

Glycoproteomic Analysis of Prostate Cancer Tissues by SWATH Mass Spectrometry Discovers N-acylethanolamine Acid Amidase and Protein Tyrosine Kinase 7 as Signatures for Tumor Aggressiveness*[§]

Yansheng Liu^{‡§}, Jing Chen^{§¶}, Atul Sethi[‡], Qing K. Li[¶], Lijun Chen[¶], Ben Collins[‡], Ludovic C. J. Gillet[‡], Bernd Wollscheid[‡], Hui Zhang^{¶||}, and Ruedi Aebersold^{‡||**}

The identification of biomarkers indicating the level of aggressiveness of prostate cancer (PCa) will address the urgent clinical need to minimize the general overtreatment of patients with non-aggressive PCa, who account for the majority of PCa cases. Here, we isolated formerly N-linked glycopeptides from normal prostate ($n = 10$) and from non-aggressive ($n = 24$), aggressive ($n = 16$), and metastatic ($n = 25$) PCa tumor tissues and analyzed the samples using SWATH mass spectrometry, an emerging data-independent acquisition method that generates a single file containing fragment ion spectra of all ionized species of a sample. The resulting datasets were searched using a targeted data analysis strategy in which an *a priori* spectral reference library representing known N-glycosites of the human proteome was used to identify groups of signals in the SWATH mass spectrometry data. On average we identified 1430 N-glycosites from each sample. Out of those, 220 glycoproteins showed significant quantitative changes associated with diverse biological processes involved in PCa aggressiveness and metastasis and indicated functional relationships. Two glycoproteins, N-acylethanolamine acid amidase and protein tyrosine kinase 7, that were significantly associated with aggressive PCa in the initial sample cohort were further validated in an independent set of patient tissues using tissue microarray analysis. The results suggest that N-acylethanolamine acid amidase and protein tyrosine kinase 7 may be used

as potential tissue biomarkers to avoid overtreatment of non-aggressive PCa. *Molecular & Cellular Proteomics* 13: 10.1074/mcp.M114.038273, 1753–1768, 2014.

Prostate cancer (PCa)¹ is the most common noncutaneous cancer and the second leading cause of cancer-related death in men in the United States (1). Most diagnosed cases represent slow-growing, non-lethal forms of cancer. Unfortunately, neither the currently available diagnostic biomarkers, such as serum prostate-specific antigen (PSA), nor histological examination of (biopsied) tumor tissue can distinguish aggressive (AG) PCa from non-aggressive (NAG) PCa. This situation leads to the undertreatment of AG PCa and, more important, the overtreatment of NAG PCa (2). In fact, up to 90% of men with PCa harbor localized tumors that are unlikely to cause significant symptoms or mortality. Of these, many are overtreated because of a lack of clear (molecular) indicators that guide physicians to the appropriate treatment. All available treatment options, including surgery, radiation therapy, and hormonal therapy, carry a risk of complications and show a range of side effects that impact the patient's long-term quality of life. There is, therefore, a pressing clinical need to identify new PCa biomarkers in clinical tissue or blood that distinguish AG from NAG prostate tumors.

PCa tissue samples (e.g. those obtained from biopsies) are routinely subjected to histopathological examination, and the

From the [‡]Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, 8093 Zurich, Switzerland; [¶]Department of Pathology, Johns Hopkins University, Baltimore, Maryland 21231; ^{**}Faculty of Science, University of Zurich, 8057 Zurich, Switzerland

✂ Author's Choice—Final version full access.

Received February 1, 2014, and in revised form, April 4, 2014

Published, MCP Papers in Press, April 16, 2014, DOI 10.1074/mcp.M114.038273

Author contributions: Y.L., J.C., H.Z., and R.A. designed research; Y.L., J.C., L.C., and L.C.G. performed research; A.S. and B.W. contributed new reagents or analytic tools; Y.L., J.C., A.S., Q.K.L., and B.C. analyzed data; Y.L., J.C., and R.A. wrote the paper.

¹ The abbreviations used are: PCa, prostate cancer; AG, aggressive; NAG, non-aggressive; FDR, false discovery rate; GO, Gene Ontology; RFIN, Reactome Functional Interaction Network; IHC, immunohistochemistry; MS, mass spectrometry; MS/MS, tandem mass spectrometry; SWATH-MS, SWATH mass spectrometry; SRM, selected reaction monitoring; LC, liquid chromatography; TOF, time-of-flight; *m/z*, mass-to-charge; SPEG, solid phase extraction of formerly N-glycosylated peptides; *de-N*-glycopeptide, *de*-glycosylated N-glycopeptide treated by PNGase F; PSM, peptide spectrum match; PSA, prostate-specific antigen; TMA, tissue microarray; NAAA, N-acylethanolamine acid amidase; PTK7, protein tyrosine kinase 7; ROC, receiver operating characteristic; DDA, data-dependent acquisition.

results are reported by a Gleason score, a grading score ranging from 2 to 10 that is calculated by adding the score of the predominant grade pattern and that of the second most common grade pattern in a specific sample. The Gleason score helps guide patient treatment, but sometimes it fails to do so sufficiently because it cannot be used to distinguish significant molecular heterogeneities of PCa and a range of clinical trajectories (3). For example, the clinical outcome is unpredictable for most Gleason 7 PCas (4, 5). Molecular-level phenotyping has been proposed as a means to develop a more highly resolving scoring system capable of correctly classifying clinically important PCa types. In principle, clinical samples can be phenotyped by different types of measurements (e.g. genomic (6), epigenomic (7), transcriptomic (3), metabolomic (8), and proteomic (5)). To date, transcript profiling has been used most extensively, mainly because of the relatively advanced maturity and accessibility of the respective measurement techniques (9). However, proteomic measurements should be equally or more informative, because proteins are more dynamic and diverse and more directly reflective of cellular physiology than nucleic-acid-based markers (10). Moreover, PSA and other approved protein markers (11) exemplify the potential information contents of proteins.

The glycoproteome represents a subproteome that is particularly relevant for clinical research because glycoproteins are usually found on the cell surface or secreted by tissues and are more likely to be detected in the blood stream as non-invasive biomarkers (12–17). In fact, all current blood tumor biomarkers, including PSA in the case of PCa, that are approved by the U.S. Food and Drug Administration are glycoproteins (14). We previously developed a protocol for the solid phase extraction of glycopeptides (SPEG) to robustly isolate the glycoproteome based on chemical immobilization and enzymatic release of *N*-linked glycopeptides with high specificity (18, 19) that was thereafter successfully applied in different cancer biomarker discovery studies (18, 20–25).

In this study we used SPEG to profile the *N*-glycoproteome of PCa histotypes to identify glycoproteins associated with tumor aggressiveness. The isolated *de-N*-glycopeptide samples from prostate tissues were analyzed via the recently developed SWATH mass spectrometry (SWATH-MS) technology. SWATH-MS is a data-independent acquisition method (26) that essentially allows one to convert all the peptides ionized from a clinical sample into a perpetually reusable digital map (27). When combined with a targeted data analysis strategy, SWATH-MS was demonstrated to achieve the favorable accuracy, dynamic range, and reproducibility of selected reaction monitoring (SRM), the gold-standard quantitative proteomic technology, while greatly extending the degree of multiplexing to thousands of peptides (26, 28, 29). We have recently demonstrated that the combination of SWATH-MS and *de-N*-glycopeptide isolation has promising quantitative performance for biomarker verification in human

plasma (28). Here we establish that this integrated technology is also highly efficient for “molecular phenotyping” of tissue specimens because, once acquired, the quantitative data files representing control and disease-affected human tissues support iterative *in silico* biomarker discovery. To facilitate the targeted analysis of SWATH maps, we generated a spectral library covering a large part of the human *N*-glycoproteome, specifically optimized for SWATH-MS analysis. This library will also provide the community with a high-quality set of reference assays for future MS analyses of the global human *N*-glycoproteome and for related clinical applications. Furthermore, our SWATH dataset led to the identification of regulated proteins and pathways that might serve a predictive role in discriminating AG and NAG PCAs.

EXPERIMENTAL PROCEDURES

Materials—Hydrazide resin was from Bio-Rad (Hercules, CA); the BCA protein assay kit, Zeba spin desalting column (7k molecular weight cut off), urea, and tris (2-carboxyethyl) phosphine were from Thermo Fisher Scientific (Waltham, MA); sequencing-grade trypsin was from Promega (Madison, WI); PNGase F was from New England Biolabs (Ipswich, MA); and monoclonal mouse anti-NAAA and anti-PTK7 primary antibody was from R&D Systems (Minneapolis, MN). All other chemicals were from Sigma-Aldrich (St. Louis, MO).

Clinical Samples—Samples and clinical information were obtained with informed consent, and procedures were performed with the approval of the Institutional Review Board of the Johns Hopkins University. NAG and AG primary prostate tumors were collected via radical prostatectomy or transurethral resection of the prostate at Johns Hopkins Hospital and Johns Hopkins Bayview Medical Center under the National Cancer Institute-funded Johns Hopkins prostate cancer SPORE project. The NAG PCa group included 22 tumor specimens with Gleason scores of 6 and 2 tumor specimens from tumors with Gleason scores of 7 with no evidence of recurrence in up to 15 years of follow-up. The AG PCa group included 11 tumor specimens with Gleason scores of 8 or 9 and 5 tumor specimens with Gleason scores of 7 from patients who either died of cancer metastasis within 6 years of surgery or were positive for metastatic tumor at the time of surgery (supplemental Table S1). The 25 metastatic tumors were from men who died of PCa and underwent autopsy as part of the Project to Eliminate Lethal Prostate Cancer rapid-autopsy program of the Johns Hopkins Autopsy Study of Lethal Prostate Cancer, initiated in 1994. All subjects underwent androgen deprivation during the course of their treatment. The 10 normal prostate tissues were from healthy transplant donors who died from accidents or suicide. The primary prostate tumor tissues were immediately frozen after resection from surgery. The normal prostate tissues were immediately frozen after resections from transplant donors. The metastatic tumor tissues were acquired via rapid autopsy (a few hours to a day after death). All specimens were snap-frozen, embedded in optimal cutting temperature compound, and stored at -80°C until use.

Isolation of *de-N*-glycopeptides and Sample Preparation—Frozen prostate tissues embedded in optimal cutting temperature compound were sectioned and stained with hematoxylin and eosin (H&E). The H&E staining was used to guide cryostat microdissection for enrichment of the tumor content of tissue. After cryostat microdissection, 6- μm -thick tissue sections for each specimen were collected in sterile screw-cap bullet tubes. Proteins were extracted using cell lysis buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100). BCA assay was performed, and 100 μg of total protein mass per specimen was used to extract formerly *N*-

linked glycopeptides via the SPEG procedure as described previously (20, 22). Briefly, the proteins were alkylated and digested into peptides, which were cleaned up by means of C18 chromatography prior to SPEG. The peptides were treated with sodium periodate to oxidize the glycan moieties of glycopeptides and purified by G-10 gel filtration cartridges (Nest Group Inc., Southborough, MA). The sample was then conjugated to Affi-gel Hydrazine resin (Bio-Rad) overnight. The unbound peptides were removed through an extensive washing procedure. *N*-linked glycopeptides were released by PNGase F. Finally, *de-N*-glycopeptides were used for downstream MS analysis. To generate a SWATH spectral library and to quantify the *de-N*-glycopeptides from each tissue group, equal amounts of peptide samples from each tissue group were pooled together and analyzed via LC-MS/MS. In addition, small-scale sample pools (from five individual samples each) were generated for non-aggressive, aggressive, and metastatic prostate tumors. Eleven retention time anchor peptides (iRT peptides, Biognosys AG, Zurich, Switzerland (30)) were added into each sample at a ratio of 1:30 v/v. For each SWATH analysis, equal amounts of sample (estimated to be roughly 1 μ g of total peptide mass) derived from pooled tissue were analyzed so that a meaningful comparison could be achieved across groups.

