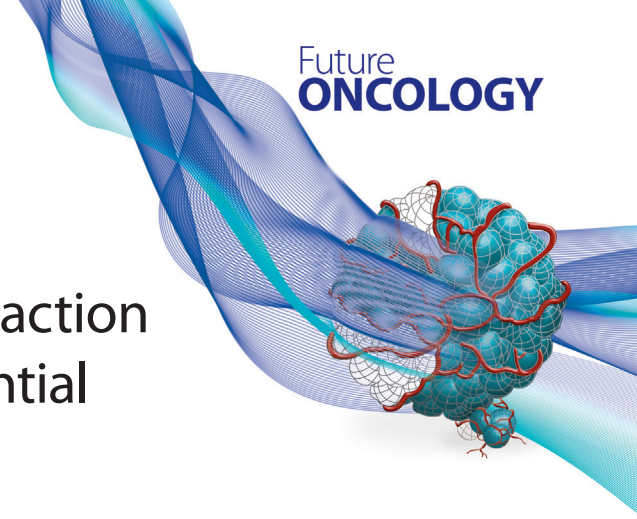
RESEARCH ARTICLE

For reprint orders, please contact: reprints@futuremedicine.com



In silico verification and parallel reaction monitoring prevalidation of potential prostate cancer biomarkers

Henry A Adeola[1,2], Bridget Calder[2], Nelson C Soares[2], Lisa Kaestner[3], Jonathan M Blackburn*[,2] & Luiz F Zerbini**[,1,2]

Purpose: Targeted proteomics of potential biomarkers is often challenging. Hence, we developed an intermediate workflow to streamline potential urinary biomarkers of prostate cancer (PCa). **Materials & methods:** Using previously discovered potential PCa biomarkers; we selected proteotypic peptides for targeted validation. Preliminary *in silico* immunohistochemical and single reaction monitoring (SRM) verification was performed. Successful PTPs were then prevalidated using parallel reaction monitoring (PRM) and reconfirmed in 15 publicly available databases. **Results:** Stringency-based targetable potential biomarkers were shortlisted following *in silico* screening. PRM reveals top 12 potential biomarkers including the top ranking seven *in silico* verification-based biomarkers. Database reconfirmation showed differential expression between PCa and benign/normal prostatic urine samples. **Conclusion:** The pragmatic penultimate screening step, described herein, would immensely improve targeted proteomics validation of potential disease biomarkers.

First draft submitted: 4 August 2015; Accepted for publication: 22 October 2015; Published online: 30 November 2015

KEYWORDS
• biomarkers • *in silico* • parallel reaction monitoring • prostate cancer • targeted proteomics

Prostate cancer (PCa) is the leading cause of cancer deaths and the most frequently diagnosed cancer in Africa, ahead of liver and lung cancer which are in second and third places, respectively [1]. In spite of this high burden, it receives suboptimal public health attention in Africa due to a concurrently high burden of infectious diseases like HIV/AIDS, tuberculosis (TB) and malaria. Putative diagnostic biomarkers like prostate-specific antigen (PSA) albeit highly useful in conjunction with other clinical tests, have fallen short in the lower reference ranges (2–10 ng/ml) in terms of specificity and negative predictive values [2]. Additionally, why men of African descent more frequently tend to carry aggressive phenotypes of PCa compared with other ethnicities is poorly understood. A few emerging molecular diagnostic biomarkers with clinical promise for risk stratification, treatment response and recurrence prediction include *PCA3*, *TMPRSS2–ERG* fusion gene, *PTEN*, *E-Cadherin* and *EZH2* [3]. Considering the low socioeconomic status of most African patients, minimally invasive and affordable novel biomarkers form early diagnosis and treatment monitoring is most desirable.

