SCIENTIFIC REPORTS

OPEN

Multi-lectin Affinity Chromatography and Quantitative Proteomic Analysis Reveal Differential Glycoform Levels between Prostate Cancer and Benign Prostatic Hyperplasia Sera

Sarah M. Totten[1], Ravali Adusumilli[1], Majlinda Kullolli[1], Cheylene Tanimoto [1], James D. Brooks[2], Parag Mallick[1] & Sharon J. Pitteri [1]

Currently prostate-specific antigen is used for prostate cancer (PCa) screening, however it lacks the necessary specificity for differentiating PCa from other diseases of the prostate such as benign prostatic hyperplasia (BPH), presenting a clinical need to distinguish these cases at the molecular level. Protein glycosylation plays an important role in a number of cellular processes involved in neoplastic progression and is aberrant in PCa. In this study, we systematically interrogate the alterations in the circulating levels of hundreds of serum proteins and their glycoforms in PCa and BPH samples using multi-lectin affinity chromatography and quantitative mass spectrometry-based proteomics. Specific lectins (AAL, PHA-L and PHA-E) were used to target and chromatographically separate core-fucosylated and highly-branched protein glycoforms for analysis, as differential expression of these glycan types have been previously associated with PCa. Global levels of CD5L, CFP, C8A, BST1, and C7 were significantly increased in the PCa samples. Notable glycoform-specific alterations between BPH and PCa were identified among proteins CD163, C4A, and ATRN in the PHA-L/E fraction and among C4BPB and AZGP1 glycoforms in the AAL fraction. Despite these modest differences, substantial similarities in glycoproteomic profiles were observed between PCa and BPH sera.

Prostate cancer (PCa) is one of the most common cancers among men in the U.S., with a projected 161,360 new cases in 2017 and an estimated 26,730 prostate cancer deaths[1]. For nearly three decades, prostate-specific antigen (PSA) has been used for prostate cancer screening, resulting in a significant increase in the number of detected cases of prostate cancer, with a shift toward detecting the cancer at earlier stages[2,3]. The beneficial effects of PSA screening are still debated, as conflicting evidence has been presented regarding whether or not it reduces prostate cancer mortality rates[3–7]. Circulating levels of PSA are also affected by other conditions of the prostate, including certain infections and inflammation, such as prostatitis, and benign enlargement of the prostate (benign prostatic hyperplasia, or BPH)[8]. Due to this lack of specificity for prostate cancer, the diagnostic capability of PSA suffers from a high number of false positives, resulting in unnecessary biopsies and overdiagnosis. Therefore there is a clinical need for a biomarker with greater specificity for prostate cancer that can distinguish between patients with benign disease from those at higher risk for prostate cancer, allowing patients to receive appropriate treatment and thus reducing the number of unnecessary biopsies, invasive surgeries, and the associated side effects.

The blood contains thousands of circulating proteins that are reflective of physiological and pathological states within the body, providing a rich source of potential biomarkers that can be easily sampled through less invasive means. However, performing in-depth plasma or serum proteomics is analytically challenging due to the

[1]Canary Center at Stanford for Cancer Early Detection, Department of Radiology, Stanford University School of Medicine, Palo Alto, CA, 94304, USA. [2]Department of Urology, Stanford University School of Medicine, Stanford, CA, 94305, USA. Correspondence and requests for materials should be addressed to S.J.P. (email: spitteri@stanford.edu)

complexity of the protein mixture and the large dynamic range of protein concentrations in the blood, which spans more than ten orders of magnitude[9]. This challenge has been met with a number of chromatography and mass spectrometry-based approaches designed to systematically achieve greater coverage deeper into the plasma/serum proteome, ranging from the selection of specific subsets of proteins through affinity chromatography, to extensive pre-fractionation at the protein and peptide level[10]. The depth of proteomic analysis has also been enhanced by technological improvements in high-performance mass spectrometers with greater scan speeds, resolving power, and sensitivity. Furthermore, it is estimated that over 50% of the human proteome is glycosylated[11]. Glycosylation is a common yet highly complex post-translational modification recognized to play an important role in a variety of biological processes, such as cell-cell communication, host-pathogen interactions, and immune response[12–14]. Glycoproteins can have multiple sites of glycosylation with varying degrees of occupancy. Additionally, a variety of different glycans can occupy a given glycosylation site, giving rise to the complex microheterogeneity that notoriously complicates the characterization and analysis of protein glycosylation. Changes in glycosylation have been correlated to disease status in a variety of cancers, including prostate cancer, and exploiting these aberrancies has shown promise for use as effective biomarkers[13,15–18]. The bulk of the research on glycosylation changes in prostate cancer has focused on characterizing the various glycoforms of PSA to improve its clinical utility[19–23]. These studies find that core-fucosylation and the sialic acid linkage of PSA glycoforms play a key role in differentiating non-PCa patients and those with BPH from low- and high-risk PCa cases. A number of glycomic and glycoproteomic studies have also looked beyond PSA for prostate cancer-specific glycosylation changes in a variety of clinical samples, including urine, seminal fluid, blood plasma/serum and tissue[24]. In a review by Drake *et al.*, it was reported that an increase in the expression of N-acetylglucosaminyl transferase V in prostate tumors leads to a subsequent increase in β1–6 branching, forming larger tri- and tetraantennary N-linked structures in prostate tissue[24]. Additionally, a recent review describing the mechanisms and clinical implications of altered glycosylation in cancer reports that an increases in glycan branching, as well as increased fucosylation and sialylation, are the most widely occurring cancer-associated alterations in protein glycosylation[13]. Although progress has been made in realizing the importance of glycosylation in the development and progression of cancer, it still remains a significant analytical challenge to obtain detailed glycan characterization while still retaining protein- and site-specific, information in large, complex biological mixtures derived from clinical samples.

In this study, we perform a quantitative glycoproteomic analysis using multi-lectin affinity chromatography (M-LAC) to compare the circulating levels of proteins and their glycoforms from the sera of men with BPH to those with prostate cancer. We use a series of chromatographic separations to simultaneously decomplex the protein mixture and to enrich for proteins with specific types of glycosylation by using lectins with affinity for specific glycan motifs. Here we chose *Aleuria aurantia* lectin (AAL) to capture core-fucosylated proteins and *Phaseolus vulgaris* leucoagglutinin erythroagglutinin, and *Phaseolus vulgaris* erythroagglutinin (jointly abbreviated as PHA-L/E) to capture highly branched glycans[25]. These lectins were chosen to target types of glycosylation reported to be aberrant in prostate cancer, while simultaneously fractionating the complex mixture of proteins to enhance depth of analysis. Using an M-LAC approach to separate glycoforms allows for a systematic way to screen for changes in glycosylation in a complex mixture while retaining protein-specific information. We identify differences at the global protein level as well as among specific glycoforms of quantitated proteins that could be used to aid in differentiating BPH from PCa cases.

Methods

Methods. *Serum Sample Collection from Benign Prostatic Hyperplasia and Prostate Cancer Patients.* De-identified serum samples used in this study were taken from an existing serum bank collected on patients immediately prior to surgery for prostate cancer, or from men with elevated serum PSA levels, known BPH, and two or more previous negative prostate biopsies. Following informed consent, all blood samples were collected in red top tubes, allowed to clot, and centrifuged. Serum was then aliquoted (500 μL) into tubes and frozen at −80 °C to limit effects of freeze/thaw cycles. Samples were retrieved and analyzed. All ten PCa samples were from men in whom the cancer volume was 1 cc or greater and showed pathological Gleason scores of $4 + 3 = 7$ (Table 1). Use of the existing serum resource has been reviewed and certified by the Stanford University Institutional Review Board (IRB). The approved IRB protocol for blood collection allows for correlation of clinical information, including disease status and follow-up, with molecular measurements in the blood, including protein analysis. The patients have provided informed consent allowing for use of their tissues/blood specimens. Samples collected prior to 1999 are considered existing samples in an established clinical and tissue database, and can be used under an IRB approved Waiver.