SWATH-MS Measurement—SWATH-MS datasets (or SWATH maps) were acquired using an AB Sciex 5600 TripleTOF mass spectrometer (Concord, Ontario, Canada) interfaced to an Eksigent NanoLC Ultra 2D Plus HPLC system (Dublin, CA) as previously described (26, 28, 29). Peptides were directly injected onto a 20-cm PicoFrit emitter (New Objective, self-packed to 20 cm with Magic C18 AQ 3- μ m, 200- \AA material) and then separated using a 120-min gradient of 2% to 35% buffer (buffer A: 0.1% (v/v) formic acid, 2% (v/v) acetonitrile; buffer B: 0.1% (v/v) formic acid, 90% (v/v) acetonitrile) at a flow rate of 300 nl/min. In SWATH-MS mode, the instrument was specifically tuned to optimize the quadrupole settings for the selection of precursor ion selection windows 25 m/z wide. Using an isolation width of 26 m/z (containing 1 m/z for the window overlap), a set of 32 overlapping windows was constructed covering the precursor mass range of 400–1200 m/z . The effective isolation windows can be considered as 399.5–424.5, 424.5–449.5, etc. SWATH MS2 spectra were collected from 100 to 2000 m/z . The collision energy was optimized for each window according to the calculation for a charge 2+ ion centered upon the window with a spread of 15 eV. An accumulation time (dwell time) of 100 ms was used for all fragment-ion scans in high-sensitivity mode, and for each SWATH-MS cycle a survey scan in high-resolution mode was also acquired for 100 ms, resulting in a duty cycle of \sim 3.4 s.

Shotgun Measurement of Isolated *de-N*-glycopeptides from PCa Samples—For shotgun acquisition, peptides from subpools of each PCa group were firstly measured on an Oribtrap XL (Thermo Scientific) to check the sample quality in collision-induced dissociation mode, and then *de-N*-glycopeptides from all four tissue groups were pooled equally as a super-mixture and analyzed using classical shotgun data acquisition on a TripleTOF 5600 instrument via four injection replicates. For measurements on the Oribtrap XL, a 90-min gradient was used for each sample using the acquisition method published previously (31). For shotgun MS/MS on the TripleTOF, the same chromatographic system and settings as described above for SWATH-MS were used. MS1 spectra were collected in the range of 360–1460 m/z for 250 ms. The 20 most intense precursors with charge states of 2 to 5 that exceeded 250 counts per second were selected for fragmentation, and MS2 spectra were collected in the range of 50–2000 m/z for 100 ms. The precursor ions were dynamically excluded from reselection for 20 s.

Shotgun Measurement of Synthetic Peptides for the Generation of an *N*-glycoprotein SWATH Atlas—We previously published an SRM assay library for 2007 human *N*-glycosylated proteins (*N*-glycoprotein

SRM Atlas) for targeted proteomic analysis (31). In that work the SRM assays were generated mainly by SRM-triggered MS2 acquisition on a QTrap instrument. For this study, all human synthetic peptides from the SRM assay library (31) were re-acquired for spectral library generation, but this time using the shotgun mode on the 5600 mass spectrometer. Basically, the peptide selection sources were the *N*-glycosites identified in large discovery-driven MS-based experiments in diverse human tissues, cell lines, and plasma and the *N*-glycosites that were selected from the UniProt database (13, 31), with additional peptide targets from recent shotgun datasets (28), yielding an identified protein list with a total of 2460 glycoproteins with the high sensitivity of the 5600 mass spectrometer. Peptides were synthesized using SPOT-synthesis technology (JPT Peptide Tech, Berlin, Germany) (32). About 800 peptides were mixed with iRT peptides together for each separate shotgun analysis using the 5600 instrument, with the same LC-MS settings described above. Shotgun measurements were repeated for certain plate-pools to maximize peptide coverage.

Spectral Library Generation—Profile-mode .wiff files from shotgun data acquisition were converted to mzML files in centroided format using AB Sciex Data Converter v.1.3 (default parameter) and then further converted to mzXML files using MSConvert v.3.04.238. The MS2 spectra were queried against the canonical Swiss-Prot complete proteome database for human (November 2012) appended with common contaminants, iRT peptide sequences, and the corresponding reversed sequence decoys (33) (40,951 protein sequences including decoys). The SEQUEST database search (34) through Sorcerer PE version 4.2 included the following criteria: trypsin as digestion enzyme; semi-tryptic peptides and peptides with up to two missed cleavages were allowed; static modifications of 57.02146 Da for cysteines; variable modifications of 15.99491 Da for methionine oxidations; and variable modifications of 0.98406 Da for asparagines (formerly *N*-glycosylated asparagines are converted to aspartic acids upon PNGase F treatment). The mass tolerances of the monoisotopic parent and fragment ions were set as 50 ppm. The identified peptides were processed and analyzed using Trans-Proteomic Pipeline 4.5.2 (35), and search results were validated using the PeptideProphet score (36). *N*-glycosylation motif information was used in PeptideProphet. For the synthetic peptides, the database for shotgun searching was generated by concatenating peptides plus scrambled pseudo protein sequences (31). The SEQUEST database searching parameters were identical to those for the shotgun analysis of the biological sample, except for the specifications of fully tryptic digestion and up to one miss cleavage. All the peptides were filtered at a false discovery rate (FDR) of 1%, as estimated by PeptideProphet at the peptide spectrum match (PSM) level (33). Peptides mapped to redundant protein identities were excluded for SWATH-MS quantification at the protein level.

The raw spectral libraries were generated from all valid PSMs in shotgun experiments of both natural samples and synthetic peptides and then refined into the nonredundant consensus libraries (29) using SpectraST (37). To generate the spectral library for SWATH-MS, the highest priority was given to the shotgun spectra acquired from the shotgun analysis on the TripleTOF 5600, because those fragment ion spectra most closely resembled those generated in SWATH-MS. This means that the spectra of unique shotgun identifications from the Oribtrap XL analyses were only accepted if the corresponding peptides were not identified with the TripleTOF. For each peptide, the retention time was mapped into the iRT space (30) with reference to a linear calibration constructed for each shotgun run, as previously described (29). The MS assays constructed from the top six most intense fragments with a Q3 range from 400 to 1200 m/z , excluding those falling in the precursor SWATH window, were used for targeted data analysis of SWATH maps.

Targeted Data Analysis for SWATH Maps—SWATH-MS .wiff files were first converted to mzXML profile using ProteoWizard (38). As described previously, the SWATH targeted data analysis was carried out using OpenSWATH (39) running on an internal computing cluster. OpenSWATH automatically integrates peak group extraction and a decoy scoring system using mProphet (40) to estimate FDR. Based on the generated spectral libraries, OpenSWATH identified the peak groups from the SWATH maps at FDR = 1% and aligned them between SWATH maps from different samples based on the clustering behaviors of the retention time in each run. Specifically, features were considered for alignment based on a nonlinear alignment algorithm (41) with a maximum FDR quality of 0.25 (quality cutoff to still consider a feature for alignment) and/or the further constraint of a retention time difference of less than 60 s in LC gradient after iRT normalization.

Statistics and Functional Annotation—The peak intensities of unique peptides were reported by OpenSWATH for label-free quantification. First, a simple global normalization based on the total intensity was done for each sample (42). Hierarchical clustering analysis was done at the *N*-glycosite level using Cluster 3.0 on the log-transformed, two-dimensional centered and normalized peptide intensities (43). The hierarchical clustering analysis result was visualized with Treeview (44). Next, to quantify the protein abundances across samples, we summed up the most abundant identified peptides for each protein (top three if more than three peptides were identified). This proxy allows one to reliably estimate global protein abundance changes, as shown in previous studies (23, 28, 41, 45). Peptides identified and aligned in <80% of samples were discarded at this step for quantitative glycoproteomic profiling. Principal component analysis was performed by R software using protein intensities. The candidates between PCa subtypes were prioritized by analysis of variance executed on Pomelo II (46), which also reported the differential glycoprotein list between AG and NAG groups. The GO cellular component classification was done by the BINGO 2.44 plugin in Cytoscape (47, 48). The annotation of biological pathways and functional processes was done using the DAVID bioinformatics resource (49), and the enrichment analysis was performed by taking the entire *N*-glycoprotein SWATHAtlas as background. The protein list in each category and the enrichment *p* values were then downloaded, manually compiled, and visualized by Cytoscape (48). SignalP 4.1 (50) and TMHMM 2.0 prediction (51) were used to predict the existence of signal peptides and transmembrane helices in a protein sequence for their possible involvement in classical secretion pathways or in integral membrane structures, with these protein features visualized by Protter (52). The receiver operating characteristic (ROC) curve analysis was done by PanelComposer (53), with values of the area under the ROC curve provided for individual proteins and the combine panel using a logistic regression model.

Network Analysis of the Relationship between Protein Sets and Public Genomic Data—The Reactome Functional Interaction Network (RFIN) (54, 55) was used to investigate the functional relationships between regulated glycoproteins and altered genes. We manually compiled a list of 351 altered genes from seven genomic, transcriptomic, and epigenomic studies of prostate cancer (6, 7, 56–60). These genes were found to be significantly mutated, bi-allelically mutated, epigenetically silenced, or recurrently fused or to have insertions and deletions in the previous studies. The statistical significance of the functional relationships between regulated glycoproteins and altered genes was assessed in RFIN and its 100 random instances. These random instances in the RFIN were created using a “switching algorithm” (61) in which the interaction partners of the nodes are randomized. The resulting network is of the same size and degree distribution as the original network, and it preserves the degree of each node. We used the switching algorithm implemented in

the Random Network Plugin for Cytoscape (55), and the network graphs were visualized in Cytoscape (48).

Immunoassay Measurements of PSA in Clinical Specimens—Proteins used for immunoassay measurements of PSA from tissue specimens were the same protein extracts used for glycopeptide isolation. After protein extracts had been adjusted to 1 μ g/ml with PBS, tissue PSA levels were measured with an Access® Hybritech PSA assay (Beckman Coulter, Inc., Brea, CA) in a lab certified by Clinical laboratory improvement amendments at Johns Hopkins Medical Institutions.

Immunohistochemical Staining and Scoring—Staining was performed on formalin-fixed paraffin-embedded prostate tissue slides from six individuals with primary prostate tumors, as well as on a 336-core prostate tissue microarray (TMA). The TMA contained 56 cases of primary PCa. Each case included four cores of the tumor regions and two cores of the matching adjacent normal prostate tissues. IHC staining was performed as previously described (62, 63). Briefly, slides with sections of tissues or TMA were deparaffinized and rehydrated. Tissue slides were incubated in antigen retrieval buffer at 92 °C to 95 °C for 10 min. Tissues were blocked by peroxidase block and 3% BSA/PBS for 30 min each followed by 2.5% horse serum blocking for 15 min at room temperature. The tissues were then incubated with a monoclonal mouse anti-NAAA or monoclonal mouse anti-PTK7 primary antibody in antibody dilution buffer at 50 μ g/ml or 10 μ g/ml, respectively, followed by incubation with anti-mouse antibody labeled polymer-HRP for 30 min each. The staining was detected with DAB chromogen. The intensity of NAAA and PTK7 was visually graded by a board-certified pathologist as 0 (no staining), 1 (weak staining), 2 (medium staining), or 3 (strong staining) in the epithelial compartments. A total of 56 cases (tumors and normal tissues) were scored. The Gleason scores of the tumor cores included in the TMA are listed in supplemental Table S2. IHC score differences were calculated by subtracting the IHC score of the matching adjacent normal cores from the tumor cores. The Wilcoxon–Mann–Whitney rank sum test was used to calculate the statistical significance between normal and tumor tissues, as well as that between tumors with Gleason scores less than or equal to 3 + 4 and tumors with Gleason scores greater than or equal to 4 + 3.

Data Availability—The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository (64) with the dataset identifier PXD000704.

