NRC·CNRC

NRC Publications Archive
Archives des publications du CNRC

Glycoproteomic comparison of clinical triple-negative and luminal breast tumors

Hill, Jennifer J.; Tremblay, Tammy-Lynn; Fauteux, François; Li, Jie; Wang, Edwin; Aguilar-Mahecha, Adriana; Basik, Mark; O'Connor-McCourt, Maureen

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.
For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1021/pr500987r
Journal of Proteome Research, 14, 3, pp. 1376-1388, 2015-03-06

NRC Publications Record / Notice d'Archives des publications de CNRC:
https://nrc-publications.canada.ca/eng/view/object/?id=dbe6fa11-7ae3-46d2-93a5-21c3cf13f49f
https://publications-cnrc.canada.ca/fra/voir/objet/?id=dbe6fa11-7ae3-46d2-93a5-21c3cf13f49f

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada | Conseil national de recherches Canada

Canada

Journal of proteome research

Article

pubs.acs.org/jpr

Glycoproteomic Comparison of Clinical Triple-Negative and Luminal Breast Tumors

Jennifer J. Hill,*[,†] Tammy-Lynn Tremblay,[†] François Fauteux,[†] Jie Li,[‡] Edwin Wang,[‡] Adriana Aguilar-Mahecha,[§] Mark Basik,[§] and Maureen O'Connor-McCourt[‡]

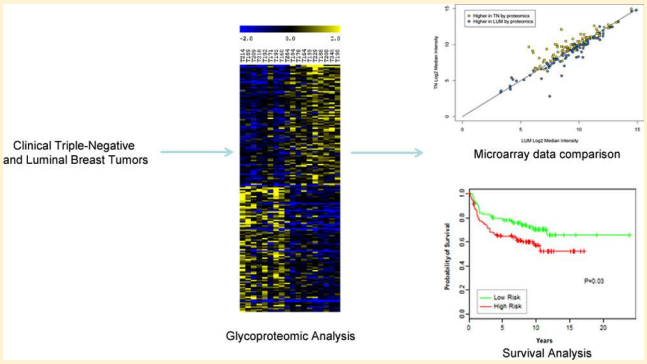
[†]Human Health Therapeutics, National Research Council Canada, 100 Sussex Drive, Ottawa, Ontario K1A 0R6, Canada

[‡]Human Health Therapeutics, National Research Council Canada, 6100 Royalmount Avenue, Montreal, Québec H4P 2R2, Canada

[§]Department of Oncology, Lady Davis Institute for Medical Research, 3755 Chemin de la Côte-Sainte-Catherine, Montreal, Québec H3T 1E2, Canada

Supporting Information

ABSTRACT: Triple-negative (TN) breast cancer accounts for ∼15% of breast cancers and is characterized by a high likelihood of relapse and a lack of targeted therapies. In contrast, luminal-type tumors that express the estrogen and progesterone receptors (ER+/PR+) and lack expression of human epidermal growth factor receptor 2 (Her2−) are treated with targeted hormonal therapy and carry a better prognosis. To identify potential targets for the development of future therapeutics aimed specifically at TN breast cancers, we have used a hydrazide-based glycoproteomic workflow to compare protein expression in clinical tumors from nine TN (Her2−/ER-/PR-) and nine luminal (Her2−/ER+/PR+) patients. Using a label-free LC−MS based approach, we identified and quantified 2264 proteins. Of these, 90 proteins were more highly expressed and 86 proteins were underexpressed in the TN tumors relative to the luminal tumors. The expression level of four of these potential targets was verified in the original set of tumors by Western blot and correlated well with our mass-spectrometry-based quantification. Furthermore, 30% of the proteins differentially expressed between luminal and TN tumors were validated in a larger cohort of 406 TN and 469 luminal tumors through corresponding differences in their mRNA expression in publically available microarray data. A group of 29 of these differentially expressed proteins was shown to correctly classify 88% of TN and luminal tumors using microarray data of their associated mRNA levels. Interestingly, even within a group of TN breast cancer patients, the expression levels of these same mRNAs were able to significantly predict patient survival, suggesting that these proteins play a role in the aggressiveness seen in TN tumors. This study provides a comprehensive list of potential targets for the development of diagnostic and therapeutic agents specifically aimed at treating TN breast cancer and demonstrates the utility of using publicly available microarray data to further prioritize potential targets.



KEYWORDS: *triple-negative breast cancer, luminal breast cancer, proteomics, glycoproteomics, mass spectrometry, tumor*

INTRODUCTION

Breast cancers can be classified in therapeutically and prognostically relevant categories by immunophenotype. Triple-negative (TN) breast cancers are defined as tumors that lack estrogen receptor (ER) and progesterone receptor (PR) expression as well as epidermal growth factor receptor 2 (HER2) amplification. These cancers account for ∼15% of all breast cancers and are more common in young women from Hispanic and African-American background.[1] TN tumors are usually high-grade and large in size and have lymph-node involvement at the time of diagnosis. Women with TN tumors have a high risk of early relapse and an increased likelihood of distant recurrence and death compared with women with other breast cancers.[2] In contrast, women with luminal tumors expressing hormonal receptors are treated with targeted hormonal therapy and generally have a good prognosis.[3] Luminal breast cancers are the most common subtype of breast cancer, accounting for ∼70% of all breast tumors.

Presently, there are no targeted therapies specifically available for TN breast cancer patients. Instead, these patients rely on conventional chemotherapeutic regimens. Although the response to chemotherapy is better in TN breast cancer than in ER+-type breast cancers, prognosis remains poor.[4] Indeed, the majority of TN tumors will develop resistance to these cytotoxic therapies and relapse. Thus, there is a particularly urgent need to develop novel targets for therapy in this subset of breast cancer.

Received: September 18, 2014
Published: February 6, 2015

ACS Publications © 2015 American Chemical Society

DOI: 10.1021/pr500987r
J. Proteome Res. 2015, 14, 1376−1388

To date, a limited number of studies have explored protein expression differences in clinical samples from triple-negative and nontriple-negative breast tumors. Two studies compared total protein expression in Her2-positive tumors with TN tumors by 2D-DIGE and spectral counting, respectively.[5,6] Recently, another 2D gel-based study identified Mage-A4 as overexpressed in TN tumors[7] compared with other breast tumor types. Lu et al. compared TN breast tumors with Her2−/ER+/PR+ tumors using hydrophobic enrichment and quantification by spectral counting,[8] leading to the identification of differentially expressed proteins including keratins, HLA-class proteins, and heat-shock proteins.

Extracellular and plasma membrane proteins make particularly promising targets for the development of therapeutic and diagnostic agents due to their inherent accessibility; however, this subset of proteins are typically difficult to identify and quantify in proteomic experiments due to their lower expression level and relative insolubility. Because glycosylation typically occurs on membrane and extracellular proteins, glycoprotein enrichment can increase coverage of this subset of proteins. Glycosylated proteins and peptides can be enriched by many different methods, including lectin affinity, HILIC and normal phase liquid chromatography, titanium dioxide, and capture on hydrazide solid support.[9−14]

In this study, we chose to use hydrazide-based capture of glycoproteins due to its high specificity.[15,16] Hydrazide capture is based on the covalent binding of oxidized glycans onto a solid support with immobilized hydrazide functional groups and has been used to profile diverse cancer-relevant samples, including serum, pleural effusion fluid, and cell lines.[17−22] A limited number of proteomic studies have used hydrazide capture to enrich glycoproteins or glycopeptides from human clinical tissue. These include studies that profiled glycopeptide expression in breast tumors and adjacent normal tissue,[23] OCT-embedded clinical prostate tumor tissue,[24] and different histological subtypes of ovarian tumors.[25]

In this study, we have compared the expression of glycoprotein-enriched fractions isolated from TN and luminal clinical breast cancer tumors with the aim of identifying potential therapeutic and diagnostic markers targeted specifically at triple-negative breast cancer.

PROCEDURES

Clinical Samples

Breast cancer tumors were obtained from the FRQS Réseau Recherche Cancer breast tumor biobank at the Jewish General Hospital (Montreal, Canada). Approval from the local ethics board was obtained for the current study, and written informed consent was obtained from all tumor donors. Primary tumors from surgical excisions were collected from the Pathology Department of the Jewish General Hospital and were snap-frozen in liquid nitrogen; ∼6 mg of frozen tumor was used for hydrazide capture from each of nine TN breast tumors (ER−, PR−, and Her2−) and nine luminal tumors (ER+, PR+, and Her2−). Receptor status was assessed by immunohistochemical analysis by a board-certified pathologist using local standards. Researchers performing the hydrazide capture and mass-spectrometry analyses were blinded to tumor type until all mass-spectrometry data were collected and quantified.

Hydrazide Capture from Clinical Breast Tumors

Breast tissue was thawed, cut into 2 mm pieces, and homogenized in coupling buffer (CB: 100 mM NaOAc, 150 mM NaCl, pH 5.5). SDS was added to a final concentration of 0.5% and samples were agitated for 1 h at RT and centrifuged at 16 000*g* for 10 min at RT to remove insoluble debris, and the protein concentration of the supernatant was determined using a DC protein assay (Bio-Rad).

The protocol for hydrazide capture was based on previously published methods[15,16] with some modifications. In brief, 1 mL of protein extract containing 500 μg of protein was oxidized using 15 mM Na-m-periodate for 1 h at RT in the dark. This reaction was quenched with 30 mM Na_2SO_4 for 10 min at RT, and protease inhibitors were added (Sigma mammalian protease inhibitor cocktail, 1:200 v/v). Samples were incubated O/N at RT with 50 μL of prewashed hydrazide beads (Bio-Rad). Beads were washed eight times with urea buffer (8 M urea, 0.4 M NH_4HCO_3) and four times with 50 mM NH_4HCO_3, followed by reduction (10 mM DTT for 1 h at 56 °C) and alkylation (25 mM iodoacetamide for 1 h at RT in the dark). Beads were then washed four times with 50 mM NH_4HCO_3/15% acetonitrile before incubation O/N at 37 °C with 10 μg of trypsin in 450 μL of 50 mM NH_4HCO_3/15% acetonitrile. Trypsin-released peptides were collected and beads were washed three times with 1.5 M NaCl, three times with 60% aceonitrile-0.1% TFA, three times with 100% MeOH, and six times with 50 mM NH_4HCO_3 before an incubation O/N at 37 °C with 4 units of PNGase F (Roche) in 100 μL of 50 mM NH_4HCO_3. N-linked peptides were collected by combining the supernatant with further elutions with 200 μL of 50 mM NH_4HCO_3 and 200 μL of 50% acetonitrile/5% acetic acid. All eluates were pooled, dried under vacuum, and resuspended in 40 μL of deionized water.