Immunodepletion of Abundant Proteins. Pooled normal human male EDTA plasma was purchased from Innovative Research and used as a reference sample pool. 200 μL of each clinical serum sample (17 in total) and 17–200 μL aliquots of the reference pool were immunodepleted using CaptureSelect™ HumanPlasma 14 affinity resin for the removal of abundant proteins (albumin, IgG, IgM, IgA, IgE, IgD, free light chains, transferrin, fibrinogen, α-1-antitrypsin, apolipoprotein A1, α-1-2-macroglobulin, α-1-acid-glycoprotein, and haptoglobin) as previously described[26]. The flow-through fractions from the CaptureSelect column containing the unbound proteins were desalted and concentrated using Amicon Ultra 15 mL 3 K NMWL centrifugal filters (Millipore). An aliquot of 10 μL from each concentrated sample was set aside for a Bradford assay to determine total protein concentration, and the remainder of the sample was diluted 1:10 in protein denaturation buffer (8 M urea, 50 mM Tris-HCl, 0.05% octyl β-D-glucopyranoside, pH 7.5, prepared in 100 mM ammonium bicarbonate).

Reduction, Alkylation, and Isotopic Labeling. Protein disulfide bonds were reduced using dithiothreitol (DTT) at a final concentration of 5.5 mM for two hours at room temperature. Cysteine residues were alkylated using

Sample Type & Number	Age	PSA (ng/mL)	Percent G4/5*	Total Cancer Volume (cc)
PCa_1	56	10.20	70	4.37
PCa_2	66	3.92	90	8.55
PCa_3	50	15.48	80	9.00
PCa_4	58	6.36	60	9.03
PCa_5	60	21.30	90	4.93
PCa_6	70	3.27	70	6.00
PCa_7	46	30.13	90	1.00
PCa_8	68	16.11	95	29.39
PCa_9	64	13.40	60	7.20
PCa_10	56	8.71	50	4.56
BPH_1	62	3.14		
BPH_2	73	4.20		
BPH_3	69	11.16		
BPH_4	71	13.80	N/A	
BPH_5	56	7.67		
BPH_6	44	3.42		
BPH_7	61	8.21		

Table 1. Clinical Characteristics of PCa and BPH samples. *Percentage of Gleason Pattern 4 or 5. The remainder is pattern 3. All PCa samples are Gleason $4 + 3 = 7$.

acrylamide for one hour at room temperature in the dark. For relative quantitation, all clinical BPH and PCa samples were alkylated with 1,2,3-^{13}C3 (heavy) acrylamide in a 7.4 mg per mg of protein ratio, and all reference samples were alkylated with 1,2,3-^{12}C3 (light) acrylamide in a 7.1 mg per mg of protein, as described in Faca et al.[27]. After alkylation, each clinical sample was then combined with a reference sample, making 7 BPH/Reference pairs and 10 PCa/Reference pairs. Each pair was concentrated and buffer exchanged into 1 mL of $1 \times$ PBS for subsequent lectin chromatography.

Multi-Lectin Affinity Chromatography. Multi-lectin affinity chromatography (M-LAC) was used to separate proteins by specific glycoforms. The following M-LAC experiments were modified for use on this specific application from previously described methodologies employing immunodepletion and M-LAC to investigate glycoproteomic changes in human plasma and serum[28–32]. *Aleuria aurantia* lectin (AAL), *Phaseolus vulgaris* leucoagglutinin, and *Phaseolus vulgaris* erythroagglutinin (PHA-L/E) were used to capture core fucosylated protein glycoforms, and glycoforms carrying highly branched complex type glycans, respectively. Agarose-bound AAL and PHA-L/E lectins were purchased from Vector Laboratories (Burlingame, CA) and gravity packed in house, as previously described[25]. All chromatography was performed on an Agilent 1260 Bio-Inert HPLC system, equipped with a quaternary pump, a manual injector, UV multiple wavelength detector, and an analytical-scale fraction collector. Reduced and alkylated protein was loaded onto the M-LAC column for fractionation. First the flow-through fraction was collected, containing the non- or otherwise-glycosylated proteins making up the unbound (UNB) fraction. Bound glycoproteins were eluted in series using competitive saccharide binding and low pH elution buffers. Core-fucosylated glycoforms bound to AAL were eluted with 200 mM L-fucose ACROS Organics). Lastly, glycoforms bound to PHA-L/E were eluted with 100 mM acetic acid, pH 3.8. Each of the three M-LAC fractions (henceforth abbreviated as UNB, AAL, and PHA) were concentrated to 250 μL using 3 K NMWL Amicon centrifugal filters.

Reversed-Phase Chromatography. Reversed-phase (RP) fractionation was performed on a 100 mm × 2.1 mm ID stainless steel column packed with POROS®R2 (Applied Biosystems), with a 2,000 Å particle size, poly styrene-divinylbenzene stationary phase. Each M-LAC fraction was further separated into 13 RP fractions using an increasing gradient of organic mobile phase as follows: 0–5 minutes at 100% buffer A (0.1% trifluoroacetic acid in water); 5–38 minutes, ramp to 90% buffer B (0.1% trifluoroacetic acid in acetonitrile); 38–40 minutes hold at 90% buffer B; 40–50 minutes hold at 95% buffer A for re-equilibration. In total, 39 fractions (13 RP fractions x 3 M-LAC fractions) were collected per patient sample. All fractions were frozen at −80 °C overnight and lyophilized. RP fractions were reconstituted in 50 μL of 50 mM ammonium bicarbonate in 4% acetonitrile and subsequently digested with 0.5 μg of trypsin at 37 °C for 18 hours.

LC – Tandem Mass Spectrometry Analysis. Tryptic peptides were analyzed by LC-MS/MS on an Ultimate 3000 RSLCnano system (Dionex) coupled to an Orbitrap Elite mass spectrometer (Thermo Fischer Scientific) with a nanospray ion source. Fifteen μL of sample (approximately 5–10 μg of peptides) was loaded onto a C18 trap column for brief desalting and concentration, then separated on a 25 cm C18 analytical column (Picofrit 75 μm ID, New Objective, packed with MagicC18 AQ resin) over a 140 minute, multi-step gradient of increasing organic phase. Each MS/MS experiment consisted of an initial MS1 scan over a mass range of 400–1800 m/z, followed

by 10 subsequent data-dependent collision-induced dissociation (CID) fragmentation events of the top 10 most intense +2 or +3 ions from the MS1 spectrum over an acquisition time of 140 minutes.

Data Processing and Statistical Analysis. The raw files obtained from LC-MS/MS were converted to mzXML format using MSconvert from the ProteoWizard software[33]. The resulting mzXML files were used to identify proteins by searching against human UniProtKB database on the LabKey server using X!Tandem algorithm[34–36]. Search results from X!Tandem were then analyzed by PeptideProphet and validated using ProteinProphet[37,38]. Protein groups and peptides with a score greater than 0.9 and 0.6, respectively, were retained for protein identification and quantitation. Heavy-to-light (H/L) ratioswere computed for cysteine-containing peptides with acrylamide labels using the Q3 quantitation algorithm[27]. The fragment ion mass accuracy was set to the default ±0.5 Da in X!Tandem. Only peptides with precursor fractional delta mass <20 ppm were used for quantitation. Peptide H/L ratios were averaged for each protein group by each M-LAC fraction resulting in three separate quantitation values for UNB, AAL and PHA. Additionally, global protein levels were calculated as an average of all glycoforms, using all quantitated peptides regardless of which M-LAC fraction the peptide was identified in. Protein groups were subsequently assembled by gene name into normalized gene groups. One gene name was picked to represent each group. H/L ratios of each group were log(2) transformed and median-centered (across all fractions per sample) for normalization. A permutation two-sample t-test (Welch's unequal variance) was performed on the mean H/L ratios for genes quantified in BPH and PCa groups using perm (R package) to randomly shuffle the data and estimate the distribution of the test statistic[39]. The p-value for the observed test statistic was then determined from the sampling distribution. False discovery rate (FDR) was estimated from the permutation test as the number of false discoveries at or above a given p-value. We used an FDR bound of 1% throughout the manuscript for differential quantification. Permutation tests were only performed on proteins that were quantified in ≥3 samples in both the BPH and PCa groups.

Method Reproducibility. The reproducibility of the entire platform described above was accessed using three identical replicates of the reference plasma designed to yield a 1:1 heavy-to-light ratio by combining 1,2,3-^{13}C3 heavy-labeled reference plasma with 1,2,3-^{12}C3 light-labeled reference plasma. After immunodepletion, reduction, and alkylation/labeling steps, each of the three replicates were separated by M-LAC, RP fractionated, digested, and analyazed by LC-MS/MS as described above. The identical replicates yielded acceptable reproducibility, and the data from these experiments is presented in the supplementary material (see Supplementary Dataset 1 for lists of identified peptides with quantitation, and Supplementary Figures S1 and S2 for spectral count and H/L ratio reproducibility, respectively).