RESULTS

Identification of Glycoproteomic Signals in SWATH Maps of PCa Tumor Tissues—The experimental scheme of the present study is shown in Fig. 1. It principally combined SPEG isolation of *de-N*-glycopeptides from selected tissue specimens with SWATH-MS to discover glycoprotein candidates; the candidates were subsequently validated by means of TMA analysis. Overall, in this study we analyzed *de-N*-glycopeptide samples from 10 cases of normal prostate tissue, 24 cases of NAG PCa tissue, 16 cases of AG PCa tissue, and 25 cases of metastatic tumor samples. The level of aggressiveness was defined based on the Gleason score and the evidence of recurrence within up to 15 years (see “Experimental Procedures”). All PCa samples selected had a high tumor cell percentage (~70%; supplemental Table S1). To gain sufficient starting material for the glycoproteomic analysis, and to focus on the general difference between normal prostate and the three different PCa types, samples were initially analyzed

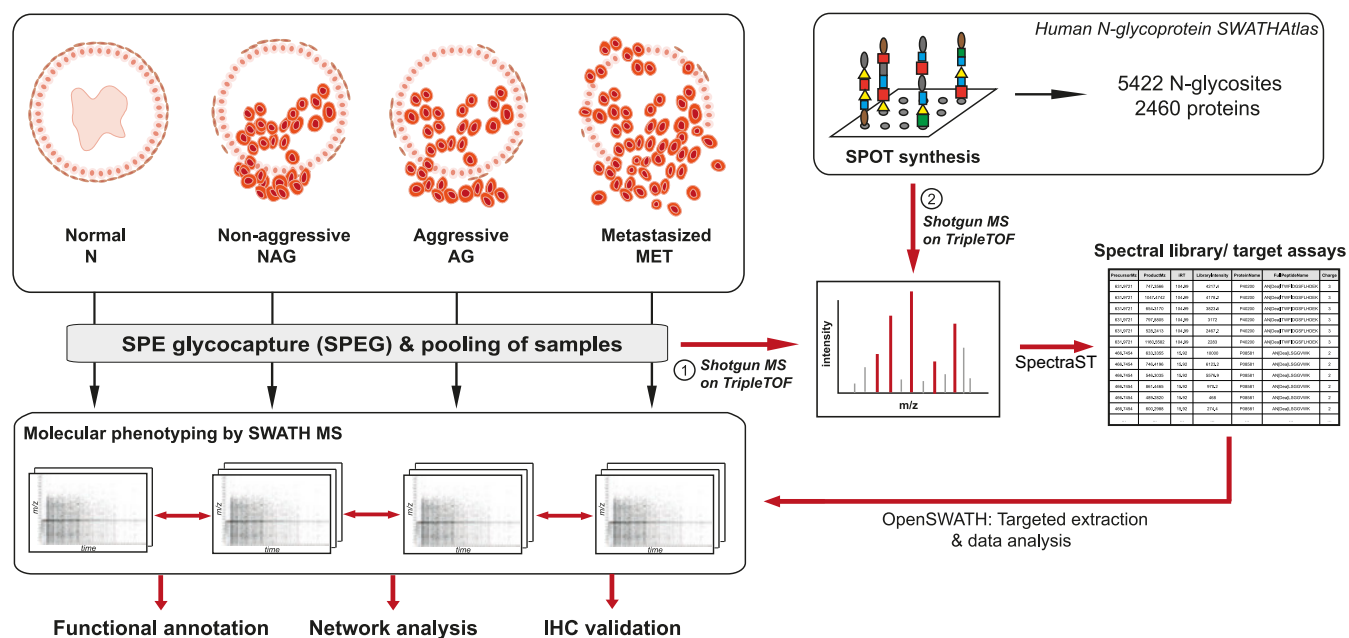


FIG. 1. **Experimental design.** Formerly N-glycosylated peptides of well-characterized PCa tissue samples were isolated via solid phase extraction of glycopeptides (SPEG) and subjected to SWATH-MS. A spectral library for the targeted identification and quantification of specific N-glycosites from the SWATH maps was generated via shotgun sequencing of the *de-N*-glycopeptides from clinical samples and synthetic reference peptides. The in-house-developed OpenSWATH software was used to identify and quantify peaks in SWATH maps, and quantification was followed by bioinformatic analysis and immunohistochemistry (IHC)-based tissue microarray validation.

by equally pooling all samples from the same tissue group for the determination of a quantitative expression ratio between different tissue groups, based on total intensity normalization. Additional samples pooled with a smaller sample size from each tissue group were then analyzed to determine the *p* values of each glycoprotein between groups (supplemental Table S1).

The targeted identification of peptides in SWATH-MS datasets (or SWATH maps) requires *a priori* generation of a spectral library that includes essential coordinates such as precursor ion masses, fragment ion masses, fragment ion intensities, and retention times for each targeted peptide (27). We thus firstly generated a spectral library through shotgun sequencing of the isolated *de-N*-glycopeptides from same prostate samples. Glycopeptide pools were analyzed using classical shotgun data acquisition on Orbitrap XL and TripleTOF 5600 instruments (see “Experimental Procedures”). A total of 1919 *N*-glycosites assigned to 548 distinct glycoproteins were identified in all the shotgun experiments (see data supplement and supplemental Tables S3 and S4 for annotated MS2 spectra and identification summary on peptide and protein levels). We then used these fragment ion spectra to build a spectral library. Using this library, we identified and quantified 882 to 1296 *N*-glycosites in the SWATH maps generated from the clinical samples at an estimated FDR of 1% (Fig. 2A).

To augment the spectral library with coordinates of *N*-glycosites that might be present in the PCa samples but not detected via the data-dependent acquisition (DDA) of *de-N*-

glycopeptides from PCa tissues, in part because of the issue of stochastic precursor-ion selection in DDA (27, 65, 66), we built a more comprehensive spectral library by generating fragment ion spectra for all synthetic peptides from the human *N*-glycoprotein SRMatlas (31) using a 5600 mass spectrometer operated in DDA mode. This resulted in reference fragment ion spectra under conditions that mimicked those encountered in SWATH-MS, thus eliminating instrument bias (26). These data resulted in the generation of a synthetic consensus spectral library containing 5422 *N*-glycosites of 2460 human glycoproteins, covering nearly 50% of the human proteins that are annotated in the UniProt database as glycoproteins. This *N*-glycoprotein SWATHAtlas library is the *de facto* model 5600 version of *N*-glycoAtlas. The coordinates that constitute a definitive mass spectrometric assay for each identified *N*-glycosite are provided in EXCEL format for future applications in supplemental Table S5. Intriguingly, by using the spectral library augmented with the spectra from synthetic peptides, we were able to detect 218 to 414 additional *N*-glycosites from the individual PCa SWATH maps, compared with the peptides identified using the tissue DDA library alone (Fig. 2A). With these approximately 30% additional peptide identifications, the augmented library achieved significantly deeper analysis of the clinical samples than the library generated from tissue DDA data only, demonstrating that SWATH-MS analysis consistently identified and quantified *N*-glycosites that were difficult to detect with DDA analysis of the same samples (27, 29). These data also illustrate the opportunity offered by SWATH-MS to iteratively reanalyze the

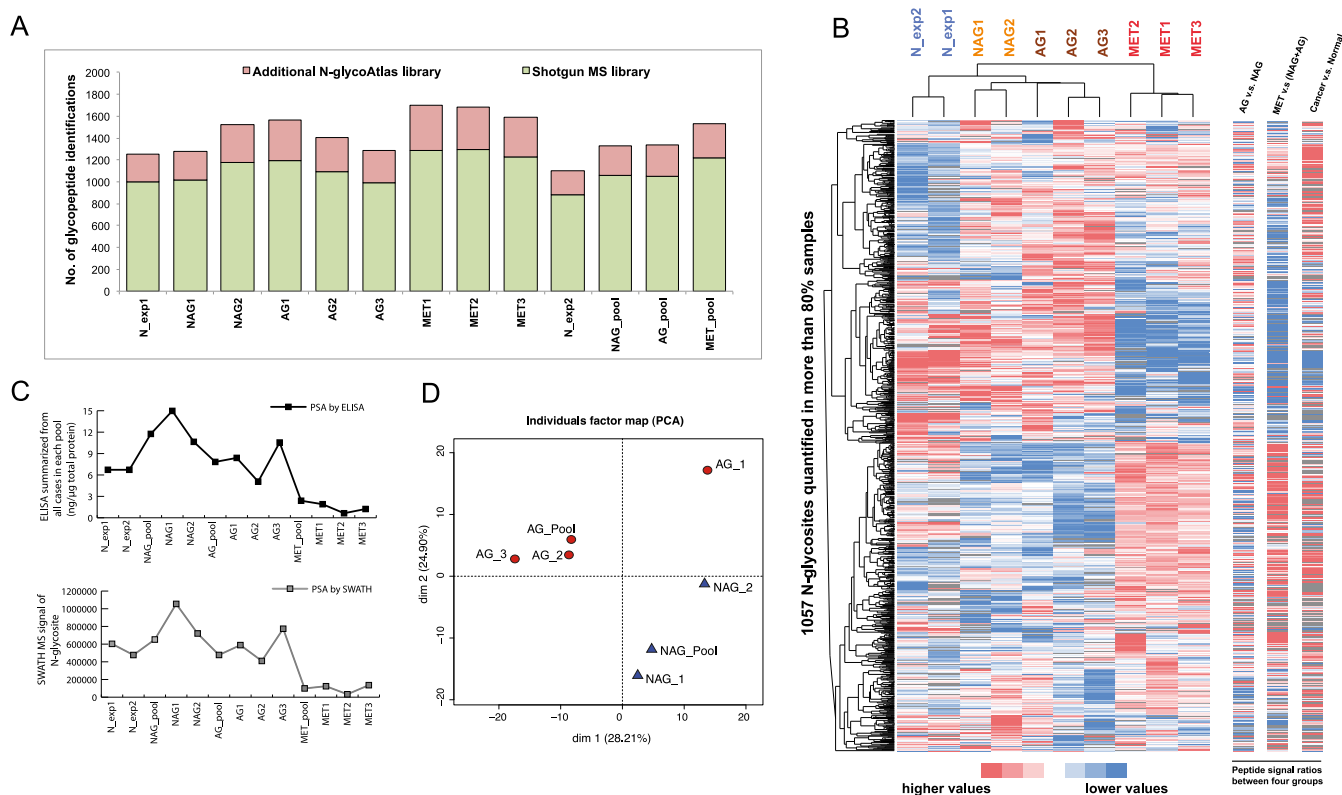


FIG. 2. Quantitative profiling of tissue N-glycoproteome between different PCa groups. *A*, numbers of N-glycosites identified from SWATH maps. Note that identification of *de-N*-glycopeptides using the SWATHAtlas library from synthetic reference peptides increased the number of identified peptides by approximately 30%. *B*, hierarchical clustering analysis of 1057 N-glycosites quantified among >80% of the samples. *C*, the accuracy of SWATH-MS quantification was revealed by the example of PSA that was also measured by the tissue ELISA. *D*, principle component analysis of NAG and AG cases tested.

same digital datasets with improved or alternative spectral libraries.

After combining the data from the tissue-derived and synthetic peptide libraries, we identified 2188 *N*-glycosites corresponding to 897 *N*-glycoproteins from all SWATH maps (supplemental Table S6) and detected on average 1430 *N*-glycosites in each sample, presenting a much deeper and more consistent glycoproteomic survey than previous studies (22, 24, 66).

Quantitative Profiling by SWATH-MS Reveals Distinct N-glycoproteomic Signatures Associated with PCa Aggressiveness—To assess the technical quality of the glycoproteomic data with respect to reproducibility and quantitative accuracy, we compared datasets from the same samples acquired at different time points and validated some data points with an orthogonal method. Pooled *N*-glycosites from normal tissue (N_exp1 & 2) were analyzed 3 months apart via SWATH-MS. The integrated peak areas of the *N*-glycosites detected in the repeat analyses highly correlated with each other ($r = 0.919$) (Fig. 2B and supplemental Fig. S1). To validate the quantification accuracy of SWATH-MS with an orthogonal method, we measured the PSA levels of all individual tissue specimens using an Access® hybritech PSA assay and compared the values to the relative peak intensities of PSA in

respective glycoproteomic SWATH maps. Fig. 2C indicates that ELISA results and SWATH signals were strongly concordant ($r = 0.973$). We conclude from these data that the quantitative values obtained via SWATH-MS accurately reflect real glycoprotein abundance changes in the clinical samples tested.

To analyze the correlation of the glycoproteome patterns of three different types of PCa and normal prostate tissues, we performed an unsupervised hierarchical clustering analysis of the samples. In total, 1057 *N*-glycosites were quantified in at least 8 out of 10 SWATH maps (Fig. 2B). The cluster graph shows that samples of the same subtype clustered together tightly. Moreover, we noticed that the overall glycoproteomes were markedly distinct among different PCa subtypes, with the most significant alteration observed between metastatic and nonmetastatic tumors. In contrast, the separation of the AG and NAG groups was moderate. Principal component analysis of the four samples of AG PCa and the three samples of NAG PCa consistently demonstrated that the AG and NAG groups could be separated by two principal components (Fig. 2D). Overall, these results position the dataset generated in this study as a valuable resource for further analysis of the glycoproteome as an indicator of the clinical behavior of various PCa subtypes.