Mass Spectrometry Analysis

Samples were analyzed by automated nanoLC−MS(/MS) on an LTQ-Orbitrap XL (Thermo Scientific) coupled to a NanoAcquity UPLC system (Waters). Peptides were trapped using an inline C8 precolumn (LC Packings, 161194) and C18 trap column (Waters, 186003514) and separated on a 10 cm × 100 μm I.D. C18 column (Waters, 1.7 μm BEH130C18, 186003546) at ∼250 nL/min using the following gradient: 1−35% solution B (100% ACN/0.1% formic acid) over 58 min and 35−60% B over 2 min, followed by a 9 min equilibration at 1% B. To minimize carryover effects, we ran blanks between samples using a 25 min gradient. MS spectra were acquired in the Orbitrap between 400 and 2000 Da m/z in profile mode at 60k resolution, while data-dependent turbo CID MS/MS scans of the top three ions were acquired concurrently in the ion trap in centroid mode with dynamic exclusion (180 s) using the following settings: isolation width = 3.0, activation Q = 0.250, activation time = 30 ms, and normalized collision energy = 35.0. These acquisition parameters provided a cycle time of 1.3 s to allow >10 points to be collected across a chromatographic peak, and the high-resolution MS scan allowed for accurate centroiding of even small peaks to increase quantitative accuracy. Three technical replicates were performed by injecting the entire set of samples on three separate occasions. The volume of tryptic digest loaded in the second and third replicates was adjusted to account for differences in the median intensity of identified proteins quantified in all samples in the first technical replicate.

Quantification by Label-Free LC−MS

Identified peptides were quantified using in-house software based on MatchRx.[26] In brief, ion current was extracted from within 12 ppm of each identified peptide present within 40 s of

DOI: 10.1021/pr500987r
J. Proteome Res. 2015, 14, 1376−1388

the retention time predicted after multirun alignment. Data were normalized separately for the N-linked and trypsin-released fraction by ensuring that the median intensity of proteins that were identified within all runs was equal. The normalized peptide level data from each fraction was then combined, and peptides were assigned to the protein that had the most other unique peptides identified. If two or more proteins match all of the assigned peptides, then these accession numbers are listed in the Table as an alternate identification. Protein intensities were calculated by summing the intensities of all unique peptide sequences derived from that protein. When calculating the median fold change between triple-negative and luminal samples, proteins found only in TN (i.e., median intensity of luminal tumors = 0) or only in luminal tumors were assigned a fold change of 8 ($\log_2$ ratio of 3) or −8 ($\log_2$ ratio of −3), respectively.

Database Searches for Peptide Identifications

All MS/MS spectra were assigned to peptide sequences using Mascot v2.3. Orbitrap data were converted to mzXML using Msconvert from the ProteoWizard version 2.0.1905 package[27] with the following parameters: −mzXML −32 −filter 'peakPicking true [2,3]'. MGF files were generated from the mzXML file using MzXML2Search from the Trans Proteomics Pipeline project (TPP-4.3.1, http://tools.proteomecenter.org/wiki/index.php?title=Software:TPP) and searched against the *Homo sapiens* SwissProt database (downloaded 09/2010, 20 359 sequences; parameters: enzyme = trypsin; modifications = carbamidomethyl (C, fixed), oxidation (M, variable), deamidation (N, variable) for N-linked fraction only; peptide tolerance = 0.5 Da (N-linked fraction), 1.2 Da (trypsin-released fraction); fragment tolerance = 1.2 Da; 1 missed cleavage allowed). Following systemic correction of masses based on the median ppm of high scoring peptides, search results were filtered to remove peptides from the trypsin-released fraction that had a delta mass >8 ppm or a score <40 or from the N-linked fraction that lacked a consensus N-glycosylation motif (N-X-[STC]), had a delta mass >8 ppm, or had a score <35. As previously described,[28] we found that these search conditions led to a lower false identification rate than setting a tighter mass tolerance during the search. The smaller 0.5 Da mass tolerance for the N-linked fraction search was required to ensure that the deamidation modification of 0.98 Da was correctly assigned. Under these conditions, the false identification rate based on decoy database searches was 1.2%.[29] All proteins identified by a single peptide were further filtered to require peptide scores of >42.5 (trypsin-released) or >37.5 (N-linked) to bring the false-positive identification rate down to <1%.

Statistics for Glycoproteomic Data

To identify differentially expressed proteins, p values were calculated for each of the three technical repeats for each protein using a delta mean permutation test. In brief, the absolute difference between the average intensity of the nine TN and nine luminal tumors was calculated. The intensities belonging to the 18 samples were then shuffled and randomly assigned to two groups and the absolute difference in average intensity of these randomized groups was calculated. This was repeated for 5000 permutations, and the p value was determined based on number of times that the randomly generated intensity difference was greater than the intensity difference between the TN and luminal groups. A combined p value for the three technical replicates was then calculated based on a binomial distribution. This combined p value was used to identify differentially expressed proteins that showed consistent directional changes (TN vs LUM) in all three technical repeats at a false discovery rate of 5% using the Benjamini−Hochberg procedure.[30]

Western Blots

Total tumor lysates were prepared and quantified as previously described. Equal protein amounts from each tumor were run on an SDS-PAGE gel and transferred to nitrocellulose membranes using standard methods. Membranes were blotted by standard methods using 5% milk in TBS as a blocking solution and HRP-conjugated secondary antibodies, followed by chemiluminescence detection using the ECL Plus Western Blotting Detection System (GE Healthcare). The following primary antibodies and dilutions were used: anti-ASAH1 (Sigma no. HPA005468, 1:1000); antihuman desmocollin-2 (R&D Systems no. AF4688, 1:1000); antihuman CRLF1 (Abcam no. ab56500, 1:500); and antihuman azurocidin (R&D Systems no. AF2200, 1:1000).

Bioinformatics-Microarray Expression in TN and Luminal Tumors

Data for 406 TN and 469 luminal breast tumors was retrieved from the Gene Expression Omnibus[31] using in-house software and the GEOmetadb database.[32] Samples with annotated ER, PR, and HER2 status were selected from the following series: GSE10281, GSE10780, GSE10810, GSE11001, GSE12276, GSE12763, GSE12917, GSE13787, GSE16391, GSE16446, GSE17700, GSE17907, GSE18728, GSE18864, GSE19615, GSE19697, GSE20685, GSE20711, GSE2109, GSE21653, GSE22035, GSE22513, GSE22544, GSE23177, GSE23593, GSE23720, GSE26639, GSE26910, GSE28796, GSE28821, GSE30010, GSE31448, GSE3526, GSE3744, GSE5460, GSE5764, GSE6532, GSE7307, GSE7515, GSE7904, GSE8977, GSE9086, and GSE9195. Microarray data analysis was performed using the R statistical software and Bioconductor.[33] In brief, data were preprocessed using the MAS 5.0 algorithm[34] and summarized using BrainArray (version 16) Entrez Gene custom chip definition file (CDF).[35] Statistically significant differences between means were identified by shrinkage t test.[36] Protein to gene mapping was done using HUGO Gene Nomenclature Committee (HGNC) curated online repository.[37]

Bioinformatics −Protein Set Analysis

Gene expression profiles and patient survival data for breast cancer data were obtained from the NCBI GEO database: GSE3494, GSE11121, GSE1456, GSE9195, GSE6532, GSE7378, and GSE12093. A normalized metadata set was generated by combining these data sets as previously described.[38] In brief, Affymetrix expression data were analyzed using the MAS5.0 algorithm. Subsequently, data were log2-transformed, and the values of the probes belonging to the same gene were averaged and then median-centered across arrays. Finally, gene expression data were z-score-normalized across arrays. TN samples were identified based on the gene expression values of ER, PR, and HER2. Cutoffs for ER, PR, and HER2 expression from microarray were derived from fitting two normal distributions to the observed distribution of Affymetrix expression values by maximum likelihood optimization using the *optim* function in R.[39] 224 TN samples were identified and extracted for survival analysis using genes encoding the differentially expressed TN/LUM proteins from Tables 2 and 3. To do so, we performed a Kaplan−Meier

DOI: 10.1021/pr500987r
J. Proteome Res. 2015, 14, 1376−1388

analysis by implementing the Cox−Mantel log-rank test using R, a statistical computing language (http://www.r-project.org/), as described previously.[38,40]

RESULTS

Proteins Identified in Triple Negative and Luminal Tumors

Breast tumors were acquired from 18 patients through the FRSQ RRCancer breast tumor biobank. As can be seen from the clinical characteristics of these patients, detailed in Table 1,

Table 1. Clinical Information for TN and Luminal Tumors

tumor no.	age	histological grade	pTNM	ER	PR	HER2
T-214	27	III	$T_2N_0M_0$	−	−	−
T-189	43	III	$T_1N_2M_x$	−	−	−
T-192	49	III	$T_3N_1M_1$	−	−	−
T-160	50	III	$T_2N_0M_0$	−	−	−
T-195	50	III	$T_1N_1M_1$	−	−	−
T-209	62	III	$T_2N_xM_0$	−	−	−
T-171	75	III	$T_3N_0M_0$	−	−	−
T-264	75	III	T_3N_1	−	−	−
T-218	84	II	$T_1N_xM_0$	−	−	−
T-164	26	III	$T_2N_1M_0$	+	+	−
T-190	44	II	$T_2N_0M_x$	+	+	−
T-159	50	I	$T_2N_0M_0$	+	+	−
T-245	50	II	T_1N_0	+	+	−
T-229	58	III	$T_2N_2M_1$	+	+	−
T-184	62	II	$T_1N_0M_0$	+	+	−
T-178	74	III	T_2N_0	+	+	−
T-188	76	II	$T_1N_1M_0$	+	+	−
T-208	83	II	$T_1N_xM_0$	+	+	−

nine of these tumors were identified as triple-negative (TN) breast tumors lacking ER, PR, and HER2 expression and nine tumors were identified as Her2−/ER+/PR+, thus sharing the luminal immunophenotype of breast tumors. The characteristics of the TN breast cancer cases are relatively homogeneous, as most cases are histological grade III, which is typical of TN breast cancers. The wider range of histological grade scores in the ER+ cancers reflects their less aggressive characteristics.