Data Availability. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Results

The overall objective of this study was to elucidate differences in the circulating levels of serum protein glycoforms that may have utility in distinguishing between PCa patients from those with BPH. Multi-dimensional chromatographic separation at the intact protein level was used to increase the depth of proteomic analysis given the dynamic range issues associated with blood plasma and serum. Using lectins to fractionate the protein mixture also provides the ability to separate protein glycoforms and to obtain an additional layer of information concerning the types of glycosylation expressed on serum glycoproteins. By using lectins with affinity for particular glycan structures, it is possible to systematically monitor global changes to the levels of specific glycoforms across hundreds of serum proteins from PCa and BPH specimens. Figure 1 shows an overview of the analytical design of these experiments, where all cases (7 BPH samples and 10 PCa samples) were quantitated relative to a pooled reference sample of male plasma that served as an internal control. Each PCa and BPH sample was isotopically labeled with heavy ^{13}C-acrylamide, and combined with the reference sample (labeled with a light ^{12}C-acrylamide isotope). Each sample was then loaded onto an M-LAC columncontaining AAL and PHA-L/E lectins. AAL preferentially binds fucose linked α1-6 to N-Acetylglucosamine (GlcNAc), and – to a lesser extent – fucose linked α1-3 to N-acetyllactosamine (LacNAc), and was used to enrich for primarily core-fucosylated N-linked glycoforms and to a lesser extent terminal fucosylated N-linked glycoforms[14]. PHA-L has affinity for tri- and tetra-antennary complex-type glycans, while PHA-E recognizes bisected di-, and triantennary complex-type N-linked glycans[14]. These two lectins were used together to enrich for protein glycoforms containing highly-branched, complex-type structures. Three separate fractions were collected from the M-LAC column including the flow-through of unbound proteins (UNB fraction), proteins bound to AAL (AAL fraction), and proteins bound to PHA-L/E (PHA fraction). For each specimen, each of the three M-LAC fractions was then further fractionated by reversed-phase chromatography, trypsinized, and analyzed by LC-MS/MS, as described in detail above. For complete lists of peptide identification and quantitation per fraction, per sample, please see Supplementary Dataset 1. Comparisons were made between PCa and BPH groups among proteins that were quantitated in at least three samples in each group. For a complete list of all protein H/L values in each sample, please refer to Supplementary Dataset 2.
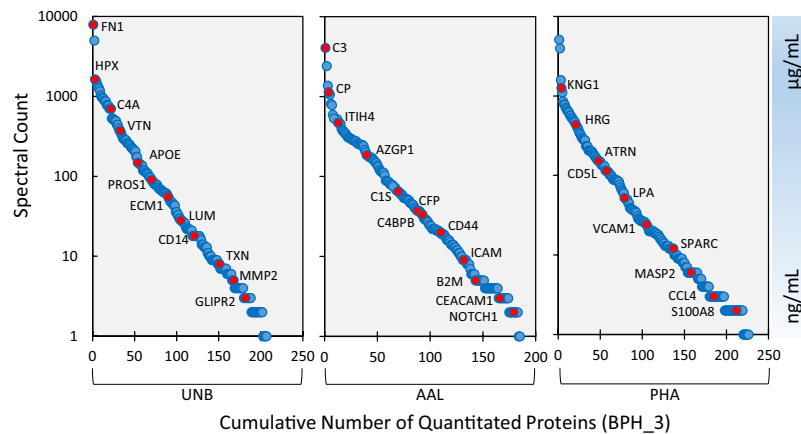
Depth of Proteomic Analysis. Due to the complexity and large dynamic range of protein concentrations inherent to the human serum and plasma proteome, we opted to pre-fractionate the protein mixture to increase lower-abundance protein identifications[10]. The use of high-resolution mass spectrometry also increases the number of identifications and the extent of protein sequence coverage. In this study, samples were extensively fractionated at the intact protein level and separated by glycoform. By LC-MS/MS analysis, on average, 187 proteins (5% FDR, collapsed by gene name, quantified in at least three samples) were quantitated in the UNB fraction, 141 in the AAL
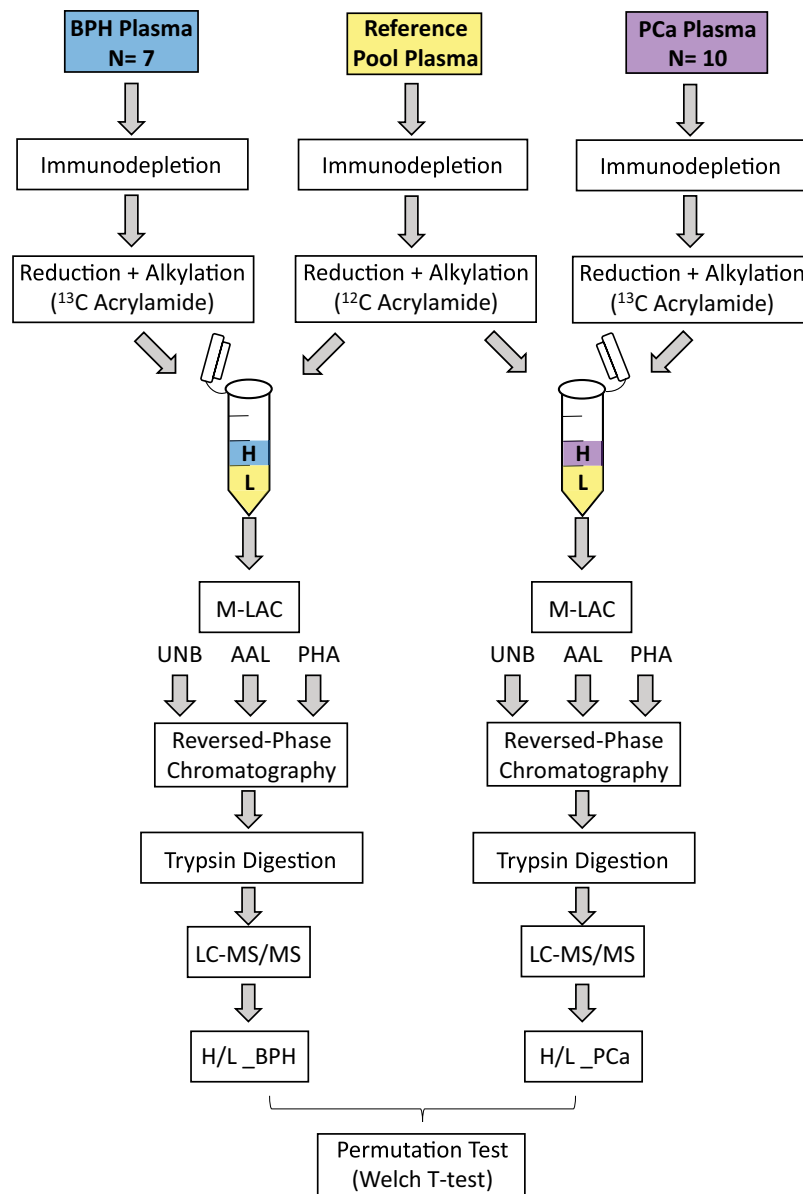
Figure 1. Experimental Design and Analytical Work Flow.

fraction, and 159 in the PHA fraction among the PCa specimens. The BPH serum samples yielded 183, 183, and 203 quantitated proteins in the UNB, AAL, and PHA fractions, respectively. These numbers of quantitated proteins results are typical for the chosen methodology, which involves the isotopic labeling of cysteine residues via alkylation with heavy and light isotopes of acrylamide, as described above. A cysteine labeling approach was utilized for efficient labeling at the protein level (versus peptide level) to ensure compatibility with the intact protein fractionation used in our sample preparation. Quantitation is therefore limited to cysteine-containing peptides, which restricts the number of proteins available for quantitation. Furthermore, the reported number of proteins quantitated across samples is a result of restricting proteins to those present in at least three samples in both the BPH and PCa groups to allow for statistical analysis. Figure 2 demonstrates the range of proteins quantitated in a single representative sample (BPH_3, Table 1), plotting spectral counts versus cumulative number of quantitated proteins. Spectral counts and known proteins concentrations in blood were used to estimate the dynamic range of the fractionated protein mixture, which contained proteins with concentrations ranging from micrograms per milliliter (such as C3, VTN, and AZGP1) down to the nanogram per milliliter level (such as MMP2, ICAM1, and CEACAM1)[9,40–42]. A complete list of the number of proteins quantified for each individual sample can be found in Supplementary Dataset 3, and a list of identified peptide sequences can be found in Supplementary Dataset 1.