Identification of Specific Glycoproteins as Candidates Distinguishing Non-aggressive, Aggressive, and Metastatic PCa—We employed analysis of variance to associate the expression patterns of specific glycoproteins with the normal, AG, NAG, and metastatic groups. Of the consistently quantified proteins, 220 proteins were differentially expressed across the groups with high significance (Table I) ($p < 0.1$; p values and fold changes summarized in supplemental Table S7). Among these, 50 proteins were significantly altered between the AG and NAG groups. Extensive literature searching uncovered evidence that 125 (*i.e.* 56.8%) of the 220 significantly altered proteins are directly related to prostate cancer or the level of aggressiveness and metastatic potential of other human cancers (supplemental Table S7). Two-thirds of the proteins (*i.e.* 142 out of 220) showed clear positive staining in $>25\%$ of the prostate cancer tissue specimens documented in the Human Protein Atlas database (supplemental Table S7) (67). Moreover, the list contains clinically applied PCa biomarkers, including PSA and prostatic acid phosphatase (68, 69).

Differentially expressed tissue glycoproteins are particularly interesting as biomarker candidates because of the high likelihood of their detectability in blood plasma. Using a bioinformatic approach, we assessed the potential detectability of our candidates in blood plasma. The program SignalP predicted 145 of the 220 (65.9%) proteins as classical secretory proteins (50) TMHMM predicted 124 of 220 (56.4%) proteins as transmembrane (51). In all, 199 of 220 proteins (90.5%) were predicted as secretory or transmembrane proteins (supplemental Table S7). These values predict a high likelihood of detecting these proteins remotely in the blood stream, and the probability is even greater than that determined from secretome studies in which conditioned medium of cultured cancer cells was analyzed (70, 71). Indeed, our list covered 25 out of 54 PCa biomarker candidates that were discovered from a previous glycoproteomic analysis of a phosphatase and tensin homolog conditional knockout mouse model (supplemental Table S7) (23). Even more intriguingly, 17 of these 25 glycoproteins were found to be detectable in human serum via SRM (Table I) (23). These include seven proteins (*i.e.* metalloproteinase inhibitor 1, attractin, asporin, cell adhesion molecule 1, biotinidase, hypoxia up-regulated protein 1, and neural cell adhesion molecule 1) that were finally verified as blood biomarkers predictive for diagnosis or grading of PCa (23). Furthermore, serum cell adhesion molecule 1, cathepsin F, and periostin were found in another study as prognostic markers for survival when serum from 57 metastatic castration-resistant prostate cancer patients was studied (72). A direct comparison of these 220 proteins to the human plasma PeptideAtlas database (73) suggests that 125 proteins were previously observed in human plasma in large-scale shotgun experiments derived from PeptideAtlas. Accordingly, $>75\%$ of the proteins mapped to PeptideAtlas are expressed at a concentration range of <100 ng/ml in human plasma (supplemental Fig. S2), indicating the strong relevance of this list for

biomarker discovery, as lots of tissue leakage products are found in blood at this concentration level (16, 28).

Table I lists representative biomarkers that were reported as diagnostic and prognostic biomarkers according to the literature, with a focus on the differentially expressed proteins between AG and NAG group and their detectability in human blood.

Annotation of Glycoproteins Significantly Changed in PCa Subtypes with Altered Pathways—We next investigated biological pathways and processes of the 220 significantly altered glycoproteins by functional GO annotation. The cellular component distribution (Fig. 3A) agreed with the prediction results above, as the majority of proteins were annotated to reside either in the plasma membrane (45.4%) or in the extracellular compartment (36.3%). The GO biological processes and PANTHER pathways associated with the altered protein set were displayed in a network with corresponding proteins (Fig. 3B) so that we could manually remove the redundant GO items annotated by the same proteins. Taking the *N*-glycoprotein SWATHAtlas as a comparison background, we found several functionally interplayed processes enriched as clusters. Notably, one cluster was related to lysosome-based proteolysis, and a second was associated with the cell adhesion process ($p < 0.01$ in all the processes; Fig. 3B). The processes of protein maturation by peptide cleavage and acute inflammatory response were also enriched. These interlaced biological processes might be helpful in synoptically delineating the molecular variability of PCa subtypes.

Reactome Network Analysis Indicates That Glycoproteomic Regulation Is Closely Connected to PCa Genomic Events—Cancer proteomic studies sometimes primarily reflect inflammation processes and acute phase responses that coincide with the primary disease, owing to the difficulty of achieving substantial analytical depth (74). Therefore, to explore whether the regulated glycoproteins identified in this study were directly associated with PCa, we related the altered protein set to the genes found altered in previous PCa genomic studies on a functional level (6, 7, 56–60). We utilized RFIN to investigate functional relationships between regulated glycoproteins and altered genes. Of the 351 altered genes, 184 were found in the RFIN. Using the RFIN network and 100 randomized instances of the same size and degree distribution, we determined that the sets of altered genes and the 220 differentially expressed glycoproteins (from analysis of variance) were significantly interconnected ($p = 0.0099$; supplemental Fig. S3). Further, the altered glycoproteins between AG and NAG groups were also found to be functionally related to the altered genes known in PCas, such as P53, PI3K, and AR mutations ($p = 0.021$; Fig. 4). The RFIN result supports the notion that our tissue glycoproteomic investigation was rather deep and correlates well with genomic events commonly encountered in PCa with statistical significance (supplemental Fig. S4), although future studies are needed to

Glycoprotein Markers for Prostate Cancer Aggressiveness

TABLE I

Selected representative biomarkers associated with differential prostate cancer subtypes. This table lists biomarkers that showed strong evidence as diagnostic and prognostic PCa biomarkers according to previous literature, with a focus on cancer aggressiveness and previous blood detection

Protein (human)	UniProt	Protein name	P value (ANOVA)	Fold change	Positive staining, % in PCa ^a	SignalP	TMHMM	PCa markers detected in blood ^b
Higher in AG/lower in NAG								
POSTN	Q15063	Periostin	0.0685	1.59	0	Y	N	Y
ASPN	Q9BXN1	Asporin	0.0754	1.44	63	Y	N	Y
LAMB2	P55268	Laminin subunit β -2	0.0002	2.1	26	Y	N	–
SERPH	P50454	Serpin H1	0.0024	1.38	39	Y	N	–
CSPG2	P13611	Versican core protein	0.0443	2.22	41	Y	N	–
ENTP1	P49961	Ectonucleoside triphosphate diphosphohydrolase 1	0.0102	1.96	42	N	Y	–
SE1L1	Q9UBV2	Protein sel-1 homolog 1	0.0341	1.67	50	Y	Y	–
ITAV	P06756	Integrin α -V	0.0435	1.54	86	Y	Y	–
FOLH1	Q04609	Glutamate carboxypeptidase 2	0.0708	6.34	94	N	Y	–
STIM1	Q13586	Stromal interaction molecule 1	0.0377	1.29	96	Y	Y	–
PTK7	Q13308	Inactive tyrosine-protein kinase 7	0.0004	1.2	100	Y	Y	–
ICOSL	O75144	ICOS ligand	0.0094	1.29	100	Y	Y	–
SPP2A	Q8TCT8	Signal peptide peptidase-like 2A	0.067	1.33	100	Y	Y	–
CD276	Q5ZPR3	CD276 antigen	0.0888	1.68	100	Y	Y	–
Higher in NAG/lower in AG								
KLK3	P07288	Prostate-specific antigen	0.0032	1.37	90	Y	N	Y
ZA2G	P25311	Zinc- α -2-glycoprotein	0.0005	3.23	95	Y	N	Y
FBN1	P35555	Fibrillin-1	0.0028	1.28	13	Y	N	–
CD38	P28907	ADP-ribosyl cyclase 1	0.0003	2.50	30	N	Y	–
CNTP2	Q9UHC6	Contactin-associated protein-like 2	0.0068	1.61	42	Y	Y	–
AMPN	P15144	Aminopeptidase N	0.0017	2.70	72	N	Y	–
FAM3B	P58499	Protein FAM3B	0.0286	1.72	72	N	Y	–
RNT2	O00584	Ribonuclease T2	0.0138	2.63	89	Y	N	–
NAAA	Q02083	N-acyl ethanolamine-hydrolyzing acid amidase	0.043	2.13	91	Y	Y	–
PPAP	P15309	Prostatic acid phosphatase	0.0001	1.56	100	N	Y	–
GSLG1	Q92896	Golgi apparatus protein 1	0.0042	1.27	100	Y	Y	–
TSN1	O60635	Tetraspanin-1	0.0707	10.00	NA	N	Y	–
DPP4	P27487	Dipeptidyl peptidase 4	0.0107	2.27	NA	N	Y	–
ST14	Q9Y5Y6	Suppressor of tumorigenicity 14 protein	0.0384	1.64	NA	N	Y	–
Higher in non-MET/lower in MET								
TIMP1	P01033	Metalloproteinase inhibitor 1	0.0101	2.70	25	Y	N	Y
CLUS	P10909	Clusterin	0.0184	2.08	20	Y	N	Y
MFAP4	P55083	Microfibril-associated glycoprotein 4	0.0634	3.70	30	Y	N	Y
AOC3	Q16853	Membrane primary amine oxidase	0	3.45	42	N	Y	Y
CADM1	Q9BY67	Cell adhesion molecule 1	0.046	2.63	60	Y	Y	Y
HYOU1	Q9Y4L1	Hypoxia up-regulated protein 1	0.0019	1.85	9	Y	Y	Y
FSTL1	Q12841	Follistatin-related protein 1	0.0063	2.50	73	Y	N	Y
KLK11	Q9UBX7	Kallikrein-11	0.0294	8.33	25	N	N	–
NBL1	P41271	Neuroblastoma suppressor of tumorigenicity 1	0.0216	4.17	70	Y	N	–
GOLM1	Q8NBJ4	Golgi membrane protein 1	0.0012	3.70	100	N	Y	–
Higher in MET/lower in non-MET								
VTNC	P04004	Vitronectin	0.0324	1.85	NA	Y	N	Y
CERU	P00450	Ceruloplasmin	0.0008	2.53	65	Y	N	Y
A1AG2	P19652	α -1-acid glycoprotein 2	0.0325	2.78	50	Y	N	–
LYVE1	Q9Y5Y7	Lymphatic vessel endothelial hyaluronic acid receptor 1	0.031	3.87	0	Y	Y	–
A1AG1	P02763	α -1-acid glycoprotein 1	0.0004	8.65	30	Y	N	–
Higher in cancer/lower in normal								
ATRNL1	O75882	Attractin	0.0683	1.28	100	N	Y	Y
CATD	P07339	Cathepsin D	0.0112	2.19	54	Y	N	Y
BTD	P43251	Biotinidase	0.0265	1.96	NA	N	N	Y
MTA1	Q13330	Metastasis-associated protein MTA1	0.0134	>10	25	N	N	–
Higher in normal/lower in cancer								
NCAM1	P13591	Neural cell adhesion molecule 1	0.0001	2.17	0	Y	Y	Y
PECA1	P16284	Platelet endothelial cell adhesion molecule	0.0634	1.56	NA	Y	Y	–

ANOVA, analysis of variance; MET, metastatic.

^a The percentages of the positive antibody staining among PCa tissue specimens were manually compiled from Protein Atlas.

^b The detection in blood for glycoproteins as PCa markers was based on Refs. 23 and 72.