For this study, we utilized a glycoproteomic approach to identify proteins in the TN and luminal tumors. Glycoproteins were enriched from ∼6 mg of tissue through the covalent binding of oxidized glycans to a solid hydrazide support, resulting in two fractions for each sample, as summarized in Figure 1A. The "trypsin-released" fraction contains non-glycopeptides that were released from the immobilized proteins by trypsin. The "N-linked" fraction contains peptides modified by an N-linked glycan that were released by PNGase F, resulting in the removal of the glycan and deamidation of the glycan-modified asparagine. To maintain information on tumor heterogeneity, we chose not to pool individual samples, analyzing the two fractions from each individual tumor in triplicate by nanoLC−MS(/MS) on an LTQ-Orbitrap XL. Representative LC−MS data for both the trypsin-released and N-linked fractions are shown in Figure 1B. As expected, we found that the N-linked fractions were considerably less complicated and more highly specific for glycoproteins than the trypsin-released fractions. The majority (88%) of peptides present in the N-linked fraction contained an N-linked glycosylation site, as evidenced by the presence of a deamidated asparagine within an N-linked glycosylation sequon (N-X-S/T/C, where X is any amino acid except P). In contrast, the trypsin-released fractions contained nonglycosylated peptides derived from both N- and O-linked glycoproteins as well as peptides derived from proteins that bind to the hydrazide bead through moieties other than traditional glycosylation, including GPI anchors, glycation, and aldehyde groups. We also see evidence of nonspecific binding of highly abundant proteins in the trypsin-released fraction.

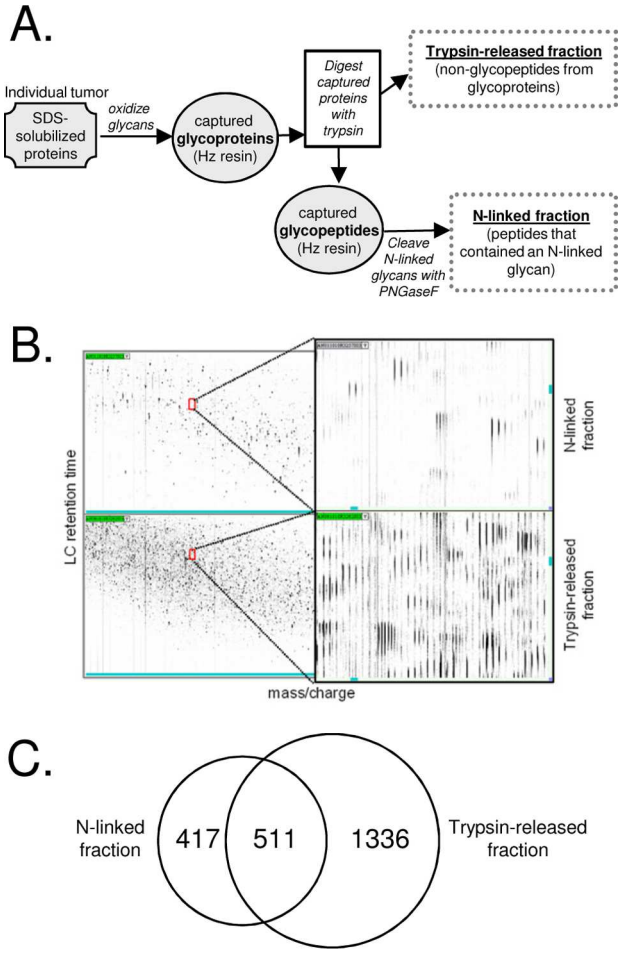


Figure 1. Overview of experimental design and mass spectrometry results. (A) Workflow used for the analysis of glycoproteins from clinical TN and luminal (Her2−/ER+/PR+) tumors. (B) 2D view of representative LC−MS data from tumor T-189 generated using MSight.[54] (C) Venn diagram showing the number of proteins identified in the N-linked and trypsin-released fractions. In total, 2263 proteins were identified in this study.

Proteins in each fraction were identified by database searching of low-resolution MS/MS data collected simultaneously to the high-resolution MS data. The N-linked and trypsin-released fractions provided greatly complementary data, with 18% of the identified proteins found only in the N-linked fraction and 59% of the proteins found only in the trypsin-released fraction (Figure 1C). A total of 2264 unique proteins were identified by combining the data for both fractions. Of these, 1477 proteins were identified by a minimum of 2 unique peptides, while the remainder were identified by only a single peptide. However, because of stringent filtering of results based on mass accuracy of the parent ion and the presence of a canonical N-linked glycosylation site (N-X-T/S/C) in the N-linked fraction, the false identification rate was estimated to be

DOI: 10.1021/pr500987r
J. Proteome Res. 2015, 14, 1376−1388

Table 2. Proteins More Highly Expressed in Triple-Negative Tumors As Identified by Glycoproteomics

acc [alternative acc's]	symbol	protein[a]	no. unique peptides	max Mascot score	fraction	ratio (median TN/median LUM) TR1	TR2	TR3	FDR (%)
P05067	APP	amyloid beta A4 protein	5	70.76	NT	1.75	1.77	1.58	0.01
P20160	AZU1	azurocidin**	10	112.55	NT	5.89	9.10	9.64	0.01
P19022	CDH2	cadherin-2**	2	60.85	N	2.80	3.46	6.25	0.01
Q9UN76	SLC6A14	sodium- and chloride-dependent neutral and basic amino acid transporter B(0+)**	2	110.65	N	7.22	9.41	10.80	0.01
P78310	CXADR	coxsackievirus and adenovirus receptor**	1	141.8	N	2.61	2.32	2.35	0.02
P08246	ELANE	neutrophil elastase**	7	114.54	NT	3.38	3.02	6.56	0.02
Q9C0H2	TTYH3	protein tweety homologue 3	2	74.67	N	1.80	1.63	2.19	0.03
P11166	SLC2A1	solute carrier family 2, facilitated glucose transporter member 1**	6	103.82	NT	2.27	2.39	2.35	0.03
P05164	MPO	myeloperoxidase	33	143.27	NT	1.92	2.16	1.81	0.04
P08779	KRT16	keratin, type I cytoskeletal 16	7	107.02	T	1.53	1.39	1.35	0.07
Q13885	TUBB2A	tubulin beta-2A chain	4	67.8	T	1.41	2.29	1.63	0.09
Q02487	DSC2	desmocollin-2	7	85.66	NT	3.26	1.17	1.79	0.11
Q9Y5X9	LIPG	endothelial lipase**	1	65.65	N	8.00	8.00		0.14
P20851	C4BPB	C4b-binding protein beta chain	4	115.11	NT	1.69	1.40	1.73	0.14
P11169 [Q8TDB8]	SLC2A3	solute carrier family 2, facilitated glucose transporter member 3	1	47.7	T	1.95	1.39	1.35	0.18
P13497	BMP1	bone morphogenetic protein 1	1	87.77	N	1.95	1.57	2.06	0.30
Q9BY67	CADM1	cell adhesion molecule 1	10	131.11	NT	1.76	1.84	2.00	0.30
O43405	COCH	cochlin**	1	73.68	NT	8.00	8.00	8.00	0.34
P24158	PRTN3	myeloblastin**	4	62.35	T	3.41	5.64	3.14	0.38
O43175	PHGDH	D-3-phosphoglycerate dehydrogenase	5	55.78	T	2.00	2.10	1.90	0.38
Q92820	GGH	gamma-glutamyl hydrolase	7	100.06	NT	1.70	1.76	1.96	0.38
P25205	MCM3	DNA replication licensing factor MCM3**	2	88.42	T	8.00	8.00	8.00	0.49
P16144	ITGB4	integrin beta-4	4	88.38	NT	2.39	1.87	2.46	0.50
P04083	ANXA1	annexin A1	12	128.65	T	1.49	1.26	1.36	0.53
P27701	CD82	CD82 antigen	2	80.48	N	1.66	1.70	1.98	0.53
Q29974 [P01911]	HLA-DRB1	HLA class II histocompatibility antigen, DRB1-16 beta chain	1	61.22	N	1.70	1.50	1.21	0.76
P22223	CDH3	cadherin-3**	1	58.57	NT	8.00	4.89	4.40	0.85
O60487	MPZL2	myelin protein zero-like protein 2**	2	88.8	N	3.60	2.61	1.30	1.04
P26368	U2AF2	splicing factor U2AF 65 kDa subunit**	3	110.89	T	2.88	1.78	4.26	1.04
P54709	ATP1B3	sodium/potassium-transporting ATPase subunit beta-3	3	93.67	NT	1.62	1.24	1.70	1.06
O43451	MGAM	maltase-glucoamylase, intestinal	1	108.05	N	8.00	8.00	8.00	1.08
P43490	NAMPT	nicotinamide phosphoribosyltransferase**	4	148.66	T	3.37	2.19		1.18
P47914	RPL29	60S ribosomal protein L29	1	71.92	T	1.60	1.58	1.57	1.42
P49746	THBS3	thrombospondin-3	1	84.15	N	1.52	1.57	1.84	1.47
P09651	HNRNPA1	heterogeneous nuclear ribonucleoprotein A1**	5	135.01	T	2.15	2.51	2.19	1.56
P00739	HPR	haptoglobin-related protein	3	143.01	T	1.67	1.77	1.55	1.64
Q9ULV4	CORO1C	coronin-1C	4	105.94	T	1.29	1.54	1.43	1.65
O95260	ATE1	arginyl-tRNA−protein transferase 1	9	107.67	T	1.53	1.22	2.18	1.72
Q08431	MFGE8	lactadherin	12	109.86	NT	1.49	1.09	1.18	1.73
P59665 [P59666]	DEFA1	neutrophil defensin 1	2	58.63	T	1.60	1.65	1.82	1.77
P48740	MASP1	mannan-binding lectin serine protease 1**	1	77.89	N	1.97	1.53	8.00	1.79
Q92974	ARHGEF2	rho guanine nucleotide exchange factor 2**	1	82.37	T	1.45	1.17	8.00	1.79
P19320	VCAM1	vascular cell adhesion protein 1	2	61.05	N	1.31	1.47	1.31	1.91
P02675	FGB	fibrinogen beta chain	29	132.3	NT	1.44	1.57	1.50	1.94
P43308	SSR2	translocon-associated protein subunit beta	2	160.93	N	1.43	1.21	1.12	2.24
Q8NFZ8	CADM4	cell adhesion molecule 4	1	78.11	N	1.78	1.68	1.80	2.37
Q30154	HLA-DRB5	HLA class II histocompatibility antigen, DR beta 5 chain	1	58.52	N	8.00	8.00		2.40
Q9Y487	ATP6 V0A2	V-type proton ATPase 116 kDa subunit a isoform 2	2	59.37	NT	1.55	1.75	2.08	2.49
Q9BXX0	EMILIN2	EMILIN-2	2	73.09	N	1.61	1.55	1.63	2.61
O00602	FCN1	ficolin-1**	2	114.75	T	2.19	1.78	3.02	2.63
Q96HE7	ERO1L	ERO1-like protein alpha	17	100.85	NT	2.07	1.85	1.39	2.75
Q9UIG0	BAZ1B	tyrosine-protein kinase BAZ1B	2	51.55	T	1.89	1.90	2.95	2.81
P06733	ENO1	alpha-enolase	19	126.4	T	1.39	1.15	1.27	3.00