Protein Expression Relative to Reference Sample. To quantitate alterations in protein abundance, cases (BPH or PCa) were isotopically labeled with heavy 1,2,3 ^{13}C-acrylamide. All cases were compared against a reference pool sample isotopically labeled with the light ^{12}C-acrylamide isotope. The mass of ^{13}C- and

Figure 2. Depth of Proteomic Analysis – the above figure plots the spectral count (per protein) against the cumulative number of quantitated proteins ordered from highest to lowest spectral count (by unique gene name). Examples of the types of proteins that were quantitated are highlighted in red for reference. Literature values of concentrations for these proteins were used to estimate the dynamic range of the protein mixture. This plot was generated from a single, representative sample with an average number of protein identifications. Across all M-LAC fractions, 315 proteins were quantitated in this sample.

^{12}C-acrylamide differs by 3 Daltons, inducing a shift in the mass spectrum of cysteine-containing peptides. The heavy-to-light (H/L) ratio is calculated based on the peak intensity of the heavy and light forms of the peptide integrated across the elution of the peak. An overview of protein quantitation relative to the reference sample is presented in Fig. 3, which features a heat map color-coded by average log2H/L values for each M-LAC fraction in PCa and BPH groups. The proteins are ordered along the horizontal axis from high to low global protein log2H/L values in the PCa group. The set of 248 proteins presented in Fig. 3 includes all quantitated proteins with an n ≥ 3 at the global protein level (i.e. there were at least three samples containing the protein in both groups). Blank (white) spaces in the heat map indicate one of two instances: (1) there were not enough data points (n < 3) in the group to perform statistical analysis, or (2) protein glycoforms were not detected in that fraction in any sample (n = 0). In the former, the H/L data points that were present contributed to the global protein level.

Both PCa and BPH demonstrate deviation from the reference sample, however, the patterns in protein expression were overall quite similar between the BPH and PCa groups. At the global protein level, the most overexpressed proteins (with respect to the reference sample) in the PCa samples were also highly overexpressed in the BPH group, including SPARC, metalloproteinases MMP2 and MMP9, ADAMTS13, MENT, COMP, SNED1, SLPI, and PCSK9. Similarly, proteins with decreased global levels in the PCa group showed comparable decreased levels in the BPH group, including SERPINA4, S100A8, and PON1, among others. The most over-expressed protein in the dataset was MMRN1 (multimerin-1), a glycoprotein with 23 N-linked glycosylation sites that was exclusively detected in the AAL fraction. Global MMRN1 levels were highly elevated in both groups, with a six-fold increase quantitated in the PCA group and a seven-fold increase in the BPH group (PCA log2H/L = 2.59, BPH log2H/L = 2.91, p-value = 0.477). In general, the same trend was found among the protein levels in individual M-LAC fractions. For most proteins, fold-changes were in concordant directions across M-LAC fractions and were similarly expressed with respect to the reference sample in the PCa and BPH groups. Additionally, protein glycoforms were typically identified in the same M-LAC fractions in the PCa and BPH groups. We did not observe any cases in which all of the PCa glycoforms of a given protein were found in one M-LAC fraction and all the BPH glycoforms of that same protein were found in another, which would have clearly indicated a blatant contrast in glycosylation patterns. However, some glycoform-specific alterations were observed that were not immediately apparent at the global protein level. For example, MINPP1 (Fig. 3) was quantitated in all three fractions, but demonstrated elevated levels in the UNB fraction, decreased levels among the AAL glycoforms, and relatively unchanged levels among the PHA glycoforms. The global protein level was slightly overexpressed (less than 1.5-fold), demonstrating which M-LAC fractions were driving the overall trend. This trend, however, was observed in both the PCa and BPH groups, rendering MINPP1 ineffective for distinguishing PCA from BPH. ADAMTSL4, SSC5D, VCAM1, and IGFBP7, to highlight a few, presented similar cases in which different M-LAC fractions showed log2H/L ratios in opposing directions – a trend that would not have been observed without glycosylation-based fractionation – but ultimately resulted in indistinguishable patterns between PCa and BPH groups.

In some cases there were insufficient data points to quantitate glycoforms of individual M-LAC fractions, but had an n ≥ 3 once values from all fractions were combined to determine a global protein level, as is shown in Fig. 3. TRIM37, JAG1, RECK, ASGR2, and SPARCL1 all demonstrated at least 1.5-fold over-expression compared to the reference sample at the global protein level in PCa and BPH groups. Some of the more striking differences were observed in TFRC and APOF, in which the BPH group demonstrated a 2-fold increase, while the PCa group was unchanged (TFRC) or decreased (APOF).
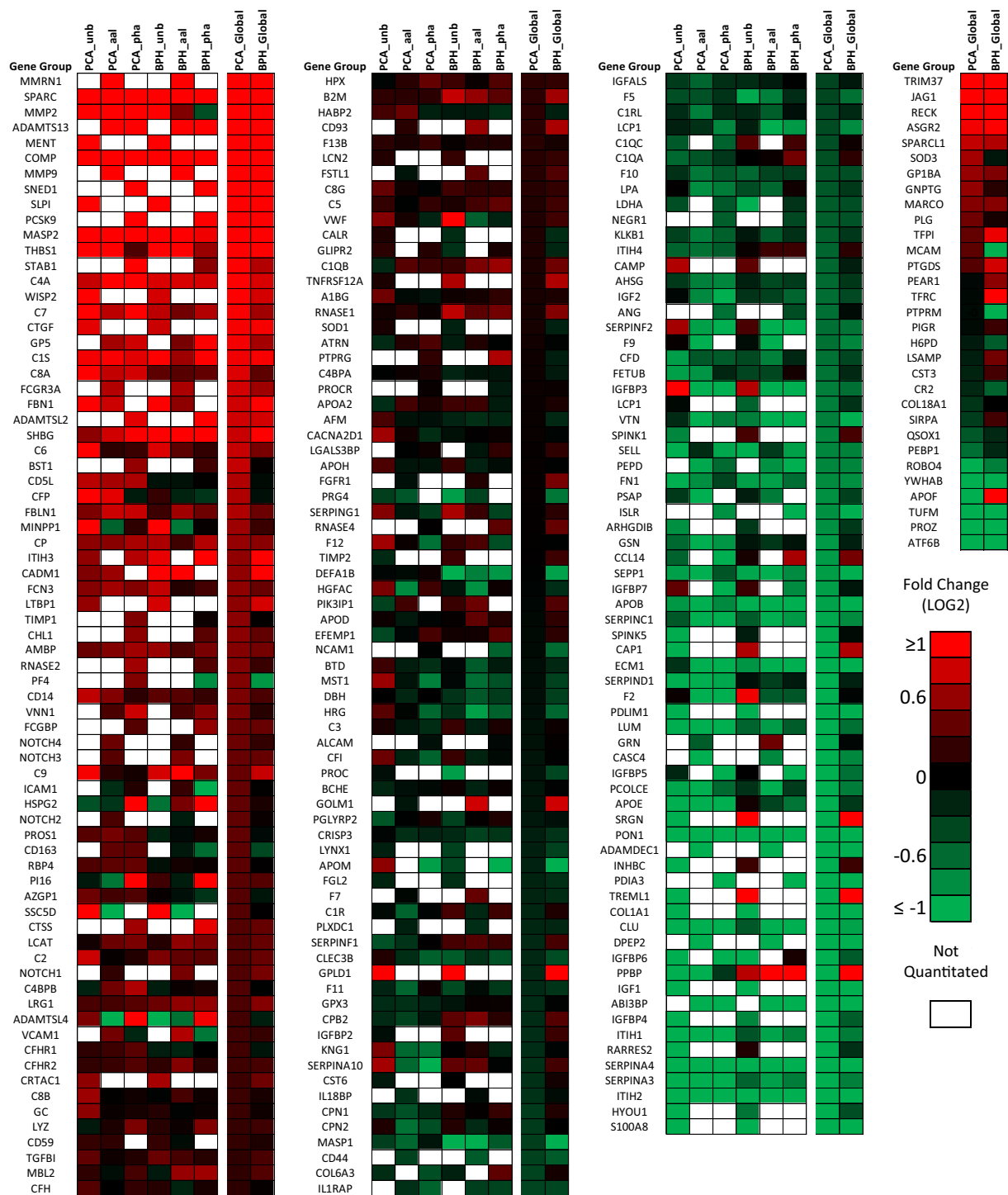
Figure 3. Heat map color-coded by mean fold change (log2H/L) in the PCa and BPH groups with respect to the reference plasma. Each M-LAC fraction is represented in its own column and proteins are ordered from largest fold change in the PCa group at the global protein level. This list of 248 proteins includes those that had an N ≥ 3 for quantitation per M-LAC fraction and/or at the global protein level (far right).

