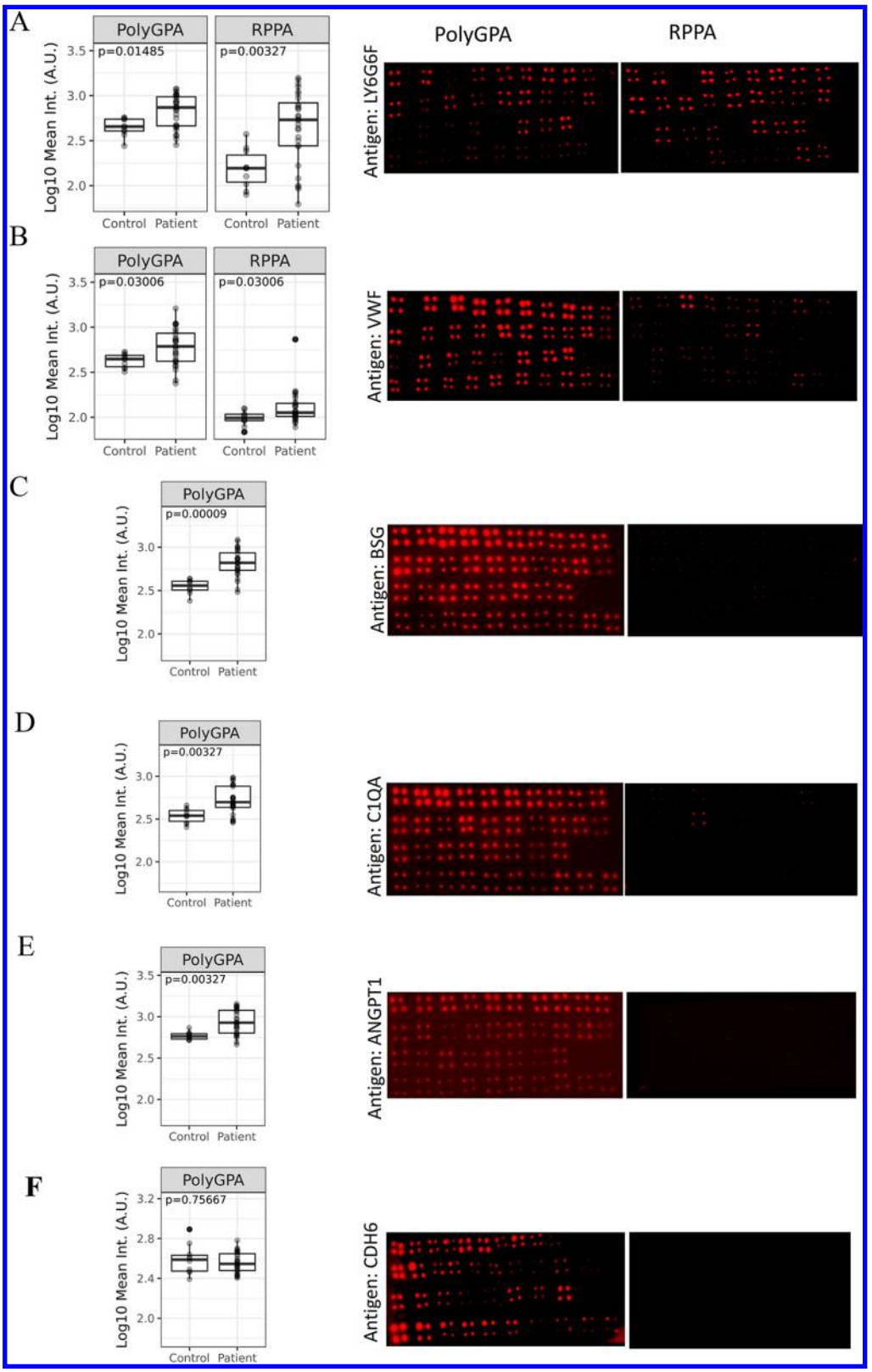


Figure 4. Verification of selected targets in plasma EVs by polyGPA. Quantification of endogenous (A) LY6G6F, (B) VWF, (C) BSG, (D) C1QA, (E) ANGPT1, and (F) CDH6 in plasma EVs. For each membrane, the top three rows were printed with 28 breast cancer samples (first two rows with 10 samples and the third row with 8 samples) and the fourth row with 10 healthy control samples, each with 4 prints per individual sample. For quantitation of signals in polyGPA, the mean intensity of 4 prints per individual was used and the distribution of log 10 (intensity) is depicted in the left pane.

hydroxyamino-dendrimer-modified nitrocellulose to covalently capture preoxidized glycans on glycoproteins, followed by on-membrane detection using the same validated antibodies as in typical reverse phase protein arrays. Although no glycosylation

DOI: 10.1021/acs.analchem.8b01090
Anal. Chem. XXXX, XXX, XXX−XXX

specific antibody or lectin is used, any change in polyGPA signal is attributed to the change in overall glycosylation of targeted glycoprotein or site differential glycosylation. In addition, we demonstrated that polyGPA's sensitivity is much higher than RPPA (over 10-fold signal increase) for the same protein concentration, likely due to improved orientation of glycoproteins during their glycan binding to the polyGPA membrane, exposing more epitopes for increased overall signal.

We prioritized the glycoproteins for further verification by polyGPA through their biological relevance to cancer in previous studies and availability of their antibodies which are validated by the Human Protein Atlas (HPA) project for high specificity. Among the glycoproteins that show significant increase in breast cancer patients (Figure 3C), some are known plasma/serum glycoproteins while others have never been detected from blood. Interestingly, 70% of the glycoproteins on the list have previously been identified from cancer tissues (Supporting Information Figure S3),[23] highlighting the important feature of this biomarker strategy which did not require an invasive biopsy but rather used EVs as the source to identify biomarkers previously reported in cancer tissue