DOI: 10.1021/pr500987r
J. Proteome Res. 2015, 14, 1376−1388

Table 2. continued

acc [alternative acc's]	symbol	protein[a]	no. unique peptides	max Mascot score	fraction	ratio (median TN/median LUM) TR1	TR2	TR3	FDR (%)
O00567	NOP56	nucleolar protein 56	6	126.47	T	1.49	1.37	1.67	3.04
Q02809	PLOD1	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1	19	128.15	NT	1.32	1.44	1.42	3.05
O95297	MPZL1	myelin protein zero-like protein 1	2	68.67	N	1.28	1.35	1.23	3.14
Q9Y295	DRG1	developmentally regulated GTP-binding protein 1	3	58.61	T	1.45	1.46	1.53	3.15
P09874	PARP1	poly [ADP-ribose] polymerase 1	11	120.4	T	1.46	1.36	1.60	3.15
O43852	CALU	calumenin	10	129.08	NT	1.43	2.03	1.58	3.17
Q8WWB7	C1ORF85	lysosomal protein NCU-G1	1	53.95	N	1.48	1.54	1.46	3.29
P30825	SLC7A1	high-affinity cationic amino acid transporter 1**	1	109.13	N	2.36	2.39	2.21	3.29
Q9UPN9	TRIM33	E3 ubiquitin-protein ligase TRIM33	1	56.74	T	1.15	2.27	1.35	3.50
P17936	IGFBP3	insulin-like growth factor-binding protein 3	2	76.55	N	1.99	1.79	1.28	3.55
P14780	MMP9	matrix metalloproteinase-9	18	101.35	NT	1.29	1.40	1.16	3.55
P80188	LCN2	neutrophil gelatinase-associated lipocalin	10	85.43	NT	1.78	1.39	1.13	3.71
A2VDJ0	KIAA0922	transmembrane protein 131-like	1	41.46	N	2.90	1.27		3.72
P04406	GAPDH	glyceraldehyde-3-phosphate dehydrogenase	15	156.93	T	1.68	1.54	1.36	3.73
P02671	FGA	fibrinogen alpha chain	25	157.26	T	1.39	1.44	1.16	3.75
P26641	EEF1G	elongation factor 1-gamma	6	102.33	T	1.31	1.27	1.31	3.85
P19440	GGT1	gamma-glutamyltranspeptidase 1	4	111.12	N	1.47	1.42	1.65	3.94
P20292	ALOX5AP	arachidonate 5-lipoxygenase-activating protein**	1	54.72	T	3.24	3.54	1.40	3.95
Q9BUP0	EFHD1	EF-hand domain-containing protein D1	2	81.28	T	1.33	1.30	1.81	4.11
P08572	COL4A2	collagen alpha-2(IV) chain	4	91.6	T	1.37	1.89	1.24	4.23
P02533	KRT14	keratin, type I cytoskeletal 14	12	135.85	T	1.28	1.33	1.28	4.39
P27797	CALR	calreticulin	9	105.22	NT	1.58	1.41	1.71	4.39
O43707	ACTN4	alpha-actinin-4	36	107.95	T	1.34	1.37	1.22	4.40
Q9Y624	F11R	junctional adhesion molecule A	5	123.26	NT	1.31	1.31	1.31	4.47
P10586	PTPRF	receptor-type tyrosine-protein phosphatase F	6	97.27	NT	1.28	1.23	1.21	4.50
O96005	CLPTM1	cleft lip and palate transmembrane protein 1	4	111.38	NT	1.46	1.49	1.80	4.51
P07996	THBS1	thrombospondin-1**	47	160.22	NT	3.18	3.04	3.48	4.52
Q96P70	IPO9	importin-9	1	71.83	T	2.04	1.79		4.53
P08581	MET	hepatocyte growth factor receptor**	1	65.78	N	8.00	8.00		4.59
O75462	CRLF1	cytokine receptor-like factor 1**	1	56.8	N	8.00	8.00	8.00	4.74
Q9UBE0	SAE1	SUMO-activating enzyme subunit 1**	1	43.18	T	1.31	1.57	8.00	4.75
P07203	GPX1	glutathione peroxidase 1	1	60.57	T	1.18	2.08	1.29	4.79
Q9NR97	TLR8	Toll-like receptor 8	5	70.59	N	1.49	1.75	1.68	4.90
Q9BU40	CHRDL1	chordin-like protein 1**	1	64.66	N	3.17	2.46	1.66	4.91
P52272	HNRNPM	heterogeneous nuclear ribonucleoprotein M	7	76.57	T	1.62	1.41	1.37	4.91
P15924	DSP	desmoplakin	7	94.08	T	2.90	1.14	1.76	4.98
P07998	RNASE1	ribonuclease pancreatic	3	80.54	NT	1.51	1.55	1.70	5.00

[a]Proteins that were validated by microarray ($p \leq 0.05$) are underlined, and proteins marked by a double asterisk (**) are members of the 29-protein subset.

<1% using a decoy database search strategy, even for the single peptide hits. The details of the peptides identified in this experiment, including their protein assignment, sequence, charge state, modifications, Mascot scores, delta mass values, and retention times are provided in Supplementary Table 1 in the Supporting Information.

Identification of Proteins Differentially Expressed between TN and Luminal Tumors

Identified peptides were quantified based on the extracted MS signal using in-house software. Duplicate samples from the four tumor lysates for which we had the most starting material were prepared and analyzed in parallel for the entire sample preparation procedure, including the multiday hydrazide capture procedure, to estimate the quantitative variability in our sample preparation and label-free MS quantification workflows. The intensity values from these true technical replicates had a median relative standard deviation of 13.7% with an interquartile range of 5.8 to 30.2%, suggesting that our sample preparation and analysis methods did not introduce large sources of quantitative variability (data not shown).

To determine proteins that were differentially expressed between TN and luminal tumors, we performed a statistical analysis based on permutation testing and false discovery rate control, as detailed in the Procedures section. This led to the identification of 90 proteins that were more highly expressed in the TN tumors and 86 proteins that were more highly expressed in the luminal tumors, with an estimated false discovery rate of 5%. These proteins are listed in Tables 2 and 3, respectively. The intensity values for each tumor in each of the three technical repeats are available in Supplementary Table 1 in the Supporting Information.

Individual tumors showed a wide variability of protein expression, as might be expected given the known heterogeneity of tumors, even within the same tumor subtype. This

DOI: 10.1021/pr500987r
J. Proteome Res. 2015, 14, 1376−1388

Table 3. Proteins More Highly Expressed in Luminal Tumors As Identified by Glycoproteomics

acc [alternative acc's]	symbol	protein[a]	no. unique peptides	max Mascot score	fraction	ratio(medianTN/median LUM) TR1	TR2	TR3	FDR (%)
P52569	SLC7A2	low affinity cationic amino acid transporter 2**	1	93.4	N	0.12	0.00	0.11	0.00
Q6ZMP0	THSD4	thrombospondin type-1 domain-containing protein 4**	1	81.06	N	0.08	0.08	0.00	0.01
P21266	GSTM3	glutathione S-transferase Mu 3	7	80.5	T	0.71	0.44	0.65	0.01
Q8NES3	LFNG	beta-1,3-*N*-acetylglucosaminyltransferase lunatic fringe	1	89.01	N	0.67	0.57	0.62	0.01
Q9UBX1	CTSF	cathepsin F	2	108	N	0.59	0.65	0.60	0.01
O43570	CA12	carbonic anhydrase 12	5	69.23	T	0.30	0.33	0.61	0.02
Q9NUM4	TMEM106B	transmembrane protein 106B	2	108.8	NT	0.35	0.28	0.34	0.03
P49961	ENTPD1	ectonucleoside triphosphate diphosphohydrolase 1	6	80.28	NT	0.51	0.58	0.49	0.03
Q5SSJ5	HP1BP3	heterochromatin protein 1-binding protein 3	9	85.82	T	0.58	0.63	0.60	0.03
O15455	TLR3	Toll-like receptor 3	2	41.54	N	0.29	0.37	0.33	0.04
Q9BXN1	ASPN	asporin	19	135.32	NT	0.33	0.37	0.32	0.07
P00846	MT-ATP6	ATP synthase subunit a	1	46.93	T	0.51	0.60	0.39	0.07
Q9NXL6	SIDT1	SID1 transmembrane family member 1**	1	86.41	N	0.00	0.00	0.27	0.08
Q13510	ASAH1	acid ceramidase	16	143.45	NT	0.48	0.42	0.40	0.14
Q09666	AHNAK	neuroblast differentiation-associated protein AHNAK	37	80.97	T	0.71	0.68	0.67	0.14
O95136	S1PR2	sphingosine 1-phosphate receptor 2	1	49.67	N	0.69	0.85	0.59	0.14
O76024	WFS1	wolframin	2	95.23	N	0.36	0.34	0.45	0.15
Q96KP4	CNDP2	cytosolic nonspecific dipeptidase	9	68.44	T	0.65	0.67	0.77	0.17
Q30201	HFE	hereditary hemochromatosis protein	1	56.55	N	0.51	0.61	0.60	0.24
Q8ND94	LRRN4CL	LRRN4 C-terminal-like protein	1	50.27	N	0.00	0.45	0.37	0.27
P10599	TXN	thioredoxin	3	74.1	T	0.51	0.53	0.44	0.27
P15586	GNS	*N*-acetylglucosamine-6-sulfatase	11	128.3	NT	0.59	0.70	0.66	0.29
P55011	SLC12A2	solute carrier family 12 member 2	1	166.93	N	0.53	0.49	0.55	0.30
Q14108	SCARB2	lysosome membrane protein 2	13	148.54	NT	0.76	0.79	0.80	0.34
P11021	HSPA5	78 kDa glucose-regulated protein	22	123.69	T	0.62	0.58	0.62	0.49
O14773	TPP1	tripeptidyl-peptidase 1	7	116	NT	0.75	0.61	0.66	0.50
P04066	FUCA1	tissue alpha-L-fucosidase	7	82.96	NT	0.59	0.90	0.58	0.51
P04899	GNAI2	guanine nucleotide-binding protein G(i) subunit alpha-2	5	95	T	0.61	0.74	0.27	0.51
Q9UIQ6	LNPEP	leucyl-cystinyl aminopeptidase	10	114.45	NT	0.77	0.81	0.47	0.53
P36405 [Q13795]	ARL3	ADP-ribosylation factor-like protein 3	1	73.81	T	0.68	0.87	0.87	0.56
P04156	PRNP	major prion protein	2	67.48	NT	0.86	0.59	0.52	0.70
Q8IUK5	PLXDC1	plexin domain-containing protein 1	1	60.13	N	0.70	0.66	0.78	0.88
Q86TX2 [P49753]	ACOT1	acyl-coenzyme A thioesterase 1	2	95.08	T	0.53	0.32	0.17	0.95
Q8IW00	C10ORF72	uncharacterized Ig-like domain-containing protein C10orf72	1	41.43	N	0.57	0.47	0.39	1.16
P15088	CPA3	mast cell carboxypeptidase A	10	102.36	NT	0.25	0.27	0.41	1.16
O60486	PLXNC1	plexin-C1	5	87.32	NT	0.78	0.68	0.60	1.17
Q7Z2K6	ERMP1	endoplasmic reticulum metallopeptidase 1	10	110.16	NT	0.53	0.54	0.42	1.19
P49748	ACADVL	very long-chain specific acyl-CoA dehydrogenase, mitochondrial	8	79.37	T	0.65	0.78	0.72	1.39
P00750	PLAT	tissue-type plasminogen activator	5	124.28	NT	0.24	0.31	0.30	1.44
Q08380	LGALS3BP	galectin-3-binding protein	20	149.53	NT	0.60	0.64	0.72	1.56
Q5ZPR3	CD276	CD276 antigen	6	125.4	NT	0.60	0.75	0.83	1.59
Q9NY47	CACNA2D2	voltage-dependent calcium channel subunit alpha-2/delta-2	3	103.43	N	0.57	0.46	0.84	1.72
P21860	ERBB3	receptor tyrosine-protein kinase erbB-3	3	135.86	N	0.60	0.62	0.64	1.86
P48735	IDH2	isocitrate dehydrogenase [NADP], mitochondrial	8	80.51	T	0.73	0.78	0.67	1.90
P14384	CPM	carboxypeptidase M	3	85.98	N	0.47	0.47	0.52	1.92
Q99574	SERPINI1	neuroserpin	1	86.37	N	0.33	0.45	0.77	1.93
Q16531	DDB1	DNA damage-binding protein 1	2	61.67	T	0.66	0.62	0.42	1.98
P55884	EIF3B	eukaryotic translation initiation factor 3 subunit B	4	64.15	T	0.45	0.64	0.80	2.31
O60343	TBC1D4	TBC1 domain family member 4	1	42.7	N	0.69	0.92	0.38	2.33