Differential Expression Between PCa and BPH. Among the quantitated proteins shown in Fig. 3, a permutation t-test was performed to determine significant differences between the BPH and PCa log2H/L ratios for those with an n ≥ 3 in each group. A t-test was done at the global protein level, which was calculated as an average of all identified glycoforms, for which 248 comparisons were made. Comparisons were also made between each of the three M-LAC fraction (for proteins with n ≥ 3 within that fraction). Altogether 153 proteins were compared in the AAL fraction, 156 in the PHA fraction, and 168 in the UNB fraction. A complete list of PCa and BPH mean log2H/L values and corresponding p-values can be found in Supplementary Dataset 4. To determine the extent
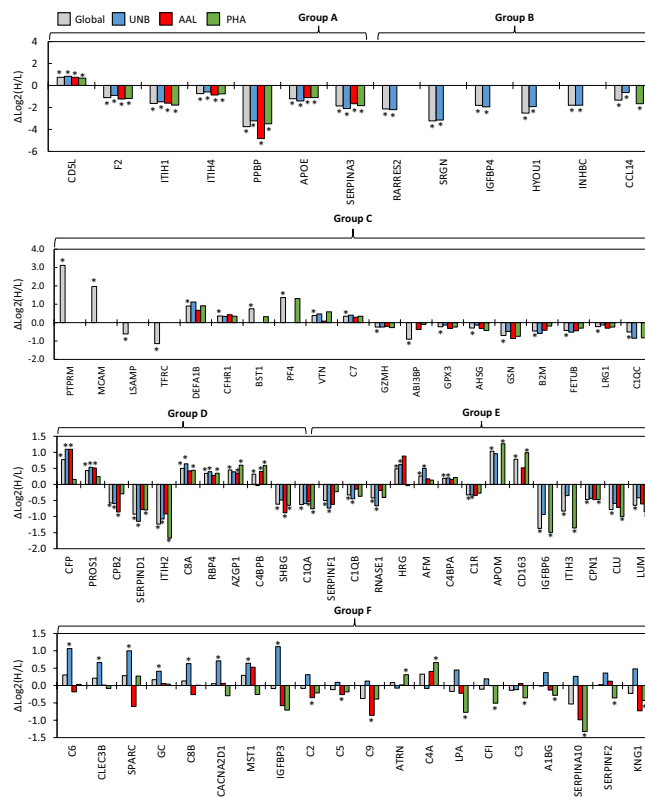
Figure 4. Difference in mean ($\Delta = \log2H/L\_PCa - \log2H/L\_BPH$) plotted by M-LAC fraction per protein. Asterisk denotes a statistically significant difference in mean ($p \leq 0.05$). Proteins are grouped by expression patterns: (**A**) Quantitated in all fractions and significantly different in all fractions; (**B**) Significantly different in every M-LAC fraction it was quantitated in; (**C**) Only significant at the global protein level; (**D**) Significantly different in multiple M-LAC fractions; (**E**) Significantly different in only one M-LAC fraction and at the global protein level; (**F**) Significantly different in one M-LAC fraction and not at the global protein level.

of differential expression in protein levels between PCa and BPH groups, the difference in mean log2H/L was calculated for each protein ($\Delta = \log2H/L_{PCa} - \log2H/L_{BPH}$) and plotted in Fig. 4. This figure illustrates a summary of the direction and magnitude of the difference in log2H/L values in each M-LAC fraction for proteins that had statistically significant differences in one, two, or all three fractions, and/or at the global protein level. Fifty-nine proteins had significantly different levels (p-values ≤ 0.05) at the global protein level. Additionally, 18 were significantly different in the AAL fraction, 30 in the PHA fraction, and 37 in the unbound UNB fraction. As shown in Fig. 4, proteins were organized into groups according to the pattern in which the protein levels were altered. In some instances, proteins yielded increased or decreased levels in the PCa relative to the BPH uniformly across all quantitated glycoforms. Seven proteins were quantitated in all three M-LAC fractions and demonstrated differential expression irrespective of glycosylation, as shown in Group A in Fig. 4. Of these, CD5L was the only protein to have significantly over-expressed protein levels that were higher in the PCa than the BPH in every fraction. Circulating levels of F2, ITIH1, ITIH4, PPBP, APOE, and SERPINA3 were lower in PCa than in the BPH across all glycoforms. Similarly, proteins in Group B (Fig. 4) yielded significantly different protein levels in PCa versus BPH in every fraction that they were quantitated in, but were not detected and/or quantitated in all three fractions. Group C (Fig. 4) contains proteins in which significant differences were only determined at the global protein level, suggesting the trend was only statistically significant when expressed as an average of all detected glycoforms due to the cumulative effect of combining all the data points and increasing the statistical power. Among these, global levels of PTPRM, MCAM, DEFA1B, CFHR1, BST1, PF4, VTN, and C7 were significantly higher in the PCa group than in the BPH group, and LSAMP, TFRC, KLKB1, ABI3BP, GPX3, AHSG, GSN, B2M, FETUB, LRG1, and C1QC were significantly lower in the PCa than in the BPH.

Glycoform-specific changes in protein expression. In many cases, glycoform-specific alterations were observed in the patterns of protein expression between the PCa and BPH groups, as is depicted in Groups D, E, and F in Fig. 4. Proteins in Group D were quantitated in all M-LAC fractions and yielded significant differences among multiple glycoforms. Properdin (CFP) showed significantly elevated levels in PCa compared to BPH among core-fucosylated glycoforms bound to AAL and among the other-wise or non-glycosylated forms quantitated in the unbound fraction. Proteins AZGP1, and C4BPB also demonstrated significantly higher levels in glycoforms containing core-fucosylated glycans in the AAL fraction and highly-branched glycans in the PHA fraction, however was not significantly different in the unbound fraction. Notably, the comparison in AZGP1 in the unbound

fraction was nearly significant (p = 0.06) and demonstrated a similar trend of increased levels in PCa patients, suggesting that AZGP1 may be over-expressed at the protein level and not necessarily displaying differential glycosylation. However, proteins in Groups E and F were only significantly different in a single M-LAC fraction (proteins in Group E were also significant at the global protein level). For example, proteins such as APOM and CD163 showed significantly increased levels in the PCa group versus the BPH group only in the PHA fraction, however a similar, although insignificant, trend was observed among the glycoforms in the UNB and AAL fractions as well. Interestingly, Group F includes 20 proteins that showed significant differences only in a single M-LAC fraction and were not significant at the global protein level, demonstrating the advantage of fractionating intact proteins by glycoform prior to digestion and MS analysis. Notable in this group is ATRN, which was significantly increased in the PHA fraction exclusively, suggesting that only the highly-branched, complex type glycoforms of this protein were expressed in higher concentrations in PCa patients. Additionally, several members of the complement family yielded lower levels of core-fucosylated glycoforms in PCa patients relative to BPH patients, including C2, C5, and C9. It should be noted that although the difference between the PCa and BPH means were negative in these three proteins (as shown in Fig. 4), all the log2H/L ratios were positive compared to the reference sample, but just to a lesser extent in the PCa group than in the BPH group. This is most notably the case in the AAL fraction of C9, for which the BPH mean was 2-fold higher than the reference sample, and the PCa mean was only 1.11-fold higher than the reference sample (p = 0.024).

In the case of KNG1, the changes in protein level was significantly different in all M-LAC fractions, but not in concordant directions. KNG1 protein levels were lower in the PCa group among glycoforms in the AAL and PHA fractions, but were elevated in the non- or otherwise-glycosylated forms in the UNB fraction. In this case, the opposing trends among the M-LAC fractions resulted in a loss of significance at the global protein level. By fractionating by glycosylation type, we were able to find differences among certain glycoforms, even when the global protein level was unchanged – a trend that would have been missed in proteomic workflows that do not target glycoproteins specifically.