Currently, the normative sequence of event in biomarker discovery pipelines is validating discovery based potential biomarkers with targeted proteomics method. However, validating multiple candidate biomarkers from high-throughput shotgun discovery proteomics can be quite

[1]International Centre for Genetic Engineering & Biotechnology, Cape Town, South Africa
[2]Institute of Infectious Diseases & Molecular Medicine, Faculty of Health Sciences, University of Cape Town, South Africa
[3]Urology Department, Grootes Schuur Hospital, Cape Town, South Africa
*Author for correspondence: jonathan.blackburn@uct.ac.za
**Author for correspondence: luiz.zerbini@icgeb.org

Future Medicine part of fsg

challenging. Considering *inter alia*, the number of 'proteotypic peptides' (PTPs), charge states and replicates needed for each biomarker. An important drawback of biomarker research has been the paucity of robust highly multiplexed quantitative methods to measure hundreds of analytes in the shortest time possible. Hence, many promising biomarkers have been discovered, however only a few are clinically useful or implementable [4–7]. With the emergence of novel high throughput omics-based technologies, there is a concomitant need for better *in silico* computational and bioinformatics tools to improve clinical inferences drawn from these huge databases generated [8]. Despite the major advances made in statistical software, workflows and algorithms for analysis and absolute quantification of targeted proteomics data, multiplexing and analysis of large numbers of potential biomarkers can be an arduous task. Hence adjunctive databases use, offer researchers a platform to address a host of biological information generated from proteomics experiments in a seamless manner.

Broadly speaking, databases for cancer proteomics research can be classified into five groups [9] which are; gene/protein expression, gene mutation and SNP, tumor antigen, cancer-associated genes and protein interaction/pathway databases. A gamut of proteomics databases have been emerging for translation of systems biology data into useable diagnostic and therapeutic tools such as CancerResource [10], Global Proteome Machine Database (GPMDB) [11], Yale Protein Expression Database (YPED) [12], NeXtProt [13] and Proteomic Identifications database (PRIDE) [11]. The Human Protein Atlas [14] is an online resource using antibody-centered proteomics to create an atlas of Ca. 400,000 high quality images of expression level and localization characteristics of over 700 human proteins in 48 and 20 different normal human and cancerous tissues, respectively [14]. Careful antibody design with rigorous recombinant affinity purification protein epitope signature tags (PrESTs) [14,15]; using only antibodies with reduced sequence homology to other human proteins and increased tissue specificity. In addition, all antibodies were tested on human tissue microarray and computationally analyzed for protein expression, localization and where possible gene expression and transcript level. Following stringent quality assurance protocols, validation scores were assigned to antibodies using bioinformatics comparison between experimental data and literature based data [14].

High-throughput *in silico* identification of peptides generated from tandem mass spectrometry experiments, algorithmically mapped to eukaryotic cell genome sequence has been made easier by the Peptide Atlas Project [16–18]. This database evolved to incorporate targeted proteomics data in its repositories [17,18]. An essential requirement of targeted proteomics provided by PeptideAtlas is large scale quantitation and compilation of proteins and PTPs across multiple experiments [17]. In 2012, the PeptideAtlas initiated a new repository of single reaction monitoring (SRM) experiments known as The PeptideAtlas SRM Experiment Library (PASSEL) which allows researchers submit and access targeted proteomics datasets generated through SRM experiments [19]. This database was also found expedient for optimizing the required number of unique transition for each peptide; using the SRMCollider software in tandem with another newly synthesized extensive SRM database (SRMAtlas) [20]. In addition, this method can accurately replicate the precision of emerging data independent acquisition (DIA) mass spectrometry methods like SWATH that combine high-throughput with consistent reproducibility. Various data builds have been established in SRMAtlas, for example, N-Glycoproteome for different cancer types have been recapitulated across different dataset in SRMAtlas [21]. Without doubt, PeptideAtlas, PASSEL and SRMAtlas are highly useful web-based resources for potential biomarker verification [22].