DOI: 10.1021/pr500987r
J. Proteome Res. 2015, 14, 1376−1388

Table 3. continued

acc [alternative acc's]	symbol	protein[a]	no. unique peptides	max Mascot score	fraction	ratio(medianTN/median LUM) TR1	TR2	TR3	FDR (%)
P15311	EZR	ezrin	6	116.06	T	0.74	0.77	0.81	2.40
P06576	ATP5B	ATP synthase subunit beta, mitochondrial	22	122.87	T	0.81	0.75	0.61	2.70
Q15661 [P20231]	TPSAB1	tryptase alpha/beta-1	8	124.27	NT	0.29	0.38	0.32	2.70
Q10589	BST2	bone marrow stromal antigen 2	5	112.32	NT	0.60	0.52	0.66	2.90
P05787	KRT8	keratin, type II cytoskeletal 8	14	101.31	T	0.51	0.62	0.44	3.05
P17301	ITGA2	integrin alpha-2	7	104.91	NT	0.67	0.67	0.65	3.15
P08758	ANXA5	annexin A5**	12	143.03	T	0.91	0.81	0.86	3.16
Q12999	TSPAN31	tetraspanin-31	1	59.19	T	0.46	0.50	0.95	3.23
Q93091	RNASE6	ribonuclease K6	2	49.67	N	0.54	0.57	0.78	3.24
P05154	SERPINA5	plasma serine protease inhibitor	6	113.13	NT	0.71	0.54	0.82	3.25
P12544	GZMA	granzyme A	1	70.02	T	0.63	0.51	0.41	3.28
Q96K49	TMEM87B	transmembrane protein 87B	2	60.97	N	0.28	0.42	0.49	3.28
P26006	ITGA3	integrin alpha-3	6	58.84	N	0.75	0.51	0.68	3.33
Q13641	TPBG	trophoblast glycoprotein	7	91.22	NT	0.79	0.67	0.82	3.40
P40763	STAT3	signal transducer and activator of transcription 3	2	85.72	T	0.67	0.72	0.61	3.46
P01033	TIMP1	metalloproteinase inhibitor 1	6	113.07	NT	0.67	0.82	0.64	3.52
P49281	SLC11A2	natural resistance-associated macrophage protein 2	2	47.43	N	0.65	0.59	0.64	3.71
Q68CP4	HGSNAT	heparan-alpha-glucosaminide *N*-acetyltransferase	3	90.61	N	0.53	0.70	0.67	3.80
Q6UX71	PLXDC2	plexin domain-containing protein 2	8	149.14	NT	0.74	0.86	0.71	3.85
P15086	CPB1	carboxypeptidase B	9	127.36	T	0.43	0.45	0.54	3.94
Q9UHG3	PCYOX1	prenylcysteine oxidase 1	14	97.2	NT	0.78	0.72	0.75	4.05
Q14204	DYNC1H1	cytoplasmic dynein 1 heavy chain 1	19	104.96	T	0.92	0.84	0.66	4.09
Q96PB1	CASD1	CAS1 domain-containing protein 1	4	107.48	NT	0.48	0.46	0.41	4.10
O43237	DYNC1LI2	cytoplasmic dynein 1 light intermediate chain 2	2	93.36	T	0.51	0.63	0.44	4.13
P30101	PDIA3	protein disulfide-isomerase A3	22	130.33	NT	0.85	0.79	0.77	4.13
P08123	COL1A2	collagen alpha-2(I) chain	16	122.41	NT	0.80	0.79	0.43	4.17
Q99571	P2RX4	P2X purinoceptor 4	3	74.07	N	0.57	0.52	0.72	4.17
Q02218	OGDH	2-oxoglutarate dehydrogenase, mitochondrial	3	57.76	T	0.79	0.81	0.77	4.49
Q9C0C4	SEMA4C	semaphorin-4C	1	57.71	N	0.39	0.44	0.95	4.49
Q9UKY0	PRND	prion-like protein doppel	1	66.92	N	0.35	0.38	0.00	4.50
P11940	PABPC1	polyadenylate-binding protein 1	3	64.14	T	0.62	0.41	0.67	4.51
P08069	IGF1R	insulin-like growth factor 1 receptor	6	97.39	N	0.50	0.70	0.63	4.51
Q3MIR4	TMEM30B	cell cycle control protein 50B	1	55.78	N	0.55	0.46	0.78	4.53
P26885	FKBP2	peptidyl-prolyl cis−trans isomerase FKBP2	1	50.86	T	0.71	0.88	0.66	4.55
P50897	PPT1	palmitoyl-protein thioesterase 1	10	152.42	NT	0.71	0.70	0.77	4.67
Q9P2C4	TMEM181	transmembrane protein 181	3	109.61	NT	0.74	0.59	0.36	4.79
P14415	ATP1B2	sodium/potassium-transporting ATPase subunit beta-2	1	38.79	N	0.34	0.30	0.25	4.88

[a]Proteins that were validated by microarray ($p \leq 0.05$) are underlined and proteins marked by a double asterisk (**) are members of the 29-protein subset.

variability can be seen by looking at the relative expression in a matrix format for the 90 proteins overexpressed and the 86 proteins underexpressed in TN tumors relative to the LUM tumors (Figure 2). This view makes it possible to see the expression of these potential markers of TN and LUM tumors in each of the 18 individual tumors analyzed. Interestingly, the TN tumor T171 appears to be more similar to the other luminal tumors in expression of these markers than to the other TN tumors.

Verification of MS Quantification by Western Blot for Selected Proteins Differentially Expressed in TN Tumors

We selected proteins for quantitative verification based on both their potential as therapeutic or diagnostic targets and the availability of antibodies. Figure 3 provides a more detailed view of the glycoproteomic MS data for the four proteins selected for verification by Western blot: acid ceramidase (ASAH1), desmocollin-2 (DSC2), cytokine receptor-like factor 1 (CRLF1), and azurocidin (AZU1). For each of these proteins the intensity in each of the nine TN and nine luminal tumors is shown for each technical replicate. Three of these proteins (DSC2, CRLF1, and AZU1) were more highly expressed in the TN than luminal tumors, while ASAH1 was more highly expressed in luminal tumors. This view clearly shows the heterogeneity of protein expression in the tumors from individual patients.

We next took advantage of this tumor heterogeneity to confirm the quantitative data calculated in our glycoproteomic study by comparing these results with those found by Western blotting. As can be seen in Figure 4, the mass-spectrometric-

DOI: 10.1021/pr500987r
J. Proteome Res. 2015, 14, 1376−1388

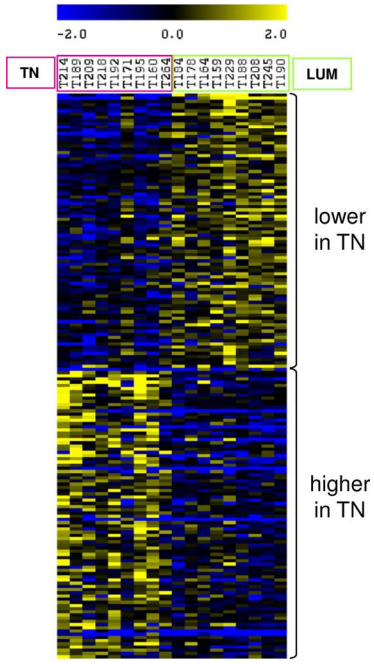


Figure 2. Proteins differentially expressed between TN and luminal tumors as determined by glycoproteomics. Matrix view of the 176 proteins found to be either under- or overexpressed in TN tumors, relative to the luminal tumors. Yellow represents expression above the median intensity value, while blue represents expression below the median intensity value as shown in $\log_2$ scaling. Intensity values were normalized to the median intensity across all of the tumors. The median intensity value from the three technical repeats is shown. Proteins with median intensities of zero were normalized to half of the maximum intensity value, and tumors showing no expression of a protein were assigned a $\log_2$ intensity of −3.