Discussion

In this study we compare alterations in protein expression in men with BPH relative to normal male plasma to the differential protein expression derived from prostate cancer patients. We hypothesized that differential expression would be observed among protein glycoforms in specific M-LAC fractions and that these differences may show potential for distinguishing BPH from PCa. We observed alterations in circulating levels of proteins both at the global level and among specific glycoforms captured by lectins with affinity for core-fucosylated species (AAL) and highly branched complex-type glycans (PHA-L/E). Quantitation relative to a pooled reference sample revealed substantial similarity in protein and glycoprotein expression among PCa and BPH groups, leaving only a small number of cancer-specific alterations. For this reason, we included proteins with a fold change of ≥1.5 (0.585 on a log2 scale) in our analysis to improve our sensitivity for smaller, more subtle changes in protein level. The extent of overlap in protein expression patterns between these two groups is not entirely unexpected, as it is known that inflammatory and immune response proteins are detectable in the blood of cancer patients in addition to proteins derived from the tumor itself[43]. These results speak to the difficulty in identifying (glyco)proteins that can effectively distinguish PCa from BPH, which is a pitfall of the currently used clinical biomarker for prostate cancer, PSA. Inflammation plays a role in the pathogenesis of both BPH and cancer[43–46], and indeed a number of inflammation-associated proteins were detected and quantitated within this sample set. Predominating among these were several glycoproteins of the complement system. Notably, complement activators MASP2, or mannan-binding lectin serine protease-2, C1S, and the subsequent pro-inflammatory cleavage product C4A all showed elevated levels in all M-LAC fractions among both PCa and BPH patients[47]. Although considerably increased compared to the reference sample, the abundances of these proteins and their separated glycoforms were similarly altered in both groups and may not serve as ideal cancer-specific markers. Notably, SPARC – a secreted glycoprotein that plays a role in tissue remodeling through interactions with the extracellular matrix – was overexpressed by at least 3-fold with respect to the reference sample in both PCa and BPH across all M-LAC fractions. Although SPARC has been reported to be dysregulated in several different types of cancers[48,49], has been associated with prostate cancer progression and bone metastasis[50–53], and has been suggested as a possible target for cancer therapeutics[48,54,55], the elevated global protein level of SPARC in this study was determined to be non-specific for prostate cancer. A significant difference between PCa and BPH was observed only in the UNB fraction, in which the levels of non- or otherwise-glycosylated forms of SPARC were significantly higher in the PCa group (PCA log2H/L = 2.58, BPH log2H/L = 1.58, p = 0.0019). Although highly overexpressed in both groups, the levels of SPARC showed trends of being more overexpressed in the PCa group in the PHA fraction and at the global protein level (not significant), and merits further study as a potential biomarker in a larger cohort with greater statistical power.

Despite very similar patterns of protein and glycoprotein expression, a number of notable statistically significant alterations were observed between the two groups that should be highlighted. On the global protein level, there were five proteins that yielded large positive differences between PCa and BPH means (i.e. the mean PCa was greater than that of BPH), were elevated at least 1.5-fold with respect to the reference sample in the PCa group (log2H/L ≥ 0.585), and were significantly different between PCa and BPH (p ≤ 0.05). These five proteins were CFP, BST1, CD5L, C8A, and C7. CD5L levels were increased 1.5-fold in every M-LAC fraction in PCa group and remained relatively unchanged in the BPH group, and yielded one of the most statistically significant differences in the dataset (p = 1.52E-07). Additionally, the increase in global levels of proteins AZGP1, and PROS1 were significantly different between PCa and BPH (p = 0.0003 and p = 2.33E-06, respectively), however it should be noted that AZGP1 and PROS1 were not particularly over-expressed in PCa with respect to the reference sample (both yielded log2H/L values around 0.35, or a 1.25-fold increase), making these proteins less ideal as potential candidates for diagnostic biomarkers.

In the above mentioned proteins, separating by glycoform using lectins did not reveal any differences in glycosylation pattern, although the additional chromatography step may have aided in detecting these moderately-abundant proteins. The real advantage of using an M-LAC approach was made evident when glycoform-specific changes in protein levels were observed between PCa and BPH groups, especially when no changes were evident at the global protein level. These alterations in glycoform expression would be missed without lectin fractionation. The most notable glycoform-specific alterations were observed in the PHA fraction, which contained glycoforms with highly-branched complex type glycans. The degree of branching among N-linked glycan structures has been previously shown to be differentially expressed in cancer, with a reported increase in the expression of complex N-glycans with a β1,6-branched GlcNAc due to the increased activity of acetylglucosaminyltransferase GNT-V[13,24]. Overexpression of tri- and tetra-antennary N-glycan structures has been observed in cell line xenograft mouse models of prostate cancer and has been associated with castration-resistant prostate cancer in human patients[18]. Here, we observe some of the most significant differences in PHA glycoforms in proteins CD163 (p = 0.034), C4A (p = 0.019), and ATRN (p = 0.007), all of which were over-expressed in the PCa group. ATRN (attractin) has been reported as an important predictor of Gleason score and has been proposed as a diagnostic marker in human sera in a study by Cima *et al*. in a genetics-guided proteomic experiment that looked for differential expression of potential glycoprotein biomarkers that were discovered in Pten cKO mice in the sera of men with localized prostate cancer compared to men with BPH[56]. CD163 has also been associated with prostate cancer. Interestingly, CD163 (hemoglobin scavenger receptor M130), is a secreted glycoprotein expressed by tumor-associated type M2 macrophages, and has been reported to be associated with prostatic inflammation and upregulated in prostate cancer with regard to tumor extension, metastasis, and biochemical recurrence[57–60]. In this study, CD163 was only detected and quantitated in the AAL and PHA fractions (no non- or otherwise glycosylated forms), suggesting the presence of core-fucosylation and extensive branching of complex type glycans. Furthermore, CD163 was significantly increased in the PHA fraction of PCa samples compared to BPH, but was only slightly over-expressed compared to the reference sample (PCA log2H/L = 0.373, BPH log2H/L = −0.610, p = 0.034). The CD163 glycoforms in the AAL fraction were also slightly elevated with respect to the reference sample, but were not statistically different from the BPH glycoforms. Given these data, CD163 glycosylation patterns and overexpression merit further investigation in a larger cohort. It should be noted that although AAL and PHA-L/E lectins are widely used, and their specificity is well documented, the exact glycan structure are not elucidated using this method, and additional glycomic or intact glycopeptide experiments would need to be performed to confirm glycan composition[14]. We have previously published a study that profiled intact glycopeptides derived from plasma glycoproteins, and have confirmed the presence of fucosylated glycopeptides were identified for CD163, C4A, and ATRN, and highly branched glycopeptides were identified for C4BPB and AZGP1[61] in plasma.

Significantly increased expression among core-fucosylated glycoforms was not commonly observed. Most statistically significant difference found in this M-LAC fraction were observed among proteins that were also elevated across all (or most) of the M-LAC fractions, and not to the AAL-bound glycoforms specifically, as shown in the proteins of Group A and D in Fig. 4. Notably, PHA and AAL glycoforms of the protein C4BPB were both elevated relative to the reference sample and were significantly higher in the PCa group than in the BPH group (PHA p = 0.006, AAL p = 0.032), but showed considerable overlap in the log2H/L ratios in the UNB fraction, suggesting increased expression only among these glycoforms. A similar expression pattern was observed in AZGP1.

Although overexpressed proteins arguably make for more definitive measurements in diagnostic biomarkers, this study revealed several proteins that demonstrated decreased expression in the PCa group but striking elevation in the BPH group. Notable among these was the proteoglycan SRGN. SRGN was the most significantly different protein in the UNB fraction, with a three-fold increase in the BPH group relative to the reference sample and a decrease equivalent in magnitude in the PCa group (PCA log2H/L = −1.53, BPH log2H/L = 1.59, p = 0.003). SRGN, or serglycin, has eight o-linked sites carrying glycosaminoglycans and no N-linked glycosylation sites (SRGN was identified exclusively in the UNB fraction). Serglycin has been described as being at the crossroads of inflammation and malignancy, playing a vital role in immune response through interactions with chemokines, cytokines and growth factors[62]. Interestingly, SRGN was only overexpressed in the BPH group in this study, and markedly decreased in the PCa group, suggesting it may be useful in positively identifying cases of inflammation associated with BPH.