studies. We selected 6 EV glycoproteins, a membrane protein lymphocyte antigen 6 complex locus protein G6f (LY6G6F), a multimeric plasma glycoprotein von willebrand factor (VWF), CD147/basigin (BSG), complement C1q subcomponent subunit A (C1QA), angiopoietin-1 (ANGPT1/Ang1), and cadherin-6 (CDH6) for further verification with another cohort of plasma samples from 28 breast cancer patients and 10 healthy controls. The 5 validated glycoproteins all have been directly linked to or implicated with cancer according to previous studies.[24−28] EVs were isolated from plasma samples, lysed, and preoxidized, and each individual sample was printed onto the polyGPA membranes and unfunctionalized membranes as in regular RPPA. Specific protein antibodies were then used to detect and quantify endogenous LY6G6F, VWF, BSG, C1QA, ANGPT1, and CDH6 signals in individual samples. As shown in Figure 4, measurements by polyGPA showed much better sensitivity because of significantly reduced sample complexity after the enrichment of glycoproteins on the functionalized membrane and better orientation of glycoproteins for epitope detection by the antibodies. This enhanced sensitivity proved to be critical for the detection of proteins with much lower abundances, such as BSG, C1QA, ANGPT1, and CDH6, and their protein signals were barely detectable in RPPA (Figure 4C−F). Five out of six glycoproteins, except CDH6, showed statistically significant specificity ($p < 0.05$) for breast cancer. The quantitative measurements with polyGPA and RPPA also allowed us to identify whether glycosylation elevation is due to changes in protein expression or changes in glycosylation. There is significant elevation in both polyGPA and RPPA for LY6G6F. The increase in breast cancer patients was clearly observed in polyGPA for VWF, but the difference is small in RPPA (the distinction is largely due to one outlier; Figure 4B), indicating that the glycosylation elevation in cancer patients is likely due to changes in patient-specific glycosylation. As stated above, due to low abundance, BSG, C1QA, ANGPT1 and CDH6 could only be quantified by polyGPA, further highlighting its uniqueness and high sensitivity for clinical samples.