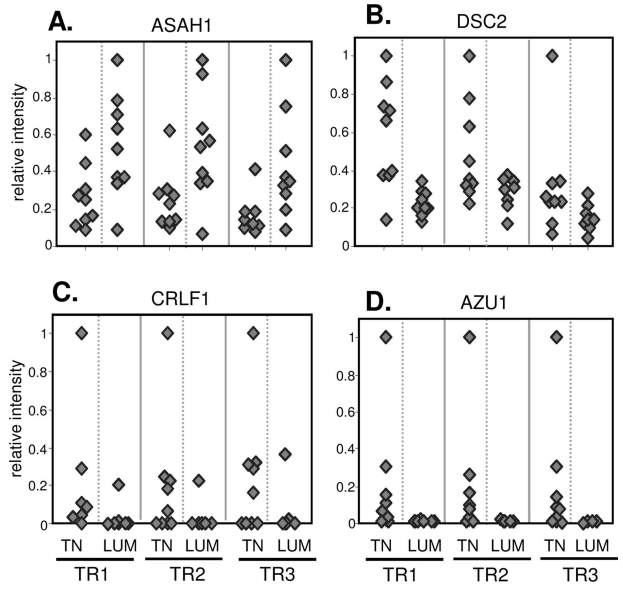


Figure 3. Relative protein intensities calculated for individual tumors in each MS technical replicate (TR1, TR2, TR3) for four proteins identified as differentially expressed between TN and luminal tumors. (A) Acid ceramidase (ASAH1), (B) desmocollin-2 (DSC2), (C) cytokine receptor-like factor 1 (CRLF1), and (D) azurocidin (AZU1).

based quantification correlates well with the Western blot results for ASAH1, DSC2, and AZU1. In the case of CRLF1,

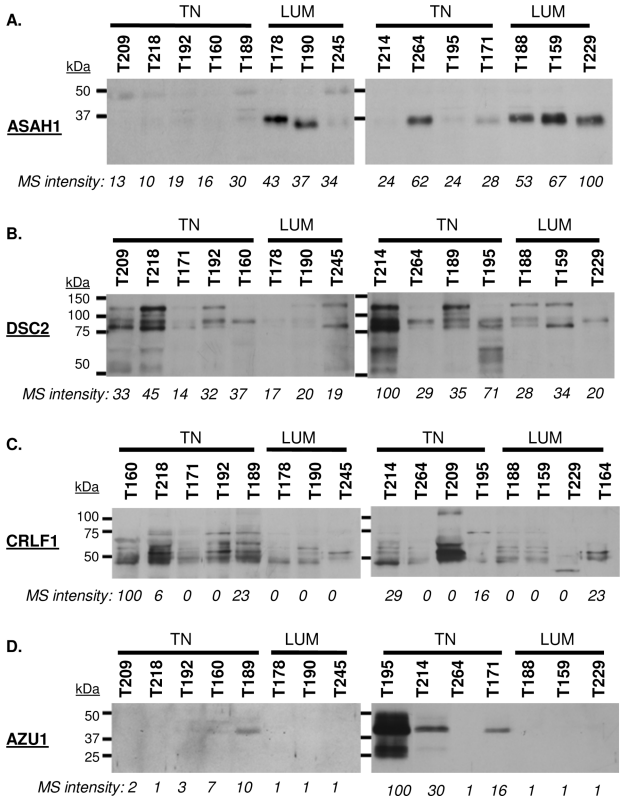


Figure 4. Western blot validation of quantitative mass spectrometry results. Western blotting was performed on lysates from the original set of TN and luminal tumors for the following proteins: (A) acid ceramidase (ASAH1), (B) desmocollin-2 (DSC2), (C) cytokine receptor-like factor 1 (CRLF1); and (D) azurocidin (AZU1). For comparison, the intensity of the protein determined by mass spectrometry is shown for each tumor as a percentage of the maximum intensity. (The median value of the three TRs is shown.) Limited or insufficient amounts of tumor lysate prevented or limited Western blot analysis of the luminal tumors T184, T208, and T164.

Western blotting confirmed our finding that CRLF1 appears to be relatively overexpressed in the TN relative to the luminal tumors. However, there were considerable differences in the relative quantification of this protein in the individual tumors between the two methods. This inconsistency could be due to a combination of factors, including the presence of multiple glycosylation states, as suggested by the banding pattern on the Western blot, which may affect binding to the hydrazide resin. In addition, the relatively low level of expression of this protein, as evidenced by the large number of tumors in which the protein was not found by mass spectrometry (MS intensity assigned to 0), likely also contributes to the observed differences.

Validation of Glycoproteomic Hits in a Larger Cohort of Samples by Comparison with Microarray Data

Although we have shown good correlation between protein expression levels determined by our glycoproteomic workflow and by Western blot, the interpretation of our glycoproteomics data is still limited by the relatively small number of samples used in our analysis. Therefore, to validate our findings in a larger cohort of tumors, we utilized publicly available microarray data. In total, mRNA expression level data were compiled from 469 Her2−/ER+/PR+ and 406 TN breast tumors. Of the 176 differentially expressed proteins identified

DOI: 10.1021/pr500987r
J. Proteome Res. 2015, 14, 1376−1388

by glycoproteomics, 10 were not mapped to a corresponding gene on the arrays. Of the remaining 166 proteins, 53 (32%) also had a corresponding difference in mRNA levels in this larger cohort of tumors using a shrinkage t test ($p \leq 0.05$). The proteins validated through this analysis are underlined in Tables 2 and 3.

We next explored whether proteins that did not have a significant expression difference at the mRNA level in this larger cohort might show smaller, nonsignificant differences in mRNA expression that concur with our glycoproteomic results. To do this, we visualized trends in mRNA expression in this larger cohort of tumors for all of the differentially expressed proteins identified in our glycoproteomic study by plotting their normalized mRNA levels in the TN and luminal groups on a log–log scale (Figure 5). For clarity, we also plotted a line

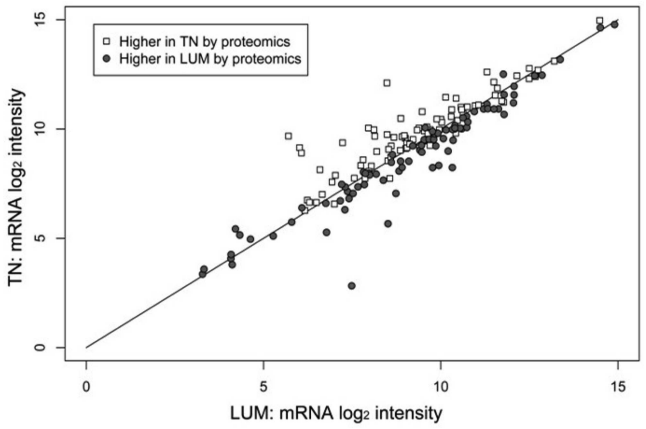


Figure 5. Log–log plot of mRNA levels for proteins found to be differentially expressed between TN and luminal tumors by glycoproteomics. Proteins identified as higher in TN by glycoproteomics (open diamonds) tend to show a corresponding difference in their mRNA level, as evidenced by their enrichment above the 1:1 ratio line. Conversely, proteins identified as more highly expressed in luminal tumors (gray squares) are enriched below the 1:1 ratio line, thus further validating these proteins in a larger cohort of tumors.

along the diagonal to represent equal mRNA levels in the TN and LUM groups. Interestingly, proteins identified as more highly expressed in TN tumors are enriched above this line, suggesting that the mRNA levels of these proteins were also, on average, higher in the TN tumors than in the luminal tumors, even if these differences did not reach statistically significant levels. Conversely, proteins that are more highly expressed in luminal tumors tended to have higher mRNA levels in luminal tumors and thus fall below the 1:1 line. This finding suggests that for the majority of proteins identified in this study there are corresponding differences in mRNA levels in this larger, unrelated cohort of samples and provides further confidence in the general applicability of our glycoproteomic results.

Identification of a 29-Member Set That Is Capable of Distinguishing TN and LUM Tumors and Predicting Outcomes in Triple-Negative Breast Cancer Patients

To determine whether the proteins identified in this study were capable of distinguishing TN and LUM tumors, we focused on the subset of the proteins that showed the largest differential expression. Thus, we selected the 31 proteins whose average expression was more than 2.5 times between the TN and luminal tumors. Of these 31 proteins, 29 were mapped onto the microarray chips, of which 25 were more highly expressed in TN and 4 were more highly expressed in luminal tumors. The proteins contained in this 29-protein set are marked by a double asterisk in Tables 1 and 2 and are also listed in Supplementary Table 2 in the Supporting Information. We next evaluated the corresponding 29-gene set, containing the genes that encode the 29 proteins selected based on our glycoproteomic results, for its ability to classify luminal ER+/PR+/Her2− and TN samples in a microarray data set (GSE25066) containing 216 ER+/PR+/Her2− and 178 TN samples. Clustering 394 samples using this gene signature, 186 of the 216 luminal (86.1%) and 161 of the 178 TN samples (90.4%) were correctly classified. Thus, this gene set is able to classify ER+/PR+/Her2− (LUM) and TN samples ($p < 3.0 \times 10^{-21}$).

We further evaluated the ability of this 29-gene signature to predict outcomes, specifically in TN breast cancer patients. To this end, we conducted a survival analysis on a collection of 224 TN samples curated from seven individual microarray studies. We found that classification of these patients using our 29-gene signature significantly predicts survival ($P = 0.03$), as shown by the ability of this classifier to separate the survival curves (Figure 6). Taken together, these results suggest that the

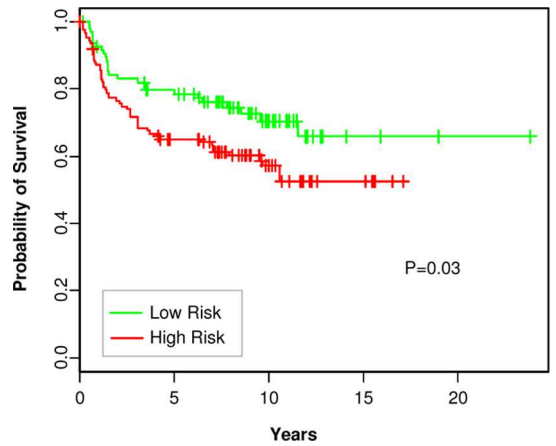


Figure 6. Subset of proteins differentially expressed between TN and luminal tumors can separate survival curves in a triple-negative patient population. TN breast tumors with associated survival information ($N = 224$) were extracted from publically available microarray data sets and split into "high risk" (red) and "low risk" (green) groups based on the mRNA levels corresponding to a 29-protein set containing proteins with ≥2.5 fold expression difference between TN and luminal tumors.

proteins encoded by this 29-gene signature may be useful as a marker set both to distinguish LUM and TN sample and to predict the outcomes of TN patients.