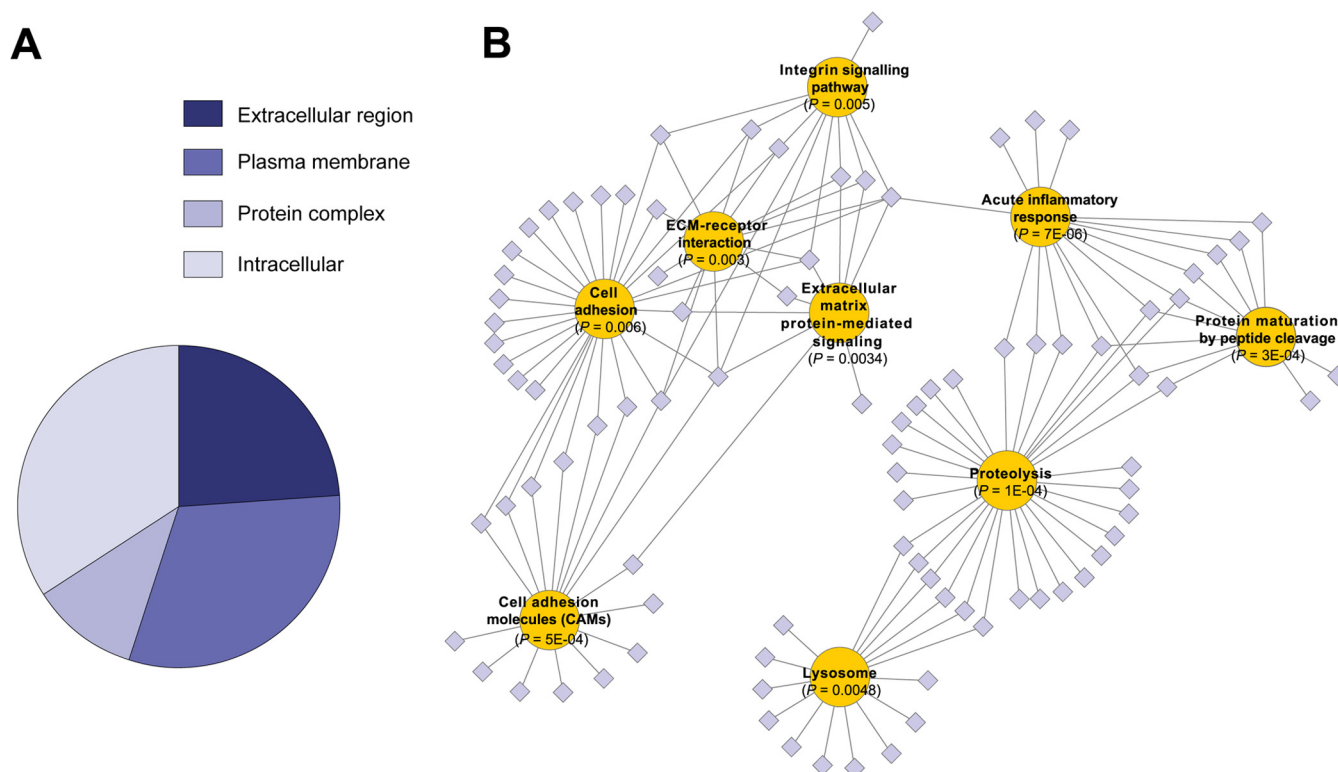


FIG. 3. **Functional annotation of significantly regulated glycoproteins ($n = 220$) between PCa groups.** A, the GO cellular component distribution. B, the pathways and biological processes enriched in the 220-protein list, according to DAVID functional annotation.

establish a causative role between proteomic and genomic observations.

Tissue Microarray Verification of NAAA and PTK7 as Signatures of PCa Aggressiveness—We next pursued TMA analysis to further validate the biomarkers of PCa aggressiveness. We selected NAAA and PTK7 in this phase based on the following considerations: 1. In the differential *N*-glycosite analysis described above, NAAA and PTK7 stood out as proteins with significantly altered expression between NAG and AG tumors. NAAA expression was significantly higher in the NAG group, whereas PTK7 was higher in AG cases (Table I and supplemental Fig. S5). 2. Both proteins showed prevalent positive staining (>90%) in PCa specimens according to the human Protein Atlas (67). 3. Both were predicted as secretory proteins (supplemental Fig. S5). 4. To the best of our knowledge, they have not been tested as tissue markers for PCa. We thus carried out preliminary IHC staining on six cases of PCa tissues. For both proteins, staining was primarily observed in the epithelial compartment, and increased expression of NAAA and PTK7 was detected in the tumor epithelium relative to normal epithelium (supplemental Fig. S6). In addition, whereas NAAA showed medium to strong staining in Gleason grade 3 tumors (Figs. 5A and 5B), the staining was rather faint in Gleason grade 4 tumors. In contrast, PTK7 staining was more intense in Gleason grade 4 tumors than in Gleason grade 3 tumors (Figs. 5C–5E).

To further evaluate NAAA and PTK7 expression in prostate tissues, TMA analysis was performed with 56 prostate adenocarcinoma cases. Each case contained four cores of tumor and two cores of patient-matched adjacent normal tissue (336 cores in total; supplemental Table S2). Among the 56 cases, 21 (37.5%) had a Gleason score of 3 + 3; 4 (7.1%) had a Gleason score of 3 + 4; 10 (17.9%) had a Gleason score of 4 + 3; 11 (19.6%) had a Gleason score of 4 + 4; 4 (7.1%) had a Gleason score of 4 + 5; 5 (8.9%) had a Gleason score of 5 + 4; and 1 (1.8%) had a Gleason score of 5 + 5. The prostate tumor cases were classified into two categories, with 25 cases (44.6%) of Gleason 3 + 4 or less and 31 cases (55.4%) of Gleason 4 + 3 and above (Table II). NAAA and PTK7 staining intensities were then evaluated and scored by a board-certified pathologist in the epithelial compartment. We firstly confirmed the different expression of these two proteins between the matching adjacent normal tissue and tumor tissues. Figs. 5F and 5G show histograms of NAAA and PTK7 IHC score differences; 66.1% and 55.4% of the cases showed an increase in NAAA and PTK7 staining, respectively, in tumors relative to normal tissues (Wilcoxon–Mann–Whitney rank sum test; $p < 0.001$ for both proteins). Most important, whereas the majority (75%) of the adjacent normal tissues stained negative or weakly for NAAA, 54.8% of the tumors (Gleason score $\geq 4 + 3$) and 88% of the tumors (Gleason score $\leq 3 + 4$) stained with medium to strong intensity. In

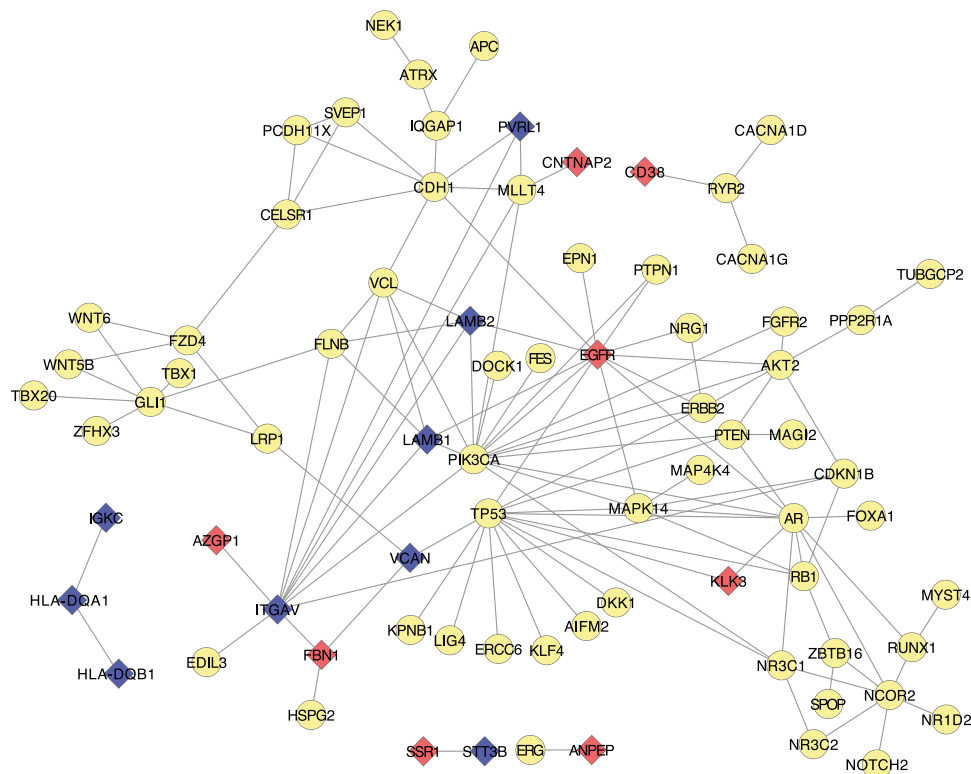


FIG. 4. An RFIN subnetwork of functional interactions between regulated glycoproteins (AG versus NAG) and genes commonly mutated in PCa. Red diamonds denote proteins up-regulated in NAG, and blue diamonds denote glycoproteins up-regulated in AG. Altered genes are shown by yellow circles. Edges denote functional relationships between the nodes they connect. Regulated glycoproteins and altered genes not connected to each other are not shown in the network. Relative to randomized RFIN networks of the same size and same connectivity, the altered PCa genes and our regulated glycoproteins are significantly interconnected to each other ($p = 0.021$).

contrast, 94.6% of the adjacent normal tissues stained negatively or weakly for PTK7, whereas 28% of the tumors (Gleason score $\leq 3 + 4$) and 38.7% of the tumors (Gleason score $\geq 4 + 3$) stained with medium to strong intensities (Table II). Relative to expression in tumors with a Gleason score equal to or less than $3 + 4$, NAAA expression was significantly decreased and PTK7 expression was significantly increased ($p < 0.05$; Fig. 5H) in tumors with a Gleason score equal to or more than $4 + 3$. From the ROC analysis based on TMA staining intensities in the tumor specimens, we found that values of the area under the ROC curve of 0.743 and 0.709 were achieved by NAAA and PTK7 expression, respectively, both with statistical significance. Interestingly, a combined panel of these two proteins showed greater predictive power (area under the curve = 0.801) than each of them alone, indicating a potential power of the two proteins combined to discriminate AG and NAG.

DISCUSSION

The molecular mechanisms of systemic diseases such as cancer are still poorly understood. Currently in clinics, disease symptoms are recorded, for example, through x-ray images, computed tomography scans, and bio-fluid test results. More accurate records of high fidelity, particularly those indicating

molecular-level differences between healthy and disease-affected cells, might be helpful in understanding disease biology and for directing optimal treatment. Because proteomic technologies have lagged behind genomic technologies in terms of throughput, sensitivity, and reproducibility, most molecular analyses of disease tissues have been carried out at the genomic or transcriptomic level, even though it is widely acknowledged that proteomic measurements reflect the functional state of a tissue more closely than genomic patterns.

In this study we used an emerging proteomic method, SWATH-MS, to quantify the N-glycoproteome of a sample cohort consisting of normal tissue as well as non-aggressive, aggressive, and metastatic PCa tissue samples with the intent to identify proteins capable of orchestrating and distinguishing these groups. In contrast to the more widely used DDA methods, in which specific precursor ions are selected from the pool of available precursors, in SWATH-MS and other data-independent acquisition methods, all the precursor ions generated by the mass spectrometer from a particular sample are fragmented, and the fragment ions are recorded in the form of convoluted composite fragment ion spectral maps. In essence, SWATH-MS acquires a complete and permanent digital record for all the detectable components of a sample (26, 27) and is therefore an appealing technique for the anal-