In summary, we successfully quantitated hundreds of glycoproteins from PCa and BPH patients to interrogate any differential expression between these two groups. Our approach combined traditional bottom-up proteomics with a series of chromatographic separations to improve the depth of analysis and to systematically screen a complex mixture of proteins for global changes in specific types of glycosylation. Using a multi-lectin approach, our method proved to be useful in separating and quantitating specific protein glycoforms, and yielded an information-rich dataset. By isotopically labeling cysteine-containing proteins, we were able to determine differences in protein expression between the PCa and BPH groups relative to a normal reference sample. Global patterns of differential expression of core-fucosylated or highly-branched glycoforms across the quantitated fraction of the serum proteome were not consistently observed as they have been previously described in glycomic studies that remove the glycan from the protein. In keeping the glycan conjugated to the protein and thus retaining protein-specific information, the breadth of the analysis is subject to the dynamic range and sampling issues associated with mass spectrometry-based proteomics. Upon removing glycans, the differential expression of specific glycan types are cumulative and more readily observed. In this study, we identified alterations among the AAL and PHA glycoforms that were localized to specific, individual proteins rather than observing any larger systemic difference in glycosylation. The expression of core-fucosylation and highly-branched glycoforms was increased in some proteins and decreased in others, with no discernable trend to differentiate the PCa and BPH groups. Among considerably similar proteomic and glycoproteomic signatures, a number of proteins showed significant differences irrespective of glycosylation, including CD5L, CFP, C8A, C7, and BST1. Additionally,

statistically significant glycoform-specific alterations were observed among M-LAC fractions, including SPARC in the UNB fraction, CD163, C4A and ATRN in the PHA fraction, and C4BPB and AZGP1 in both the AAL and PHA fractions. These alterations in protein expression and differential glycosylation patterns may warrant further study in a larger cohort to investigate the reproducibility of our findings, however, a clinically useful biomarker would ideally yield a more dramatic fold-change between PCa and BPH than was observed in any proteins in this study. A strategy for future validation of any glycoproteins of interest may include the use of antibody-lectin assays built to target specific protein glycoforms, such as those use by Sinha, J. et al.[63]. Additionally, further experiments for analysis of intact glycopeptides may also provide more detailed annotation of protein glycosylation status and reveal more specific alterations in glycan structure between PCa and BPH, as lectins, although useful tools for separation, provide limited information regarding glycan structure.

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians* **67**, 7–30, https://doi.org/10.3322/caac.21387 (2017).
2. Welch, H. G. & Albertsen, P. C. Prostate Cancer Diagnosis and Treatment After the Introduction of Prostate-Specific Antigen Screening: 1986–2005. *JNCI: Journal of the National Cancer Institute* **101**, 1325–1329, https://doi.org/10.1093/jnci/djp278 (2009).
3. Howrey, B. T., Kuo, Y.-F., Lin, Y.-L. & Goodwin, J. S. The Impact of PSA Screening on Prostate Cancer Mortality and Overdiagnosis of Prostate Cancer in the United States. *The Journals of Gerontology: Series A* **68**, 56–61, https://doi.org/10.1093/gerona/gls135 (2013).
4. Hayes, J. H. & Barry, M. J. Screening for prostate cancer with the prostate-specific antigen test: A review of current evidence. *JAMA* **311**, 1143–1149, https://doi.org/10.1001/jama.2014.2085 (2014).
5. Andriole, G. L. et al. Prostate Cancer Screening in the Randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: Mortality Results after 13 Years of Follow-up. *JNCI: Journal of the National Cancer Institute* **104**, 125–132, https://doi.org/10.1093/jnci/djr500 (2012).
6. Wilt, T. J. & Dahm, P. PSA Screening for Prostate Cancer: Why Saying No is a High-Value Health Care Choice. *Journal of the National Comprehensive Cancer Network* **13**, 1566–1574, https://doi.org/10.6004/jnccn.2015.0182 (2015).
7. Schröder, F. H. et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *The Lancet* **384**, 2027–2035, https://doi.org/10.1016/S0140-6736(14)60525-0 (2014).
8. Chen, Z., Chen, H. & Stamey, T. A. Prostate Specific Antigen in Benign Prostatic Hyperplasia: Purification and Characterization. *The Journal of Urology* **157**, 2166–2170, https://doi.org/10.1016/S0022-5347(01)64704-9 (1997).
9. Anderson, N. L. & Anderson, N. G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Molecular & Cellular Proteomics* **1**, 845–867, https://doi.org/10.1074/mcp.R200007-MCP200 (2002).
10. Hanash, S. M., Pitteri, S. J. & Faca, V. M. Mining the plasma proteome for cancer biomarkers. *Nature* **452**, 571–579 (2008).
11. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database11Dedicated to Prof. Akira Kobata and Prof. Harry Schachter on the occasion of their 65th birthdays. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1473**, 4–8, https://doi.org/10.1016/S0304-4165(99)00165-8 (1999).
12. Ohtsubo, K. & Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **126**, 855–867, https://doi.org/10.1016/j.cell.2006.08.019 (2006).
13. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* **15**, 540–555, https://doi.org/10.1038/nrc3982 (2015).
14. Varki, A. et al. *Essentials of Glycobiology*. 2nd edn, (Cold Spring Harbor Laboratory Press, 2009).
15. Adamczyk, B., Tharmalingam, T. & Rudd, P. M. Glycans as cancer biomarkers. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1820**, 1347–1353, https://doi.org/10.1016/j.bbagen.2011.12.001 (2012).
16. Pierce, M. & Taniguchi, N. Glycan BiomarkerDiscovery. *PROTEOMICS – Clinical Applications* **7**, 595–596, https://doi.org/10.1002/prca.201370054 (2013).
17. Ruhaak, L. R., Miyamoto, S. & Lebrilla, C. B. Developments in the Identification of Glycan Biomarkers for the Detection of Cancer. *Molecular & Cellular Proteomics* **12**, 846–855, https://doi.org/10.1074/mcp.R112.026799 (2013).
18. Munkley, J., Mills, I. G. & Elliott, D. J. The role of glycans in the development and progression of prostate cancer. *Nat Rev Urol* **13**, 324–333, https://doi.org/10.1038/nrurol.2016.65 (2016).
19. Gilgunn, S., Conroy, P. J., Saldova, R., Rudd, P. M. & O'Kennedy, R. J. Aberrant PSA glycosylation[mdash]a sweet predictor of prostate cancer. *Nat Rev Urol* **10**, 99–107 (2013).
20. Peracaula, R. et al. Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins. *Glycobiology* **13**, 457–470, https://doi.org/10.1093/glycob/cwg041 (2003).
21. Vermassen, T., Speeckaert, M. M., Lumen, N., Rottey, S. & Delanghe, J. R. Glycosylation of prostate specific antigen and its potential diagnostic applications. *Clinica Chimica Acta* **413**, 1500–1505, https://doi.org/10.1016/j.cca.2012.06.007 (2012).
22. Yoneyama, T. et al. Measurement of aberrant glycosylation of prostate specific antigen can improve specificity in early detection of prostate cancer. *Biochemical and Biophysical Research Communications* **448**, 390–396, https://doi.org/10.1016/j.bbrc.2014.04.107 (2014).
23. Llop, E. et al. Improvement of Prostate Cancer Diagnosis by Detecting PSA Glycosylation-Specific Changes. *Theranostics* **6**, 1190–1204, https://doi.org/10.7150/thno.15226 (2016).
24. Drake, R. R., Jones, E. E., Powers, T. W. & Nyalwidhe, J. O. In *Advances in Cancer Research* Vol. 126 (eds Richard R. Drake & Lauren E. Ball) 345–382 (Academic Press, 2015).
25. Totten, S. M., Kullolli, M. & Pitteri, S. J. In *Proteomics: Methods and Protocols* (eds Lucio Comai, Jonathan E. Katz, & Parag Mallick) 99–113 (Springer New York, 2017).
26. Kullolli, M., Warren, J., Arampatzidou, M. & Pitteri, S. J. Performance evaluation of affinity ligands for depletion of abundant plasma proteins. *Journal of Chromatography B* **939**, 10–16, https://doi.org/10.1016/j.jchromb.2013.09.008 (2013).
27. Faca, V. et al. Quantitative Analysis of Acrylamide Labeled Serum Proteins by LC−MS/MS. *Journal of Proteome Research* **5**, 2009–2018, https://doi.org/10.1021/pr060102+ (2006).
28. Gbormittah, F. O., Hincapie, M. & Hancock, S. W. Development of an improved fractionation of the human plasma proteome by a combination of abundant proteins depletion and multi-lectin affinity chromatography. *Bioanalysis* **6**, 2537–2548, https://doi.org/10.4155/bio.14.217 (2014).
29. Zeng, Z. et al. A Proteomics Platform Combining Depletion, Multi-lectin Affinity Chromatography (M-LAC), and Isoelectric Focusing to Study the Breast Cancer Proteome. *Analytical Chemistry* **83**, 4845–4854, https://doi.org/10.1021/ac2002802 (2011).
30. Kullolli, M., Hancock, W. S. & Hincapie, M. Preparation of a high-performance multi-lectin affinity chromatography (HP-M-LAC) adsorbent for the analysis of human plasma glycoproteins. *Journal of Separation Science* **31**, 2733–2739, https://doi.org/10.1002/jssc.200800233 (2008).
31. Yang, Z. & Hancock, W. S. Monitoring glycosylation pattern changes of glycoproteins using multi-lectin affinity chromatography. *Journal of Chromatography A* **1070**, 57–64, https://doi.org/10.1016/j.chroma.2005.02.034 (2005).