DISCUSSION

To date, there are a limited number of published proteomic studies exploring differences in protein expression in clinical breast tissue from TN and Her2−/ER+/PR+ tumors, the comparison that we used here for our glycoproteomic approach. Lu et al. used hydrophobic protein enrichment and spectral counting to identify proteins that may be differentially expressed between these tumor types; however, overlap with our data set is limited, likely due to the different protein enrichment methods used. Both studies identified HLA class-II proteins, alpha-enolase, fibrinogen, and annexin A1 as more highly expressed in TN tumors and galectin-3 binding protein

DOI: 10.1021/pr500987r
J. Proteome Res. 2015, 14, 1376−1388

as more highly expressed in luminal tumors. More recently, a couple of comprehensive proteomic studies using cancer-derived cell lines have been published. Geiger et al. published a comprehensive SILAC-based proteomic data set on a series of cell lines, a subset of which were originally isolated from TN tumors while others were from Her2+ tumors.[41] Because this study did not include any Her2−/ER+/PR+-derived cells and focused on cancer-stage progression rather than subtype, direct comparison with our data is difficult. Another study profiled the conditioned media from five luminal-derived and five TN-derived cell lines using lectin purification. Although none of the cell lines used were classified as Her2−/ER+/PR+, 7 of the 83 proteins identified as highly expressed in the TN cell lines in this published study were also found in our glycoproteomic comparison, including APP, BMP1, CDH2, IGFBP3, MET, MFGE8, and THBS3.[42]

In this study, we chose to analyze each of the individual tumors separately rather than pooling them prior to analysis. Although this approach added greatly to the mass spectrometry analysis time, it had the advantage of allowing us to explore the heterogeneity in protein expression displayed by individual tumors, which we found to be considerable. This heterogeneity complicates the statistical analysis. Although we have set the false discovery rate based on permutation testing at 5%, this is likely to be an underestimate because it does not account for proteins that may not be truly differentially expressed when tested in a larger cohort of samples, even though they are differentially expressed between the nine TN and nine luminal tumors analyzed in this study. To partially alleviate this concern, we used publically available microarray data to validate the differential expression of 30% of the proteins identified in this study through corresponding changes in mRNA level in a separate, larger set of TN and LUM tumors. These data overcome the major limitation of the glycoproteomic study, namely, the limited number of individual samples.

Because proteomic studies, even with enrichment techniques such as hydrazide capture, tend to identify and quantify more abundant proteins, the proteins both identified by glycoproteomics and validated by microarray represent excellent potential targets for therapeutic or diagnostic reagents. Despite the value of using microarray data in this way, it is important to note that although a corresponding difference in mRNA and protein levels is reassuring the lack of a significant difference in mRNA levels does not preclude the use of the corresponding protein as a useful target of TNBC. First, there is known to be a relatively weak correlation between mRNA and protein levels (reviewed in ref 43). Second, because of a lack of statistical power, it is difficult for small differences in mRNA levels to be detected as statistically significant, despite the fact that these small differences may lead to even larger differences in protein concentration due to amplification. In fact, when we plotted the mRNA levels for all of the proteins identified as differentially expressed by glycoproteomics, it became clear that many of these proteins show small but statistically insignificant differences in their mRNA transcript levels.

Patients diagnosed with luminal Her2−/ER+/PR+ tumors generally have a higher 5-year overall survival rate when compared with TN tumors.[2] In this study, we showed that the expression of mRNAs of the 29 proteins that were most differentially expressed between TN and luminal tumors can separate survival curves within a TN patient population, stratifying tumors into a better prognosis group or a worse prognosis group. Thus, even in tumors classified as TN, expression profiles that are more similar to luminal tumors can impart a survival advantage. This suggests that TN tumors can show varying degrees of "luminal-like" characteristics and that these characteristics may have prognostic value. It further suggests that at least some of the proteins represented in the 29-gene signature may play an important role in the lower survival rates seen in TN cancers.

Molecular profiling based on microarray data has identified five subtypes of breast cancer.[44] The majority of TN tumors fall into the basal-like breast cancer subtype, which express proteins characteristic of basal epithelial cells.[45,46] The proteins that we have identified in this study as differentially expressed between TN and luminal (Her2−, PR+, ER+) breast cancer tumors include some proteins characteristic of the basal-like cancers, such as cytokeratin 14 (KRT14) and *p*-cadherin (CDH3) as well as keratin 8, a known marker of luminal breast cancers.[47,48] However, the majority of these proteins do not have a previous association with TNBC.

Several proteins identified in this study as highly expressed in TN tumors are associated with degranulating neutrophils, suggesting a role for the innate immune response in TN breast cancer biology. These proteins include azurocidin, neutrophil defensin, neutrophil gelatinase-associated lipocalin, myeloperoxidase, maltase-glucoamylase, MMP-9, and myeloblastin. In fact, when we compared our results with proteins identified in a proteomic study of neutrophil granule proteins, 17 (18.9%) of the proteins more highly expressed in TN were also identified in the neutrophil granules, while only 5 (5.8%) of proteins that were highly expressed in luminal tumors were found.[49] Interestingly, a high level of intratumoral neutrophils has recently been associated with poor prognosis in renal cell carcinoma, nonsmall cell lung carcinoma, and melanoma,[50−52] suggesting that neutrophils likely play an important role in cancer progression. Furthermore, clustering-analysis-based gene expression profiles of TN breast tumors identified an immunomodulatory subtype.[53] Thus, it seems reasonable to hypothesize that increased neutrophil infiltration in TN breast cancer may account for some of the difference in survival seen between luminal and TN breast cancers.

In this study, we have identified many potential markers of TN breast cancer. These proteins may represent potential therapeutic targets for TN breast cancer, a disease that currently lacks targeted therapies. In addition, given the strong bias toward extracellular and secreted proteins introduced by our focus on glycosylated proteins, a subset of these proteins may prove to be diagnostic or predictive biomarkers depending on their levels in serum. Future experiments will be focused on the function of a subset of these proteins in cancer and their expression in the serum of breast cancer patients.

■ ASSOCIATED CONTENT

Supporting Information

Supplementary Table 1: Detailed identification and quantification data for all identified proteins. Supplementary Table 2: Members of the 29-protein set. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: 613-993-7206. Fax: 613-952-9052.

Notes

The authors declare no competing financial interest.

DOI: 10.1021/pr500987r
J. Proteome Res. 2015, 14, 1376−1388

■ ACKNOWLEDGMENTS

We thank John Kelly for scientific and administrative support of the mass spectrometry facility, Denis Bourbeau for his project management expertise, Marguerite Buchanan for maintaining the tumor biobank, and Luc Tessier and Anna Robotham for maintenance of the LC−MS systems. We also acknowledge the FRQS-Réseau de Recherche Cancer axe cancer du sein/ovaire and the Quebec Breast Cancer Foundation for support of the biobank.

■ REFERENCES

(1) National Cancer Institute. Triple-Negative Breast Cancer Disproportionately Affects African American and Hispanic Women. *NCI Cancer Bull.* **2007**, *22* (4), 7.

(2) Dent, R.; Trudeau, M.; Pritchard, K. I.; Hanna, W. M.; Kahn, H. K.; Sawka, C. A.; Lickley, L. A.; Rawlinson, E.; Sun, P.; Narod, S. A. Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin. Cancer Res.* **2007**, *13* (15 Pt 1), 4429−4434.

(3) Brenton, J. D.; Carey, L. A.; Ahmed, A. A.; Caldas, C. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J. Clin. Oncol.* **2005**, *23* (29), 7350−7360.

(4) Rouzier, R.; Perou, C. M.; Symmans, W. F.; Ibrahim, N.; Cristofanilli, M.; Anderson, K.; Hess, K. R.; Stec, J.; Ayers, M.; Wagner, P.; Morandi, P.; Fan, C.; Rabiul, I.; Ross, J. S.; Hortobagyi, G. N.; Pusztai, L. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin. Cancer Res.* **2005**, *11* (16), 5678−5685.

(5) Schulz, D. M.; Bollner, C.; Thomas, G.; Atkinson, M.; Esposito, I.; Hofler, H.; Aubele, M. Identification of differentially expressed proteins in triple-negative breast carcinomas using DIGE and mass spectrometry. *J. Proteome Res.* **2009**, *8* (7), 3430−3438.

(6) He, J.; Whelan, S. A.; Lu, M.; Shen, D.; Chung, D. U.; Saxton, R. E.; Faull, K. F.; Whitelegge, J. P.; Chang, H. R. Proteomic-based biosignatures in breast cancer classification and prediction of therapeutic response. *Int. J. Proteomics* **2011**, *2011*, 896476.

(7) Cabezon, T.; Gromova, I.; Gromov, P.; Serizawa, R.; Timmermans Wielenga, V.; Kroman, N.; Celis, J. E.; Moreira, J. M. Proteomic profiling of triple-negative breast carcinomas in combination with a three-tier orthogonal technology approach identifies Mage-A4 as potential therapeutic target in estrogen receptor negative breast cancer. *Mol. Cell. Proteomics* **2013**, *12* (2), 381−394.

(8) Lu, M.; Whelan, S. A.; He, J.; Saxton, R. E.; Faull, K. F.; Whitelegge, J. P.; Chang, H. R. Hydrophobic Proteome Analysis of Triple Negative and Hormone-Receptor-Positive-Her2−Negative Breast Cancer by Mass Spectrometer. *Clin. Proteomics* **2010**, *6* (3), 93−103.

(9) Wei, X.; Li, L. Comparative glycoproteomics: approaches and applications. *Briefings Funct. Genomics* **2008**, *8* (2), 104−113.

(10) Drake, R. R.; Schwegler, E. E.; Malik, G.; Diaz, J.; Block, T.; Mehta, A.; Semmes, O. J. Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. *Mol. Cell. Proteomics* **2006**, *5* (10), 1957−67.

(11) Ito, S.; Hayama, K.; Hirabayashi, J. Enrichment strategies for glycopeptides. *Methods Mol. Biol.* **2009**, *534*, 195−203.

(12) Ding, W.; Hill, J. J.; Kelly, J. Selective enrichment of glycopeptides from glycoprotein digests using ion-pairing normal-phase liquid chromatography. *Anal. Chem.* **2007**, *79* (23), 8891−8899.

(13) Palmisano, G.; Lendal, S. E.; Engholm-Keller, K.; Leth-Larsen, R.; Parker, B. L.; Larsen, M. R. Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc.* **2010**, *5* (12), 1974−1982.

(14) Pan, S.; Chen, R.; Aebersold, R.; Brentnall, T. A. Mass spectrometry based glycoproteomics–from a proteomics perspective. *Mol. Cell. Proteomics* **2011**, *10* (1), R110 003251.

(15) Sun, B.; Ranish, J. A.; Utleg, A. G.; White, J. T.; Yan, X.; Lin, B.; Hood, L. Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics. *Mol. Cell, Proteomics* **2007**, *6* (1), 141−149.

(16) 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.

(17) Arcinas, A.; Yen, T. Y.; Kebebew, E.; Macher, B. A. Cell surface and secreted protein profiles of human thyroid cancer cell lines reveal distinct glycoprotein patterns. *J. Proteome Res.* **2009**, *8* (8), 3958−3968.

(18) Cao, J.; Shen, C.; Wang, H.; Shen, H.; Chen, Y.; Nie, A.; Yan, G.; Lu, H.; Liu, Y.; Yang, P. Identification of N-glycosylation sites on secreted proteins of human hepatocellular carcinoma cells with a complementary proteomics approach. *J. Proteome Res.* **2009**, *8* (2), 662−672.