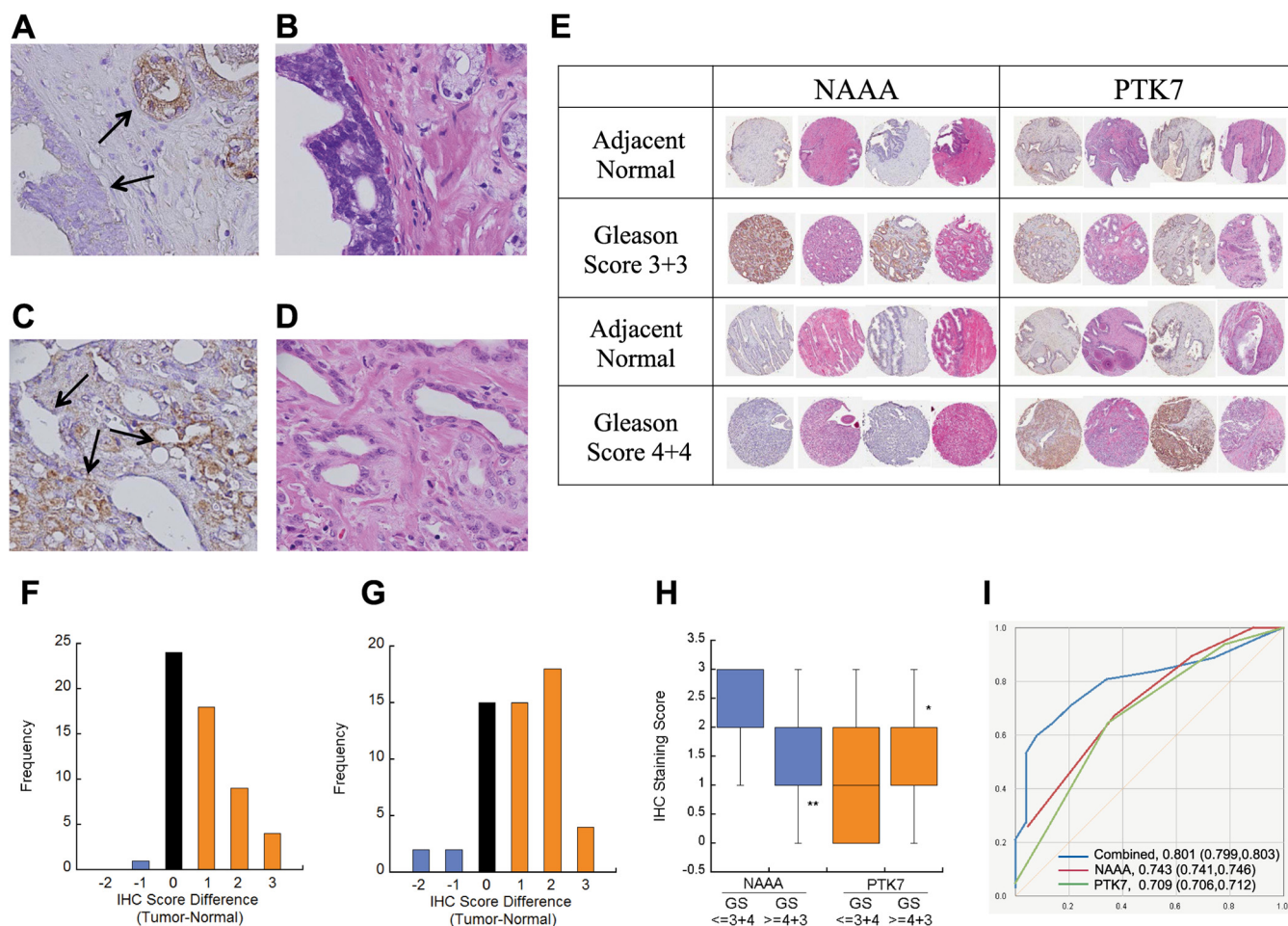


FIG. 5. TMA analysis for NAAA and PTK7. A, B, representative NAAA staining showing strong staining in Gleason score 3 tumors and faint staining in Gleason score 4 tumors. Faint or no staining was observed in adjacent normal tissues. C, D, representative PTK7 staining showing strong staining in Gleason score 4 tumors and less staining in Gleason score 3 tumors. Faint staining was observed in adjacent normal tissues. E, representative NAAA and PTK7 staining in TMA with corresponding H&E staining. F, histogram of the difference in IHC score between tumor and its matched adjacent normal tissue for NAAA. G, histogram of the difference in IHC score between tumor and its matched adjacent normal tissue for PTK7. Wilcoxon signal rank order test (paired, two-sided) was performed for NAAA and PTK7 between tumors and matched adjacent normal tissues ($p < 0.0001$). H, a box-plot was generated for NAAA and PTK7 between tumors with a Gleason score less than or equal to 3 + 4 tumor and tumors with a Gleason score greater than or equal to 4 + 3 tumor. * $p < 0.05$; ** $p < 0.01$. I, ROC analysis of NAAA and PTK7 based on their IHC staining intensities in the tumor specimens.

TABLE II

Epithelium IHC scoring of NAAA and PTK7. Immunohistochemical staining was performed for N-acyl ethanolamine acid amidase (NAAA) and protein tyrosine kinase 7 (PTK7) on tissue microarray (TMA) with 56 cases of prostate cancer tissues. From each case, four cores of tumor tissues and two cores of adjacent normal prostate tissues were included in the TMA. The numbers of cases with pathologist-read epithelium IHC scores of 0, 1, 2, and 3 are listed for adjacent normal tissue and tumor tissues with a Gleason score less than or equal to 3 + 4 and more than or equal to 4 + 3 for both NAAA and PTK7

Epithelium IHC score	NAAA				PTK7			
	0	1	2	3	0	1	2	3
Adjacent normal	22 (39.3%)	20 (35.7%)	9 (16.1%)	5 (8.9%)	38 (67.9%)	15 (26.8%)	3 (5.4%)	0 (0%)
Gleason score $\leq 3 + 4$	0 (0%)	3 (12%)	8 (32%)	14 (56%)	10 (40%)	8 (32%)	6 (24%)	1 (4%)
Gleason Score $\geq 4 + 3$	6 (19.4%)	8 (25.8%)	10 (32.3%)	7 (22.6%)	2 (6.5%)	17 (54.8%)	8 (25.8%)	4 (12.9%)

ysis of unique, nonrenewable clinical samples (e.g. biopsied tissues). In this study, SWATH-MS was applied to the analysis of glycoproteins because this subproteome was assumed to be enriched for potential biomarkers (13, 16, 17). The data

demonstrate that the combination of SPEP and SWATH-MS achieved a high degree of reproducibility for quantifying glycoproteins (Fig. 2) and correlated well with clinical quantitation of PSA. This is consistent with our previous observation

that SWATH-MS allows equal variability ($cv < 20\%$) and similar accuracy relative to SRM (28).

Because SWATH-MS generates composite fragment ion spectra of multiple precursor ions that are concurrently fragmented, we implemented a targeted data analysis strategy to reliably identify and quantify specific peptides from the acquired datasets. The strategy requires the generation of a high-quality spectral library as prior information for the quantification of peptides from SWATH-MS datasets (27). In essence, an ensemble of fragment ion coordinates, including the m/z ratio of fragment ions, their relative signal intensity, the elution patterns of the fragment ion signals, etc., is extracted from the spectral library and used to confidently identify a target peptide in the SWATH-MS map using a dedicated search tool such as OpenSWATH (39). To support this type of data analysis for the *N*-glycoproteome, we generated an *N*-glycoprotein SWATHatlas derived from synthetic peptides as an easily transferrable resource that will also support future targeted proteomic studies focused on the *N*-glycoproteome. The assays from the *N*-glycoprotein SWATHatlas provided an unprecedented chance to identify and quantify 5422 *N*-glycosites of 2460 glycoproteins that essentially cover $\sim 50\%$ of the annotated human *N*-glycoproteome. The use of these assays helped us to retrieve 30% more identifications in every SWATH map as an extra benefit. More important, this increase highlights the possibility of data re-mining in the one-time acquired SWATH maps when there are more MS assays available for novel protein targets. This data re-mining feature is a unique advantage of SWATH-MS over traditional DDA or SRM approaches.

It should be noted that the addition of unique shotgun identifications from an instrument other than that used for the SWATH-MS analysis (e.g. Orbitrap XL-derived spectra added to a library generated on a 5600 instrument) is not ideal because the fragment ion spectra are less similar across instrument types than within instrument types (75). In this study, because of the limited amount of peptide available from the microdissected clinical samples, we nevertheless extended the 5600 instrument-derived spectral library with spectra generated on an Orbitrap XL instrument. The direct transferability for targeted proteomics of peptide fragmentation data across instrument platforms and the outcome were assessed in a recent study (75) in which Toprak *et al.* reported that TripleTOF shotgun peptide fragmentation provided the most similar fragmentations to that observed with SWATH-MS targeted data analysis, compared with other shotgun solutions (e.g. those provided by an Orbitrap instrument). Therefore, we again recommend that one acquire, when possible, instrument-specific libraries or specifically tune the fragmentation parameters of the MS instrument available.

We previously determined the limit of detection of SWATH-MS in human plasma *N*-glycoproteome at 5 ng/ml, a level close to the PSA concentration in blood of PCa patients (28). Here,

substantial analytical depth of the tissue glycoproteome was also achieved in glycoproteomic SWATH analysis, which uncovered pathways (Fig. 3) associated with PCa progression including cell adhesion (76), integrin signaling (77), and proteolysis (78), indicating possible linkage between glycoproteomic regulations and common genomic mutations in PCa (Fig. 4). These results shed more light on the disease biology related to different PCa types, which might be interesting for future investigations. Although we deployed a greedy literature search for each putative biomarker candidate from the discovery phase (supplemental Table S7), defining the mechanism of their relationship with prostate cancer aggressiveness and metastasis is beyond the scope of this study. Nevertheless, our SWATH dataset nicely overlapped with already described PCa markers and, again, allows for the re-quantification and verification of newly discovered biomarkers or proteins of interest at any time in the future.

To summarize the above discussions, the combination of SPEG and SWATH-MS provides a reproducible, deep, and quantitative glycoproteomic reference that can be easily reexamined and therefore provides a significant advance for biomarker studies. The high specificity (75% to 90%) of SPEG was repeatedly tested in many previous studies (18–25). Here, according to our shotgun data, we found that only 2.6% of the identified *de-N*-glycopeptides can present a natural, detectable un-deamidated form, demonstrating a practically high specificity of glycoproteomic survey by our combined method. In the instrument configuration for SWATH-MS used in this study, 1 to 3 μg of total peptide mass represented an optimal sample load. This translates into a need for about 10^5 to 10^6 cells for proteomic profiling and at least 10^7 cells for the glycoproteomic survey by SWATH analysis following SPEG enrichment (19, 79). Thus, a more efficient protein extraction strategy and a specific SPEG protocol (80), both optimized for small amounts of clinical samples such as those derived from microdissected tissues or biopsied samples, make an important contribution toward the reproducible glycoproteomic phenotyping of clinical samples via SWATH-MS.

Two glycoproteins, NAAA and PTK7, were for the first time reported and verified as novel potential biomarkers for PCa aggressiveness. NAAA is a lysosomal enzyme that degrades bioactive fatty acid amides to their corresponding acids, with primary reference to the anti-inflammatory substance *N*-palmitoylethanolamine. NAAA also exhibits weak hydrolytic activity against ceramides *N*-lauroylsphingosine and *N*-palmitoylsphingosine (81). It has been reported that NAAA is present in cell lines derived from human blood cells (82). Wang and coworkers found that among human tissues, the prostate showed the highest NAAA mRNA level (83). In addition, IHC staining images from the human Protein Atlas also showed stronger staining with NAAA antibody relative to all the other 19 types of human cancers (67). Moreover, Wang *et al.* also reported that NAAA is functionally active in PCa cells and is released as a secretory protein. Interest-

ingly, they illustrated that the NAAA expression in LNCap cells, a PCa cell line that is responsive to androgen stimulation, is higher than that in PC-3 and DU-145 cells, which are androgen-insensitive cells representing the androgen-refractory phase of advanced PCa (83). In this study, we showed that NAAA has a significantly lower expression in PCa with a Gleason score $\geq 4 + 3$, which for the first time established NAAA not only as a potential PCa biomarker, but also as a promising signature for tumor aggressiveness. However, the physiological role of lower NAAA expression in high-grade PCa tissues is unclear and requires further studies.

PTK7, also known as colon carcinoma kinase-4, is an inactive tyrosine kinase involved in various processes. PTK7 is an essential component of the Wnt/planar cell polarity pathway that controls tissue polarity and cell movement. Unlike NAAA, PTK7 has been previously annotated as a cancer-associated protein. In previous reports, the expression of PTK7 has been shown to be increased in leukemia cell lines (84), acute myeloid leukemia specimens (85), esophageal squamous cell carcinoma tissues (86), and colorectal carcinomas (87). Also, PTK7 was reported as a potential prognosis biomarker for a number of cancers, such as gastric cancer (88), lung adenocarcinoma (89), and triple-negative breast cancer treated with chemotherapy (90). In the case of gastric cancer, positive staining of PTK7 was noted in 114 of 201 tissue samples (88). However, until now PTK7 expression has not been reported to be associated with prostate cancer. We found that PTK7 was dysregulated in PCa tumors relative to the matched adjacent normal tissues. More important, PTK7 was up-regulated in aggressive PCa tumors relative to non-aggressive PCa. The exact functional role of PTK in PCa progression remains to be determined.

We observed a certain degree of synergy in applying these two markers for identifying NAG cases (Figs. 5H and 5I). It therefore would be very interesting to seek the possibility of establishing a biomarker panel based on these two markers in future studies to avoid overtreatment of NAG PCa. Because both NAAA and PTK7 were predicted as secretory proteins (Table I), we envision further investigations into their detectability in human plasma and their discriminating power between PCa cases in blood. Also, the relationship of their expression with PCa metastasis needs to be ascertained. TMA analyses (56 cases, 336 cores) were performed for NAAA and PTK7, and the expressions of both proteins were significantly altered between AG and NAG samples. These data suggest that NAAA and PTK7 might be potential IHC markers for staging prostate cancer. However, larger scaled TMA analyses with different cohorts need to be done to validate their clinical utility, because of the practical difficulty of discriminating AG and NAG PCa (as we note in Fig. 2). All the newly identified markers for PCa aggressiveness have to be validated in prospective studies in the future.