CONCLUSION

Development of new glycoproteins as potential biomarkers has struggled due to the lack of good analytical tools. Here, we reported an in-depth analysis of *N*-glycoproteomes in plasma EVs and demonstrated the feasibility of developing EV glycoproteins as potential breast cancer biomarkers.

This study also addresses a major issue in the development of glycoproteins for biomarker discovery, i.e., how to validate specific glycoproteins in high throughput. We introduced polyGPA as an alternative and novel high throughput method for simple, sensitive quantification of glycoproteins in array format. Using glyco-specific, 3-dimensional functionalized membrane to capture glycoproteins followed by detection using high-quality antibodies, the new platform allowed us to measure glycoproteins in multiple clinical samples in parallel. However, the limitation of polyGPA should be taken into consideration for clinical validation applications. PolyGPA only measures the overall glycosylation in a protein. For a glycoprotein with multiple glycosylation sites, polyGPA may not be sensitive enough to a glycosylation change on a specific site.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.8b01090.

Online methods description and figures showing EV isolation characterization, glycoproteins comparison, and hierarchical clustering analysis (PDF)

Tables listing identified glycopeptides, quantifiable glycosylation sites, significant increased glycosylation sites, quantifiable proteins, and polyGPA results (XLSX)

AUTHOR INFORMATION

Corresponding Authors

*(W.A.T.) E-mail: watao@purdue.edu.

*(Y.Z.) E-mail: ying@fudan.edu.cn.

ORCID

Ying Zhang: 0000-0003-0509-1098

W. Andy Tao: 0000-0002-5535-5517

Notes

The authors declare the following competing financial interest(s): I-H.C., H.A.A., A.B.I., and W.A.T. are entitled to royalties that may result from licensing related patent applications according to IP policies of Purdue Research Foundation.

ACKNOWLEDGMENTS

This project has been funded by NIH Grants 5R01GM088317, 1R01GM111788, and S10 RR025044 and NSF Grant 1506752. Additional support was provided by the Purdue University Center for Cancer Research (Grant P30 CA023168).

REFERENCES

(1) Yong, E. *Nature* **2014**, *511*, 524−526.

(2) Pinho, S. S.; Reis, C. A. *Nat. Rev. Cancer* **2015**, *15*, 540−555.

(3) Blixt, O.; Westerlind, U. *Curr. Opin. Chem. Biol.* **2014**, *18*, 62−69.

(4) Hanash, S. M.; Pitteri, S. J.; Faca, V. M. *Nature* **2008**, *452*, 571−579.

(5) Geyer, P. E.; Kulak, N. A.; Pichler, G.; Holdt, L. M.; Teupser, D.; Mann, M. *Cell Syst* **2016**, *2*, 185−195.

(6) Melo, S. A.; Luecke, L. B.; Kahlert, C.; Fernandez, A. F.; Gammon, S. T.; Kaye, J.; LeBleu, V. S.; Mittendorf, E. A.; Weitz, J.;