Traditionally, a triple quadrupole (QQQ) mass spectrometer is described as the workhorse of a targeted proteomics method known as the Single Reaction Monitoring (SRM) in which the 1st and 3rd quadrupoles as specified *m/z* filters, while the second Quadrupole acts as a collision cell [23]. To overcome some of the setbacks of SRM, Parallel Reaction Monitoring (PRM) using a high resolution, high mass accuracy hybrid Quadrupole-Orbitrap mass spectrometer such as the QExactive™ has been employed. This instrument allows a highly multiplexed simultaneous identification and quantitation of multiple transitions in a single run. The *modus operandi* is in some respects similar to the QQQ, albeit the third Q is replaced by an Orbitrap mass analyzer [24,25]. PRM has been reported

to provide comparable performance metrics with SRM in terms of precision, linearity and dynamic range [26]. We have provided herein, a judicious intermediate step in the biomarker discovery and validation pipeline using a combination of *in silico* database verification and PRM prevalidation. This could represent an essential penultimate screening step prior to definitive targeted validation.

Materials & methods

• Sample source

Urine samples were collected from PCa and benign prostatic hyperplasia (BPH) patients as well as normal healthy (NC) individuals attending urology clinic at Grootes Schuur Hospital and two satellite hospitals in Cape Town with full ethical consent (HREC 454/2012). These samples were processed routinely for shotgun discovery proteomics and analyzed using ultra-HPLC and mass spectrometry. Postanalytic statistical evaluation revealed a total of 73 potential PCa biomarkers as well as nine biomarkers which demonstrated potential ethnic trends in South Africa [27]. Two pooled samples were prepared for PRM using peptides from 15 PCa samples and 15 normal control samples, respectively. All PCa patients used for this study had localized primary disease (≤TNM stage III) and evidence of metastatic disease was an exclusion criterion. Detailed selection criteria and clinicopathologic features such as age, race, PSA level and Gleason scores of the individual PCa patients used in this study can be found in our previously published work [27] and shown in here in the online supplementary materials **(Supplementary Table 1)**. Patient who would be undergoing Transurethral resection of prostate (TURP), without history of cancer at any other site or history of any other major comorbidities like essential arterial hypertension or diabetes were selected for the study. Patients who were unwilling to participate due to religious or cultural reasons were excluded as well. Patients were drawn from a heterogeneous cohort of South African Prostate cancer patients made of Indigenous black African, Caucasian South African and mixed ancestry patients. Prior to PRM experiments, *in silico* methods were employed to identify top ranking potential PCa biomarkers; after which they were evaluated retrospectively in our shotgun database as well as other urinary proteomics databases **(Figure 1)**.

• Human Protein Atlas database biomarker verification

Characterization of 82 potential biomarkers was performed using the Human Protein Atlas (HPA) database. Immunohistochemical procedure was carried out in a high throughput and highly automated manner as described in detail elsewhere [14]. In brief, standardized immunohistochemical analysis was performed on a specially designed tissue microarray (TMA) and optimized with multiple antibody dilutions. Both positive and negative controls were spotted on the recipient TMA blocks. Antigen retrieval was performed in citrate buffer (pH 6.0) using pressure boiler heat method. Optimized monospecific primary antibody dilution was used for each unique antibody and secondary antibodies were selected in a host species dependent manner. Slides were first developed with diaminobezindine and then counterstained with hematoxylin. Immunohistochemical images were collected with an automated slide scanning system at 40× magnification. Using web-based annotation software, the images are scored by calibrated certified histopathologists. Differential immunohistochemical staining of cancerous and normal prostate tissue was assessed by searching either the name or the gene symbol of the protein of interest. Expression levels of protein and target mRNA in cancerous and normal prostate tissue and human cell lines, respectively, were assessed, albeit it is well established that mRNA levels does not always correlate to protein expression levels due to factors such as translational control and posttranslational modification. In addition, subcellular location of proteins can be verified from this database. Taking into account that many of our previously discovered biomarkers were predicted membrane-bound or extracellular matrix (ECM), we took into consideration membrane and extracellular/stromal staining. Potential biomarkers with differential staining in cancerous and normal prostate tissue were considered for further targeted proteomics analysis.