Acknowledgments—We thank G. Steven Bova for curating and defining the frozen prostate tumor specimens; Lorenz Blum and

Hannes Roest for the help in SWATH data processing; Ruth Hüttenhain and Brunner Roland for the sample preparation of synthetic glycopeptides; Ralph Schiess, H. Alexander Ehardt, and Peter Wild for their helpful comments on the study; and Oliver M. Bernhardt and Lukas Reiter (Spectronaut developers from Biognosys CH) for helpful discussions. We thank the PRIDE team for data deposition.

* This work was supported by NCI, National Institutes of Health; the Early Detection Research Network (EDRN, U01CA152813); the Clinical Proteomic Tumor Analysis Consortium (CPTAC, U24CA160036); NHLBI, National Institutes of Health; Johns Hopkins Proteomics Center (N01-HV-00240); Programs of Excellence in Glycosciences (PEG, P01HL107153); the Swiss National Science Foundation (Grant No. 310030B 147086); SystemsX.ch (PhosphoNetPPM); the European Research Council (FP7 Grant No. 233226); and PRIME-XS (EC).

§ This article contains supplemental material.

|| To whom correspondence should be addressed: Hui Zhang, Department of Pathology, Clinical Chemistry Division, Cancer Research Building II, Johns Hopkins University, 1550 Orleans Street, Room 3M03, Baltimore, MD 21231, Fax: 1-443-287-6388, E-mail: zhhang32@jhmi.edu; Ruedi Aebersold, Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Wolfgang-Pauli-Str.16, 8093 Zurich, Switzerland, Tel.: 41-44-633-3170, Fax: 41-44-633-1051, E-mail: aebersold@imsb.biol.ethz.ch.

§ These authors contributed to this work equally.

REFERENCES

1. Siegel, R., Naishadham, D., and Jemal, A. (2013) Cancer statistics, 2013. *CA* **63**, 11–30
2. Andriole, G. L., Crawford, E. D., Grubb, R. L., 3rd, Buys, S. S., Chia, D., Church, T. R., Fouad, M. N., Gelmann, E. P., Kvale, P. A., Reding, D. J., Weissfeld, J. L., Yokochi, L. A., O'Brien, B., Clapp, J. D., Rathmell, J. M., Riley, T. L., Hayes, R. B., Kramer, B. S., Izmirlian, G., Miller, A. B., Pinsky, P. F., Prorok, P. C., Gohagan, J. K., and Berg, C. D. (2009) Mortality results from a randomized prostate-cancer screening trial. *N. Engl. J. Med.* **360**, 1310–1319
3. Lapointe, J., Li, C., Higgins, J. P., van de Rijn, M., Bair, E., Montgomery, K., Ferrari, M., Egevad, L., Rayford, W., Bergerheim, U., Ekman, P., DeMarzo, A. M., Tibshirani, R., Botstein, D., Brown, P. O., Brooks, J. D., and Pollack, J. R. (2004) Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 811–816
4. Sakr, W. A., Tefilli, M. V., Grignon, D. J., Banerjee, M., Dey, J., Gheiler, E. L., Tiguert, R., Powell, I. J., and Wood, D. P. (2000) Gleason score 7 prostate cancer: a heterogeneous entity? Correlation with pathologic parameters and disease-free survival. *Urology* **56**, 730–734
5. Pin, E., Fredolini, C., and Petricoin, E. F., 3rd (2013) The role of proteomics in prostate cancer research: biomarker discovery and validation. *Clin. Biochem.* **46**, 524–538
6. Berger, M. F., Lawrence, M. S., Demichelis, F., Drier, Y., Cibulskis, K., Sivachenko, A. Y., Sboner, A., Esgueva, R., Pflueger, D., Sougnez, C., Onofrio, R., Carter, S. L., Park, K., Habegger, L., Ambrogio, L., Fennell, T., Parkin, M., Saksena, G., Voet, D., Ramos, A. H., Pugh, T. J., Wilkinson, J., Fisher, S., Winckler, W., Mahan, S., Ardlie, K., Baldwin, J., Simons, J. W., Kitabayashi, N., MacDonald, T. Y., Kantoff, P. W., Chin, L., Gabriel, S. B., Gerstein, M. B., Golub, T. R., Meyerson, M., Tewari, A., Lander, E. S., Getz, G., Rubin, M. A., and Garraway, L. A. (2011) The genomic complexity of primary human prostate cancer. *Nature* **470**, 214–220
7. Lin, P. C., Giannopoulou, E. G., Park, K., Mosquera, J. M., Sboner, A., Tewari, A. K., Garraway, L. A., Beltran, H., Rubin, M. A., and Elemento, O. (2013) Epigenomic alterations in localized and advanced prostate cancer. *Neoplasia* **15**, 373–383
8. McDunn, J. E., Li, Z., Adam, K. P., Neri, B. P., Wolfert, R. L., Milburn, M. V., Lotan, Y., and Wheeler, T. M. (2013) Metabolomic signatures of aggressive prostate cancer. *Prostate* **73**, 1547–1560
9. Soon, W. W., Hariharan, M., and Snyder, M. P. (2013) High-throughput

- sequencing for biology and medicine. *Mol. Syst. Biol.* **9**, 640
10. Aebersold, R., Anderson, L., Caprioli, R., Druker, B., Hartwell, L., and Smith, R. (2005) Perspective: a program to improve protein biomarker discovery for cancer. *J. Proteome Res.* **4**, 1104–1109
 11. Gutman, S., and Kessler, L. G. (2006) The US Food and Drug Administration perspective on cancer biomarker development. *Nat. Rev. Cancer* **6**, 565–571
 12. Zhang, H., Yan, W., and Aebersold, R. (2004) Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Curr. Opin. Chem. Biol.* **8**, 66–75
 13. Zhang, H., Loriaux, P., Eng, J., Campbell, D., Keller, A., Moss, P., Bonneau, R., Zhang, N., Zhou, Y., Wollscheid, B., Cooke, K., Yi, E. C., Lee, H., Peskind, E. R., Zhang, J., Smith, R. D., and Aebersold, R. (2006) UniPep, a database for human N-linked glycosites: a resource for biomarker discovery. *Genome Biol.* **7**, R73
 14. Zhang, H., and Chan, D. W. (2007) Cancer biomarker discovery in plasma using a tissue-targeted proteomic approach. *Cancer Epidemiol. Biomarkers Prev.* **16**, 1915–1917
 15. Roth, J. (2002) Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. *Chem. Rev.* **102**, 285–303
 16. Schiess, R., Wollscheid, B., and Aebersold, R. (2009) Targeted proteomic strategy for clinical biomarker discovery. *Mol. Oncol.* **3**, 33–44
 17. Zhang, H., Liu, A. Y., Loriaux, P., Wollscheid, B., Zhou, Y., Watts, J. D., and Aebersold, R. (2007) Mass spectrometric detection of tissue proteins in plasma. *Mol. Cell. Proteomics* **6**, 64–71
 18. Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **21**, 660–666
 19. Tian, Y., Zhou, Y., Elliott, S., Aebersold, R., and Zhang, H. (2007) Solid-phase extraction of N-linked glycopeptides. *Nat. Protoc.* **2**, 334–339
 20. Chen, J., Xi, J., Tian, Y., Bova, G. S., and Zhang, H. (2013) Identification, prioritization, and evaluation of glycoproteins for aggressive prostate cancer using quantitative glycoproteomics and antibody-based assays on tissue specimens. *Proteomics* **13**, 2268–2277
 21. Zhao, J., Simeone, D. M., Heidt, D., Anderson, M. A., and Lubman, D. M. (2006) Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: application to pancreatic cancer serum. *J. Proteome Res.* **5**, 1792–1802
 22. Tian, Y., Bova, G. S., and Zhang, H. (2011) Quantitative glycoproteomic analysis of optimal cutting temperature-embedded frozen tissues identifying glycoproteins associated with aggressive prostate cancer. *Anal. Chem.* **83**, 7013–7019
 23. Cima, I., Schiess, R., Wild, P., Kaelin, M., Schuffler, P., Lange, V., Picotti, P., Ossola, R., Templeton, A., Schubert, O., Fuchs, T., Leippold, T., Wyler, S., Zehetner, J., Jochum, W., Buhmann, J., Cerny, T., Moch, H., Gilllessen, S., Aebersold, R., and Krek, W. (2011) Cancer genetics—guided discovery of serum biomarker signatures for diagnosis and prognosis of prostate cancer. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3342–3347
 24. Chen, R., Tan, Y., Wang, M., Wang, F., Yao, Z., Dong, L., Ye, M., Wang, H., and Zou, H. (2011) Development of glycoprotein capture-based label-free method for the high-throughput screening of differential glycoproteins in hepatocellular carcinoma. *Mol. Cell. Proteomics* **10**, M110.006445
 25. Zeng, X., Hood, B. L., Sun, M., Conrads, T. P., Day, R. S., Weissfeld, J. L., Siegfried, J. M., and Bigbee, W. L. (2010) Lung cancer serum biomarker discovery using glycoprotein capture and liquid chromatography mass spectrometry. *J. Proteome Res.* **9**, 6440–6449
 26. Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, O111.016717
 27. Liu, Y., Huttenhain, R., Collins, B., and Aebersold, R. (2013) Mass spectrometric protein maps for biomarker discovery and clinical research. *Expert Rev. Mol. Diagn.* **13**, 811–825
 28. Liu, Y., Huttenhain, R., Surinova, S., Gillet, L. C., Mouritsen, J., Brunner, R., Navarro, P., and Aebersold, R. (2013) Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. *Proteomics* **13**, 1247–1256
 29. Collins, B. C., Gillet, L. C., Rosenberger, G., Rost, H. L., Vichalkovski, A., Gstaiger, M., and Aebersold, R. (2013) Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14–3-3 system. *Nat. Methods* **10**, 1246–1253
 30. Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., MacCoss, M. J., and Rinner, O. (2012) Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **12**, 1111–1121
 31. Huttenhain, R., Surinova, S., Ossola, R., Sun, Z., Campbell, D., Cerciello, F., Schiess, R., Bausch-Fluck, D., Rosenberger, G., Chen, J., Rinner, O., Kusebauch, U., Hajdich, M., Moritz, R. L., Wollscheid, B., and Aebersold, R. (2013) N-glycoprotein SRMATlas: a resource of mass spectrometric assays for N-glycosites enabling consistent and multiplexed protein quantification for clinical applications. *Mol. Cell. Proteomics* **12**, 1005–1016
 32. Frank, R. (2002) The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports—principles and applications. *J. Immunol. Methods* **267**, 13–26
 33. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214
 34. Yates, J. R., 3rd, Eng, J. K., McCormack, A. L., and Schieltz, D. (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **67**, 1426–1436
 35. Keller, A., Eng, J., Zhang, N., Li, X. J., and Aebersold, R. (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.* **1**, 2005.0017
 36. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392
 37. Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K., King, N., Stein, S. E., and Aebersold, R. (2007) Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* **7**, 655–667
 38. Kessner, D., Chambers, M., Burke, R., Agus, D., and Mallick, P. (2008) ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534–2536
 39. Rost, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinovic, S. M., Schubert, O. T., Wolski, W., Collins, B. C., Malmstrom, J., Malmstrom, L., and Aebersold, R. (2014) OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat. Biotechnol.* **32**, 219–223
 40. Reiter, L., Rinner, O., Picotti, P., Huttenhain, R., Beck, M., Brusniak, M. Y., Hengartner, M. O., and Aebersold, R. (2011) mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat. Methods* **8**, 430–435
 41. Weisser, H., Nahnsen, S., Grossmann, J., Nilse, L., Quandt, A., Brauer, H., Sturm, M., Kenar, E., Kohlbacher, O., Aebersold, R., and Malmstrom, L. (2013) An automated pipeline for high-throughput label-free quantitative proteomics. *J. Proteome Res.* **12**, 1628–1644
 42. Quackenbush, J. (2002) Microarray data normalization and transformation. *Nat. Genet.* **32 Suppl**, 496–501
 43. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863–14868
 44. Juan, H. F., and Huang, H. C. (2007) Bioinformatics: microarray data clustering and functional classification. *Methods Mol. Biol.* **382**, 405–416
 45. Ludwig, C., Claassen, M., Schmidt, A., and Aebersold, R. (2012) Estimation of absolute protein quantities of unlabeled samples by selected reaction monitoring mass spectrometry. *Mol. Cell. Proteomics* **11**, M111.013987
 46. Morrissey, E. R., and Diaz-Uriarte, R. (2009) Pomelo II: finding differentially expressed genes. *Nucleic Acids Res.* **37**, W581–W586
 47. Maere, S., Heymans, K., and Kuiper, M. (2005) BINGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449
 48. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504
 49. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics re-