32. Drake, P. M. *et al*. A lectin affinity workflow targeting glycosite-specific, cancer-related carbohydrate structures in trypsin-digested human plasma. *Analytical Biochemistry* **408**, 71–85, https://doi.org/10.1016/j.ab.2010.08.010 (2011).

33. Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534–2536, https://doi.org/10.1093/bioinformatics/btn323 (2008).

34. Craig, R. & Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467, https://doi.org/10.1093/bioinformatics/bth092 (2004).

35. Rauch, A. *et al*. Computational Proteomics Analysis System (CPAS): An Extensible, Open-Source Analytic System for Evaluating and Publishing Proteomic Data and High Throughput Biological Experiments. *Journal of Proteome Research* **5**, 112–121, https://doi.org/10.1021/pr0503533 (2006).

36. Nelson, E. K. *et al*. LabKey Server: An open source platform for scientific data integration, analysis and collaboration. *BMC Bioinformatics* **12**, 71, https://doi.org/10.1186/1471-2105-12-71 (2011).

37. Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Analytical Chemistry* **74**, 5383–5392, https://doi.org/10.1021/ac025747h (2002).

38. Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Analytical Chemistry* **75**, 4646–4658, https://doi.org/10.1021/ac0341261 (2003).

39. Fay, M. P. & Shaw, P. A. Exact and Asymptotic Weighted Logrank Tests for Interval Censored Data: The interval R package. *Journal of statistical software* **36**, i02 (2010).

40. Yu, H., Yu, J., Ren, Y., Yang, Y. & Xiao, X. Serum CEACAM1 Level Is Associated with Diagnosis and Prognosis in Patients with Osteosarcoma. *PLoS ONE* **11**, e0153601, https://doi.org/10.1371/journal.pone.0153601 (2016).

41. Farrah, T. *et al*. A High-Confidence Human Plasma Proteome Reference Set with Estimated Concentrations in PeptideAtlas. *Molecular & Cellular Proteomics* **10**, https://doi.org/10.1074/mcp.M110.006353 (2011).

42. Sörensen-Zender, I. *et al*. Zinc-alpha2-glycoprotein in patients with acute and chronic kidney disease. *BMC Nephrology* **14**, 145, https://doi.org/10.1186/1471-2369-14-145 (2013).

43. Kelly-Spratt, K. S. *et al*. Plasma Proteome Profiles Associated with Inflammation, Angiogenesis, and Cancer. *PLOS ONE* **6**, e19721, https://doi.org/10.1371/journal.pone.0019721 (2011).

44. Hanahan, D. & Weinberg, Robert A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674, https://doi.org/10.1016/j.cell.2011.02.013.

45. Chughtai, B., Lee, R., Te, A. & Kaplan, S. Role of Inflammation in Benign Prostatic Hyperplasia. *Reviews in Urology* **13**, 147–150 (2011).

46. Nickel, J. C. Inflammation and Benign Prostatic Hyperplasia. *The Urologic clinics of North America* **35**, 109–115, https://doi.org/10.1016/j.ucl.2007.09.012 (2008).

47. Janeway, C. A. Jr. *et al*. *Immunobiology: The Immune System in Health and Disease*. 5 edn, (Garland Science, 2001).

48. Tai, I. T. & Tang, M. J. SPARC in cancer biology: Its role in cancer progression and potential for therapy. *Drug Resistance Updates* **11**, 231–246, https://doi.org/10.1016/j.drup.2008.08.005 (2008).

49. Chen, J. *et al*. SPARC Is a Key Regulator of Proliferation, Apoptosis and Invasion in Human Ovarian Cancer. *PLOS ONE* **7**, e42413, https://doi.org/10.1371/journal.pone.0042413 (2012).

50. Mateo, F. *et al*. SPARC mediates metastatic cooperation between CSC and non-CSC prostate cancer cell subpopulations. *Molecular Cancer* **13**, 237, https://doi.org/10.1186/1476-4598-13-237 (2014).

51. Thomas, R., True, L. D., Bassuk, J. A., Lange, P. H. & Vessella, R. L. Differential Expression of Osteonectin/SPARC during Human Prostate Cancer Progression. *Clinical Cancer Research* **6**, 1140–1149 (2000).

52. Sharma, S. *et al*. Secreted Protein Acidic and Rich in Cysteine (SPARC) Mediates Metastatic Dormancy of Prostate Cancer in the Bone. *Journal of Biological Chemistry*, https://doi.org/10.1074/jbc.M116.737379 (2016).

53. De, S. *et al*. Molecular Pathway for Cancer Metastasis to Bone. *The Journal of biological chemistry* **278**, 39044–39050, https://doi.org/10.1074/jbc.M304494200 (2003).

54. Feng, J. & Tang, L. SPARC in Tumor Pathophysiology and as a Potential Therapeutic Target. *Current Pharmaceutical Design* **20**, 6182–6190 (2014).

55. Vaz, J., Ansari, D., Sasor, A. & Andersson, R. SPARC: A Potential Prognostic and Therapeutic Target in Pancreatic Cancer. *Pancreas* **44**, 1024–1035, https://doi.org/10.1097/MPA.0000000000000409 (2015).

56. Cima, I. *et al*. Cancer genetics-guided discovery of serum biomarker signatures for diagnosis and prognosis of prostate cancer. *Proceedings of the National Academy of Sciences* **108**, 3342–3347, https://doi.org/10.1073/pnas.1013699108 (2011).

57. Lundholm, M. *et al*. Secreted Factors from Colorectal and Prostate Cancer Cells Skew the Immune Response in Opposite Directions. *Scientific Reports* **5**, 15651, https://doi.org/10.1038/srep15651 (2015).

58. Lanciotti, M. *et al*. The Role of M1 and M2 Macrophages in Prostate Cancer in relation to Extracapsular Tumor Extension and Biochemical Recurrence after Radical Prostatectomy. *BioMed Research International* **2014**, 6, https://doi.org/10.1155/2014/486798 (2014).

59. Yamamichi, F., Shigemura, K., Arakawa, S., Tanaka, K. & Fujisawa, M. CD-163 correlated with symptoms (pain or discomfort) of prostatic inflammation. *International Journal of Clinical and Experimental Pathology* **8**, 2408–2414 (2015).

60. Topf, M. C., Tuluc, M., Harshyne, L. A. & Luginbuhl, A. Macrophage type 2 differentiation in a patient with laryngeal squamous cell carcinoma and metastatic prostate adenocarcinoma to the cervical lymph nodes. *Journal for ImmunoTherapy of Cancer* **5**, 60, https://doi.org/10.1186/s40425-017-0264-z (2017).

61. Totten, S. M., Feasley, C. L., Bermudez, A. & Pitteri, S. J. Parallel Comparison of N-Linked Glycopeptide Enrichment Techniques Reveals Extensive Glycoproteomic Analysis of Plasma Enabled by SAX-ERLIC. *Journal of Proteome Research* **16**, 1249–1260, https://doi.org/10.1021/acs.jproteome.6b00849 (2017).

62. Korpetinou, A. *et al*. Serglycin: At the Crossroad of Inflammation and Malignancy. *Frontiers in Oncology* **3**, 327, https://doi.org/10.3389/fonc.2013.00327 (2013).

63. Sinha, J. *et al*. A Gastric Glycoform of MUC5AC Is a Biomarker of Mucinous Cysts of the Pancreas. *PLOS ONE* **11**, e0167070, https://doi.org/10.1371/journal.pone.0167070 (2016).

Acknowledgements

Author Contributions

S.M.T. contributed to experimental design, performed experiments, analyzed and interpreted the data, and wrote the manuscript. M.K. contributed to experimental design and initial sample preparation. R.A.

performed statistical analysis and contributed to data presentation and visualization. C.T. contributed to sample preparation. J.D.B. contributed to study design, selected and provided clinical samples, and gave feedback on data interpretation. P.M. contributed to the statistical analysis, data presentation, and visualization. S.J.P. contributed to study design, data generation, analysis, and interpretation. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-24270-w.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.