DOI: 10.1021/acs.analchem.8b01090
Anal. Chem. XXXX, XXX, XXX−XXX

Rahbari, N.; Reissfelder, C.; Pilarsky, C.; Fraga, M. F.; Piwnica-Worms, D.; Kalluri, R. *Nature* **2015**, *523*, 177−182.
(7) Gonzales, P. A.; Pisitkun, T.; Hoffert, J. D.; Tchapyjnikov, D.; Star, R. A.; Kleta, R.; Wang, N. S.; Knepper, M. A. *J. Am. Soc. Nephrol.* **2009**, *20*, 363−379.
(8) Boukouris, S.; Mathivanan, S. *Proteomics: Clin. Appl.* **2015**, *9*, 358−367.
(9) Moreno-Gonzalo, O.; Villarroya-Beltri, C.; Sanchez-Madrid, F. *Front. Immunol.* **2014**, *5*, 383.
(10) Chen, I. H.; Xue, L.; Hsu, C. C.; Paez, J. S.; Pan, L.; Andaluz, H.; Wendt, M. K.; Iliuk, A. B.; Zhu, J. K.; Tao, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 3175−3180.
(11) Gerlach, J. Q.; Griffin, M. D. *Mol. BioSyst.* **2016**, *12*, 1071−1081.
(12) Sok Hwee Cheow, E.; Hwan Sim, K.; de Kleijn, D.; Neng Lee, C.; Sorokin, V.; Sze, S. K. *Mol. Cell. Proteomics* **2015**, *14*, 1657−1671.
(13) Pan, L.; Aguilar, H. A.; Wang, L.; Iliuk, A.; Tao, W. A. *J. Am. Chem. Soc.* **2016**, *138*, 15311−15314.
(14) Ruhaak, L. R.; Miyamoto, S.; Lebrilla, C. B. *Mol. Cell. Proteomics* **2013**, *12*, 846−855.
(15) Zhang, Y.; Jiao, J.; Yang, P.; Lu, H. *Clin Proteomics* **2014**, *11*, 18.
(16) Xiao, H.; Wu, R. *Chem. Sci.* **2017**, *8*, 268−277.
(17) Krishnamoorthy, L.; Mahal, L. K. *ACS Chem. Biol.* **2009**, *4*, 715−732.
(18) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. *Nat. Biotechnol.* **2003**, *21*, 660−666.
(19) Masuda, T.; Sugiyama, N.; Tomita, M.; Ishihama, Y. *Anal. Chem.* **2011**, *83*, 7698−7703.
(20) Vizcaino, J. A.; Cote, R. G.; Csordas, A.; Dianes, J. A.; Fabregat, A.; Foster, J. M.; Griss, J.; Alpi, E.; Birim, M.; Contell, J.; O'Kelly, G.; Schoenegger, A.; Ovelleiro, D.; Perez-Riverol, Y.; Reisinger, F.; Rios, D.; Wang, R.; Hermjakob, H. *Nucleic Acids Res.* **2013**, *41*, D1063−D1069.
(21) Gelderman, M. P.; Simak, J. *Methods Mol. Biol.* **2008**, *484*, 79−93.
(22) Kowal, J.; Arras, G.; Colombo, M.; Jouve, M.; Morath, J. P.; Primdal-Bengtson, B.; Dingli, F.; Loew, D.; Tkach, M.; Thery, C. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E968−977.
(23) Hill, J. J.; Tremblay, T. L.; Fauteux, F.; Li, J.; Wang, E.; Aguilar-Mahecha, A.; Basik, M.; O'Connor-McCourt, M. *J. Proteome Res.* **2015**, *14*, 1376−1388.
(24) De Vet, E. C.; Aguado, B.; Campbell, R. D. *Biochem. J.* **2003**, *375*, 207−213.
(25) Franchini, M.; Frattini, F.; Crestani, S.; Bonfanti, C.; Lippi, G. *Thromb. Res.* **2013**, *131*, 290−292.
(26) Kanekura, T.; Chen, X. *J. Dermatol. Sci.* **2010**, *57*, 149−154.
(27) Bulla, R.; Tripodo, C.; Rami, D.; Ling, G. S.; Agostinis, C.; Guarnotta, C.; Zorzet, S.; Durigutto, P.; Botto, M.; Tedesco, F. *Nat. Commun.* **2016**, *7*, 10346.
(28) Metheny-Barlow, L. J.; Li, L. Y. *Cell Res.* **2003**, *13*, 309−317.

DOI: 10.1021/acs.analchem.8b01090
Anal. Chem. XXXX, XXX, XXX−XXX