- sources. *Nat. Protoc.* **4**, 44–57
50. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786
 51. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580
 52. Omasits, U., Ahrens, C. H., Muller, S., and Wollscheid, B. (2013) Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* **30**, 884–886
 53. Jeong, S. K., Na, K., Kim, K. Y., Kim, H., and Paik, Y. K. (2012) PanelComposer: a web-based panel construction tool for multivariate analysis of disease biomarker candidates. *J. Proteome Res.* **11**, 6277–6281
 54. Wu, G., Feng, X., and Stein, L. (2010) A human functional protein interaction network and its application to cancer data analysis. *Genome Biol.* **11**, R53
 55. Haw, R., Hermjakob, H., D'Eustachio, P., and Stein, L. (2011) Reactome pathway analysis to enrich biological discovery in proteomics data sets. *Proteomics* **11**, 3598–3613
 56. Kumar, A., White, T. A., MacKenzie, A. P., Clegg, N., Lee, C., Dumpit, R. F., Coleman, I., Ng, S. B., Salipante, S. J., Rieder, M. J., Nickerson, D. A., Corey, E., Lange, P. H., Morrissey, C., Vessella, R. L., Nelson, P. S., and Shendure, J. (2011) Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17087–17092
 57. Weischenfeldt, J., Simon, R., Feuerbach, L., Schlangen, K., Weichenhan, D., Minner, S., Wuttig, D., Warnatz, H. J., Stehr, H., Rausch, T., Jager, N., Gu, L., Bogatyrova, O., Stutz, A. M., Claus, R., Eils, J., Eils, R., Gerhauser, C., Huang, P. H., Hutter, B., Kabbe, R., Lawrenz, C., Radomski, S., Bartholomae, C. C., Falth, M., Gade, S., Schmidt, M., Amschler, N., Hass, T., Galal, R., Gjoni, J., Kuner, R., Baer, C., Masser, S., von Kalle, C., Zichner, T., Benes, V., Raeder, B., Mader, M., Amstislavskiy, V., Avci, M., Lehrach, H., Parkhomchuk, D., Sultan, M., Burkhardt, L., Graefen, M., Huland, H., Kluth, M., Krohn, A., Sirma, H., Stumm, L., Steurer, S., Grupp, K., Sultmann, H., Sauter, G., Plass, C., Brors, B., Yaspo, M. L., Korb, J. O., and Schlomm, T. (2013) Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. *Cancer Cell* **23**, 159–170
 58. Kannan, K., Wang, L., Wang, J., Ittmann, M. M., Li, W., and Yen, L. (2011) Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9172–9177
 59. Barbieri, C. E., Baca, S. C., Lawrence, M. S., Demichelis, F., Blattner, M., Theurillat, J. P., White, T. A., Stojanov, P., Van Allen, E., Stransky, N., Nickerson, E., Chae, S. S., Boyesen, G., Auclair, D., Onofrio, R. C., Park, K., Kitabiyashi, N., MacDonald, T. Y., Sheikh, K., Vuong, T., Guiducci, C., Cibulskis, K., Sivachenko, A., Carter, S. L., Saksena, G., Voet, D., Hussain, W. M., Ramos, A. H., Winckler, W., Redman, M. C., Ardlie, K., Tewari, A. K., Mosquera, J. M., Rupp, N., Wild, P. J., Moch, H., Morrissey, C., Nelson, P. S., Kantoff, P. W., Gabriel, S. B., Golub, T. R., Meyerson, M., Lander, E. S., Getz, G., Rubin, M. A., and Garraway, L. A. (2012) Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **44**, 685–689
 60. Grasso, C. S., Wu, Y. M., Robinson, D. R., Cao, X., Dhanasekaran, S. M., Khan, A. P., Quist, M. J., Jing, X., Lonigro, R. J., Brenner, J. C., Asangani, I. A., Ateeq, B., Chun, S. Y., Siddiqui, J., Sam, L., Anstett, M., Mehra, R., Prensner, J. R., Palanisamy, N., Ryslik, G. A., Vandin, F., Raphael, B. J., Kunju, L. P., Rhodes, D. R., Pienta, K. J., Chinnaiyan, A. M., and Tomlins, S. A. (2012) The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243
 61. Milo, R., Kashtan, N., Itzkovitz, S., Newman, M. E. J., and Alon, U. (2003) On the uniform generation of random graphs with prescribed degree sequences. *Cond. Matter* **2**, 1–4
 62. Liu, Y., Luo, X., Hu, H., Wang, R., Sun, Y., Zeng, R., and Chen, H. (2012) Integrative proteomics and tissue microarray profiling indicate the association between overexpressed serum proteins and non-small cell lung cancer. *PLoS One* **7**, e51748
 63. Liu, Y., Li, C., Xing, Z., Yuan, X., Wu, Y., Xu, M., Tu, K., Li, Q., Wu, C., Zhao, M., and Zeng, R. (2010) Proteomic mining in the dysplastic liver of WHV/c-myc mice—insights and indicators for early hepatocarcinogenesis. *FEBS J.* **277**, 4039–4053
 64. Vizcaino, J. A., Cote, R. G., Csordas, A., Dienes, 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., and Hermjakob, H. (2013) The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069
 65. Lambert, J. P., Ivosev, G., Couzens, A. L., Larsen, B., Taipale, M., Lin, Z. Y., Zhong, Q., Lindquist, S., Vidal, M., Aebersold, R., Pawson, T., Bonner, R., Tate, S., and Gingras, A. C. (2013) Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nat. Methods* **10**, 1239–1245
 66. Domon, B., and Aebersold, R. (2010) Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* **28**, 710–721
 67. Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., Wernerus, H., Bjorling, L., and Ponten, F. (2010) Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250
 68. Prensner, J. R., Rubin, M. A., Wei, J. T., and Chinnaiyan, A. M. (2012) Beyond PSA: the next generation of prostate cancer biomarkers. *Sci. Translat. Med.* **4**, 127rv123
 69. Robey, E. L., Schellhammer, P. F., Wright, G. L., Jr., and el-Mahdi, A. M. (1985) Cancer serum index and prostatic acid phosphatase for detection of progressive prostatic cancer. *J. Urol.* **134**, 787–790
 70. Caccia, D., Zanetti Domingues, L., Micciche, F., De Bortoli, M., Carniti, C., Mondellini, P., and Bongarzone, I. (2011) Secretome compartment is a valuable source of biomarkers for cancer-relevant pathways. *J. Proteome Res.* **10**, 4196–4207
 71. Luo, X., Liu, Y., Wang, R., Hu, H., Zeng, R., and Chen, H. (2011) A high-quality secretome of A549 cells aided the discovery of C4b-binding protein as a novel serum biomarker for non-small cell lung cancer. *J. Proteomics* **74**, 528–538
 72. Kalin, M., Cima, I., Schiess, R., Fankhauser, N., Powles, T., Wild, P., Templeton, A., Cerny, T., Aebersold, R., Krek, W., and Gillessen, S. (2011) Novel prognostic markers in the serum of patients with castration-resistant prostate cancer derived from quantitative analysis of the pten conditional knockout mouse proteome. *Eur. Urol.* **60**, 1235–1243
 73. Farrah, T., Deutsch, E. W., Omenn, G. S., Campbell, D. S., Sun, Z., Bletz, J. A., Mallick, P., Katz, J. E., Malmstrom, J., Ossola, R., Watts, J. D., Lin, B., Zhang, H., Moritz, R. L., and Aebersold, R. (2011) A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell. Proteomics* **10**, M110.006353
 74. Zhang, H., Yi, E. C., Li, X. J., Mallick, P., Kelly-Spratt, K. S., Masselon, C. D., Camp, D. G., 2nd, Smith, R. D., Kemp, C. J., and Aebersold, R. (2005) High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Mol. Cell. Proteomics* **4**, 144–155
 75. Toprak, U. H., Gillet, L. C., Maiolica, A., Navarro, P., Leitner, A., and Aebersold, R. (2014) Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. *Mol. Cell. Proteomics* doi:10.1074/mcp.O113.036475
 76. Cohen, M. B., Griebing, T. L., Ahaghotu, C. A., Rokhlin, O. W., and Ross, J. S. (1997) Cellular adhesion molecules in urologic malignancies. *Am. J. Clin. Pathol.* **107**, 56–63
 77. Goel, H. L., Li, J., Kogan, S., and Languino, L. R. (2008) Integrins in prostate cancer progression. *Endocrine Relat. Cancer* **15**, 657–664
 78. Jin, S., and White, E. (2007) Role of autophagy in cancer: management of metabolic stress. *Autophagy* **3**, 28–31
 79. Wisniewski, J. R., Ostasiewicz, P., Dus, K., Zielinska, D. F., Gnad, F., and Mann, M. (2012) Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol. Syst. Biol.* **8**, 611
 80. Chen, J., Shah, P., and Zhang, H. (2013) Solid phase extraction of N-linked glycopeptides using hydrazide tip. *Anal. Chem.* **85**, 10670–10674
 81. Tsuboi, K., Sun, Y. X., Okamoto, Y., Araki, N., Tonai, T., and Ueda, N. (2005) Molecular characterization of N-acyl ethanolamine-hydrolyzing acid amidase, a novel member of the cholesteryl glycerol hydrolase family with structural and functional similarity to acid ceramidase. *J. Biol. Chem.* **280**, 11082–11092
 82. Sun, Y. X., Tsuboi, K., Zhao, L. Y., Okamoto, Y., Lambert, D. M., and Ueda, N. (2005) Involvement of N-acyl ethanolamine-hydrolyzing acid amidase in the degradation of anandamide and other N-acyl ethanolamines in

- macrophages. *Biochim. Biophys. Acta* **1736**, 211–220
83. Wang, J., Zhao, L. Y., Uyama, T., Tsuboi, K., Wu, X. X., Kakehi, Y., and Ueda, N. (2008) Expression and secretion of N-acyl ethanolamine-hydrolysing acid amidase in human prostate cancer cells. *J. Biochem.* **144**, 685–690
84. Shanguan, D., Cao, Z., Meng, L., Mallikaratchy, P., Sefah, K., Wang, H., Li, Y., and Tan, W. (2008) Cell-specific aptamer probes for membrane protein elucidation in cancer cells. *J. Proteome Res.* **7**, 2133–2139
85. Muller-Tidow, C., Schwable, J., Steffen, B., Tidow, N., Brandt, B., Becker, K., Schulze-Bahr, E., Halfter, H., Vogt, U., Metzger, R., Schneider, P. M., Buchner, T., Brandts, C., Berdel, W. E., and Serve, H. (2004) High-throughput analysis of genome-wide receptor tyrosine kinase expression in human cancers identifies potential novel drug targets. *Clin. Cancer Res.* **10**, 1241–1249
86. Shin, W. S., Kwon, J., Lee, H. W., Kang, M. C., Na, H. W., Lee, S. T., and Park, J. H. (2013) Oncogenic role of protein tyrosine kinase 7 in esophageal squamous cell carcinoma. *Cancer Sci.* **104**, 1120–1126
87. Shin, W. S., Maeng, Y. S., Jung, J. W., Min, J. K., Kwon, Y. G., and Lee, S. T. (2008) Soluble PTK7 inhibits tube formation, migration, and invasion of endothelial cells and angiogenesis. *Biochem. Biophys. Res. Commun.* **371**, 793–798
88. Lin, Y., Zhang, L. H., Wang, X. H., Xing, X. F., Cheng, X. J., Dong, B., Hu, Y., Du, H., Li, Y. A., Zhu, Y. B., Ding, N., Du, Y. X., Li, J. Y., and Ji, J. F. (2012) PTK7 as a novel marker for favorable gastric cancer patient survival. *J. Surg. Oncol.* **106**, 880–886
89. Endoh, H., Tomida, S., Yatabe, Y., Konishi, H., Osada, H., Tajima, K., Kuwano, H., Takahashi, T., and Mitsudomi, T. (2004) Prognostic model of pulmonary adenocarcinoma by expression profiling of eight genes as determined by quantitative real-time reverse transcriptase polymerase chain reaction. *J. Clin. Oncol.* **22**, 811–819
90. Ataseven, B., Angerer, R., Kates, R., Gunesch, A., Knyazev, P., Hogel, B., Becker, C., Eiermann, W., and Harbeck, N. (2013) PTK7 expression in triple-negative breast cancer. *Anticancer Res.* **33**, 3759–3763