• PTP selection

The specificity of selected peptide for target protein is largely dependent on its fragmentation pattern. Shorter unmodified peptides unique to a specific or single isoform of the target protein with imino acid proline are generally better SRM targets than poorly ionized longer peptides. The permissible charge states were +2 and

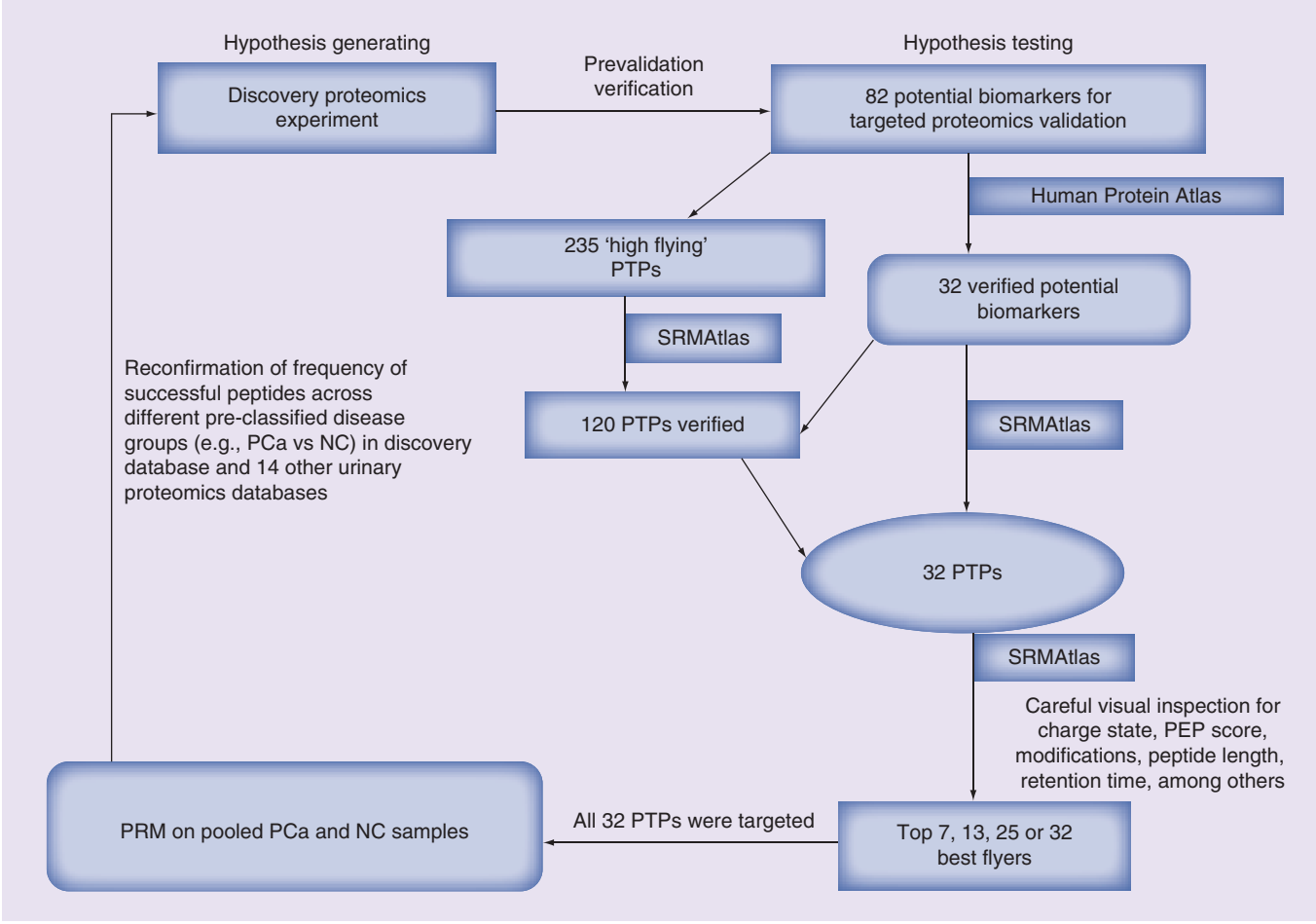


Figure 1. Experimental workflow for potential prostate cancer biomarker prevalidation. Discovered potential biomarkers from previous shotgun proteomics experiments were subjected to two steps of *in silico* screening *viz* SRMAtlas and Human Protein Atlas databases. Top ranking potential biomarkers from these *in silico* steps were experimentally prevalidated using PRM and potential PCa biomarker frequency is reconfirmed in shotgun databases for differential expression between PCa and healthy controls.
NC: Normal healthy; PCa: Prostate cancer; PEP: Posterior error probability; PRM: Parallel reaction monitoring; PTP: Proteotypic peptide