(19) Soltermann, A.; Ossola, R.; Kilgus-Hawelski, S.; von Eckardstein, A.; Suter, T.; Aebersold, R.; Moch, H. N-glycoprotein profiling of lung adenocarcinoma pleural effusions by shotgun proteomics. *Cancer* **2008**, *114* (2), 124−133.

(20) Hill, J. J.; Moreno, M. J.; Lam, J. C.; Haqqani, A. S.; Kelly, J. F. Identification of secreted proteins regulated by cAMP in glioblastoma cells using glycopeptide capture and label-free quantification. *Proteomics* **2009**, *9* (3), 535−549.

(21) Whelan, S. A.; Lu, M.; He, J.; Yan, W.; Saxton, R. E.; Faull, K. F.; Whitelegge, J. P.; Chang, H. R. Mass spectrometry (LC-MS/MS) site-mapping of N-glycosylated membrane proteins for breast cancer biomarkers. *J. Proteome Res.* **2009**, *8* (8), 4151−4160.

(22) Zeng, X.; Hood, B. L.; Sun, M.; Conrads, T. P.; Day, R. S.; Weissfeld, J. L.; Siegfried, J. M.; Bigbee, W. L. Lung cancer serum biomarker discovery using glycoprotein capture and liquid chromatography mass spectrometry. *J. Proteome Res.* **2010**, *9* (12), 6440−6449.

(23) Tian, Y.; Esteva, F. J.; Song, J.; Zhang, H. Altered expression of sialylated glycoproteins in breast cancer using hydrazide chemistry and mass spectrometry. *Mol. Cell. Proteomics* **2012**, *11* (6), M111 011403.

(24) Tian, Y.; Bova, G. S.; Zhang, H. Quantitative glycoproteomic analysis of optimal cutting temperature-embedded frozen tissues identifying glycoproteins associated with aggressive prostate cancer. *Anal. Chem.* **2011**, *83* (18), 7013−7019.

(25) Tian, Y.; Yao, Z.; Roden, R. B.; Zhang, H. Identification of glycoproteins associated with different histological subtypes of ovarian tumors using quantitative glycoproteomics. *Proteomics* **2011**, *11* (24), 4677−4687.

(26) Haqqani, A. S.; Kelly, J. F.; Stanimirovic, D. B. Quantitative protein profiling by mass spectrometry using label-free proteomics. *Methods Mol. Biol.* **2008**, *439*, 241−256.

(27) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M. Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918−920.

(28) Hsieh, E. J.; Hoopmann, M. R.; MacLean, B.; MacCoss, M. J. Comparison of database search strategies for high precursor mass accuracy MS/MS data. *J. Proteome Res.* **2010**, *9* (2), 1138−1143.

(29) Elias, J. E.; Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, *4* (3), 207−214.

(30) Benjamini, Y.; Hockberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc., Ser. B* **1995**, *57* (1), 289−300.

(31) Edgar, R.; Domrachev, M.; Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30* (1), 207−210.

DOI: 10.1021/pr500987r
J. Proteome Res. 2015, 14, 1376−1388

(32) Zhu, Y.; Davis, S.; Stephens, R.; Meltzer, P. S.; Chen, Y. GEOmetadb: powerful alternative search engine for the Gene Expression Omnibus. *Bioinformatics* **2008**, *24* (23), 2798−800.
(33) Gentleman, R. C.; Carey, V. J.; Bates, D. M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; Hornik, K.; Hothorn, T.; Huber, W.; Iacus, S.; Irizarry, R.; Leisch, F.; Li, C.; Maechler, M.; Rossini, A. J.; Sawitzki, G.; Smith, C.; Smyth, G.; Tierney, L.; Yang, J. Y.; Zhang, J. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **2004**, *5* (10), R80.
(34) Hubbell, E.; Liu, W. M.; Mei, R. Robust estimators for expression analysis. *Bioinformatics* **2002**, *18* (12), 1585−1592.
(35) Dai, M.; Wang, P.; Boyd, A. D.; Kostov, G.; Athey, B.; Jones, E. G.; Bunney, W. E.; Myers, R. M.; Speed, T. P.; Akil, H.; Watson, S. J.; Meng, F. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* **2005**, *33* (20), e175.
(36) Opgen-Rhein, R.; Strimmer, K. Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Stat. Appl. Genet. Mol. Biol.* **2007**, *6*, No. 9.
(37) Shows, T. B.; Alper, C. A.; Bootsma, D.; Dorf, M.; Douglas, T.; Huisman, T.; Kit, S.; Klinger, H. P.; Kozak, C.; Lalley, P. A.; Lindsley, D.; McAlpine, P. J.; McDougall, J. K.; Meera Khan, P.; Meisler, M.; Morton, N. E.; Opitz, J. M.; Partridge, C. W.; Payne, R.; Roderick, T. H.; Rubinstein, P.; Ruddle, F. H.; Shaw, M.; Spranger, J. W.; Weiss, K. International system for human gene nomenclature (1979) ISGN (1979). *Cytogenet. Cell Genet.* **1979**, *25* (1−4), 96−116.
(38) Li, J.; Lenferink, A. E.; Deng, Y.; Collins, C.; Cui, Q.; Purisima, E. O.; O'Connor-McCourt, M. D.; Wang, E. Identification of high-quality cancer prognostic markers and metastasis network modules. *Nat. Commun.* **2010**, *1*, 34.
(39) Karn, T.; Metzler, D.; Ruckhaberle, E.; Hanker, L.; Gatje, R.; Solbach, C.; Ahr, A.; Schmidt, M.; Holtrich, U.; Kaufmann, M.; Rody, A. Data-driven derivation of cutoffs from a pool of 3,030 Affymetrix arrays to stratify distinct clinical types of breast cancer. *Breast Cancer Res. Treat.* **2010**, *120* (3), 567−579.
(40) Cui, Q.; Ma, Y.; Jaramillo, M.; Bari, H.; Awan, A.; Yang, S.; Zhang, S.; Liu, L.; Lu, M.; O'Connor-McCourt, M.; Purisima, E. O.; Wang, E. A map of human cancer signaling. *Mol. Syst. Biol.* **2007**, *3*, 152.
(41) Geiger, T.; Madden, S. F.; Gallagher, W. M.; Cox, J.; Mann, M. Proteomic portrait of human breast cancer progression identifies novel prognostic markers. *Cancer Res.* **2012**, *72* (9), 2428−2439.
(42) Drake, P. M.; Schilling, B.; Niles, R. K.; Prakobphol, A.; Li, B.; Jung, K.; Cho, W.; Braten, M.; Inerowicz, H. D.; Williams, K.; Albertolle, M.; Held, J. M.; Iacovides, D.; Sorensen, D. J.; Griffith, O. L.; Johansen, E.; Zawadzka, A. M.; Cusack, M. P.; Allen, S.; Gormley, M.; Hall, S. C.; Witkowska, H. E.; Gray, J. W.; Regnier, F.; Gibson, B. W.; Fisher, S. J. Lectin chromatography/mass spectrometry discovery workflow identifies putative biomarkers of aggressive breast cancers. *J. Proteome Res.* **2012**, *11* (4), 2508−2520.
(43) Maier, T.; Guell, M.; Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **2009**, *583* (24), 3966−3973.
(44) Perou, C. M.; Sorlie, T.; Eisen, M. B.; van de Rijn, M.; Jeffrey, S. S.; Rees, C. A.; Pollack, J. R.; Ross, D. T.; Johnsen, H.; Akslen, L. A.; Fluge, O.; Pergamenschikov, A.; Williams, C.; Zhu, S. X.; Lonning, P. E.; Borresen-Dale, A. L.; Brown, P. O.; Botstein, D. Molecular portraits of human breast tumours. *Nature* **2000**, *406* (6797), 747−752.
(45) Kreike, B.; van Kouwenhove, M.; Horlings, H.; Weigelt, B.; Peterse, H.; Bartelink, H.; van de Vijver, M. J. Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. *Breast Cancer Res.* **2007**, *9* (5), R65.
(46) Thike, A. A.; Cheok, P. Y.; Jara-Lazaro, A. R.; Tan, B.; Tan, P.; Tan, P. H. Triple-negative breast cancer: clinicopathological characteristics and relationship with basal-like breast cancer. *Mod. Pathol.* **2010**, *23* (1), 123−133.
(47) Abd El-Rehim, D. M.; Pinder, S. E.; Paish, C. E.; Bell, J.; Blamey, R. W.; Robertson, J. F.; Nicholson, R. I.; Ellis, I. O. Expression of luminal and basal cytokeratins in human breast carcinoma. *J. Pathol.* **2004**, *203* (2), 661−671.
(48) Rastelli, F.; Biancanelli, S.; Falzetta, A.; Martignetti, A.; Casi, C.; Bascioni, R.; Giustini, L.; Crispino, S. Triple-negative breast cancer: current state of the art. *Tumori* **2010**, *96* (6), 875−888.
(49) Lominadze, G.; Powell, D. W.; Luerman, G. C.; Link, A. J.; Ward, R. A.; McLeish, K. R. Proteomic analysis of human neutrophil granules. *Mol. Cell. Proteomics* **2005**, *4* (10), 1503−1521.
(50) Donskov, F.; von der Maase, H. Impact of immune parameters on long-term survival in metastatic renal cell carcinoma. *J. Clin. Oncol.* **2006**, *24* (13), 1997−2005.
(51) Ilie, M.; Hofman, V.; Ortholan, C.; Bonnetaud, C.; Coelle, C.; Mouroux, J.; Hofman, P. Predictive clinical outcome of the intratumoral CD66b-positive neutrophil-to-CD8-positive T-cell ratio in patients with resectable nonsmall cell lung cancer. *Cancer* **2012**, *118* (6), 1726−1737.
(52) Jensen, T. O.; Schmidt, H.; M?ller, H. J.; Donskov, F.; Hoyer, M.; Sjoegren, P.; Christensen, I. J.; Steiniche, T. Intratumoral neutrophils and plasmacytoid dendritic cells indicate poor prognosis and are associated with pSTAT3 expression in AJCC stage I/II melanoma. *Cancer* **2012**, *118* (9), 2476−2485.
(53) Lehmann, B. D.; Bauer, J. A.; Chen, X.; Sanders, M. E.; Chakravarthy, A. B.; Shyr, Y.; Pietenpol, J. A. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **2011**, *121* (7), 2750−2767.
(54) Palagi, P. M.; Walther, D.; Quadroni, M.; Catherinet, S.; Burgess, J.; Zimmermann-Ivol, C. G.; Sanchez, J. C.; Binz, P. A.; Hochstrasser, D. F.; Appel, R. D. MSight: an image analysis software for liquid chromatography-mass spectrometry. *Proteomics* **2005**, *5* (9), 2381−2384.

DOI: 10.1021/pr500987r
J. Proteome Res. 2015, 14, 1376−1388