+3 for each PTP. Peptides containing too many units of tryptophan (W) or methionine (M) were avoided because of the high propensity of artefactual side chain oxidation. Depending on adjacent sequences, glutamine (Q) and asparagine (N) were carefully considered due to their chemical instability; particularly N-terminal glutamine can change to pyroglutamate when exposed to acidic treatment. Maximum missed cleavages permitted per peptide were two. The chances of having missed cleavages are known to be increase in sequences containing two adjacent terminal basic amino acids such as -KR, -KK or -RR. Due to low ion current in comparison to their tryptic peptide counterpart, nontryptic peptides are generally avoided in SRMs, albeit they can be used where no good tryptic peptides are available. Carefully following the above-mentioned considerations, three PTPs were selected per protein from the evidence file (**txt*) generated from the Maxquant analysis of raw Xcalibur data. Less than three PTPs were accepted for proteins with fewer identified peptides from shotgun assays.

• SRMAtlas database verification of PTPs

SRMAtlas is a compendium of high quality SRM assays for identification and quantitation of proteins. This database has identified over 170,000 human proteome peptides with Ca. 99.9% coverage for human proteins [19]. Our *a priori* expectation is that previously assayed and documented peptides in the SRMAtlas database represent a reliable cohort of targetable peptides, particularly if evident in our shotgun proteomics experiments. All selected PTPs from our

previous shotgun proteomics assay were searched against the SRMAtlas database using the following 'search SRM Assays' parameters. After filling the protein accession identification number, 'Ion Trap' was selected as 'transition sources' because of its similarity with the Orbitrap mass analyzer. Specified m/z range was 200–2000 m/z while only y- and b- ions were permitted without neutral loss. SwissProt, Ensembl and IPI were the selected target databases and default settings were accepted for all other parameters before activating the query.

• Pilot Parallel reaction monitoring

Following verification of discovery MS-based candidate biomarker PTPs with Human Protein Atlas databases and SRMAtlas, selected 'best flying' PTPs were further prevalidated by PRM. Initial PRM runs on the QExactive Mass spectrometer is aimed at narrowing the list of candidate biomarkers. A spectral library is first created using data from the original shotgun proteomics assay, followed by isolation list generation in Skyline (MacCoss Lab software) an open sources tool for targeted proteomics analysis [28]. Isolation list generation involves creating a randomized order of precursors, ensuring that no subsequent precursor *m/z* values are similar as well as randomizing the order of +2 or +3 charge states of the same precursor ions. These experiments were not scheduled because observed retention time range varied quite widely across all potential PCa biomarkers **(Supplementary Figure 1)**. Samples were singly injected because we did not observe much variability in multiply injected technical replicates in our previous PRM experiments on urine samples of HIV-infected patients, possibly due to the high resolution and mass accuracy of the instrument used.

Ultra-HPLC

We ensured that liquid chromatography gradients and column used in these pilot PRMs were similar to those used in the shotgun discovery proteomics experiments. Pooled PCa (n = 15) and NC (n = 15) samples at 50 ng/µl in HPLC grade water containing 0.1% (v/v) Formic acid were subjected to Nanoflow ultra-HPLC inline on a Dionex UltiMate® 3500 RSnano UPLC system (Thermo Fisher, CA, USA) equipped with a 100 µm × 5 cm; 5 µm; 100Å; C-18 precolumn and a 75 µm × 50 cm; 5 µm; 100Å; C-18 analytic column. Temperature for gradient chromatography was set at 23°C using a flow rate of 300 nl/min. Peptides elution spanned over 0–180 min at a 5–80% water to acetonitrile gradient. The binary mobile phase system used was as follows: buffer A contained water and 0.1% formic acid, while buffer B contained acetonitrile and 0.1% formic acid. Elution gradient for peptides was 5% B from 0–20 min, then increasing to 80% B over 180 min. Following each run, the flow rate was increased to 450 nl/min at 50% B to equilibrate the analytic column and then dropped to 300 nl and 5% B again prior to the next sample run.

QExactive quadrupole-orbitrap analyzes

Parallel reaction monitoring was carried out on a QExactive Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher) which combines high-resolution quadrupole precursor ion selection with accurate mass Orbitrap detection. For the nanoelectrospray ionization source, an ionization voltage of 1.86 kV and spray current of 0.1 µA was used at capillary temperature of 320°C and S-lens RF level of 50.0. We used an unscheduled *4- plex* MS2 targeted methodology at a mass spectra acquisition resolution of 35,000. Full scan All Ion Fragmentation (AIF) mode was used at a scan range of 79–1945 m/*z* with positive polarity, maximum time per peptide of 30 ms and charge exclusion z = 2. Requisite cycle time was 2 s with a total scan time of 140 ms. Automatic gain control (AGC) target was fixed at 5 × E6. Normalized Collision Energy (NCE) set at 27eV with in-source High-energy Collision Dissociation (HCD) was used for peptide fragmentation. The method of acquisition comprises of targeted PRM scan events directed at the +2 and +3 charge states of the PTP precursor. An isolation window of 2 *m/z* units was used for each precursor ion at a starting *m/z* of 80 and final *m/z* is automatically derived from the charge state and *m/z* of the PTP precursor. Once MS2 data is generated by PRM, raw files were imported into Skyline for further analyzes. This is repeated with the isolation list in a different order to confirm that interference from multiplexing had not biased our results and further quantitative analysis to compare biomarker transitions between pooled PCa and Normal control samples.

• Discovery database verification of successful biomarkers

Successful candidate biomarkers using HPA, SRMAtlas and PRM were identified and

assessed in previous discovery shotgun analysis. These were also searched against other urinary proteomics databases to assess their differential expression and ability to distinguish prostate cancer from normal healthy individuals or patients with benign prostatic hyperplasia.

Results

• Potential biomarker screening

We screened 82 previously discovered potential urinary protein biomarkers of PCa. Even though we aimed at 3 PTPs selection per protein, we found a total of 235 PTPs (Supplementary Table 2). Out of 235 PTPs, a total of 115 (48.9%) PTPs have not been previously reported; while there were 120 (51.1%) PTP found in the SRMAtlas database. Considering peptide length, posterior error probability (PEP) score, charge state, retention time range, modification status, HPA analysis and SRMAtlas verification, the list of 82 biomarkers were downsized to 32 top ranking potential biomarkers which were found most suitable for targeted proteomics analysis (Table 1). Some potential protein biomarkers were not found in the Human Protein Atlas database (Supplementary Table 2) and new unreported PTPs in SRMAtlas were found. An overview of the workflow which involved *in silico* verification steps, experimental PRM prevalidation, and database reconfirmation of the top ranking biomarkers is presented in Figure 1.

• Immunohistochemical patterns for potential biomarkers

Differential immunohistochemical staining patterns between cancerous and normal prostate tissue were found in 32 of the 82 potential biomarkers from shotgun experiment, using the human protein atlas database. We focused on the top ranking seven potential biomarkers (Table 1) and prospectively, five other good biomarkers by PRM (Supplementary Figure 2). For PGLYRP2, there is heavy staining of the ductal acinar cells and nonspecific staining of the glandular stroma of PCa tissue in contrast to a low level of staining of the ductal epithelium and nonspecific staining of the stroma in normal prostate tissue. HPR demonstrated a medium staining with PCa tissue, while the normal tissue stained negative (Figure 2A & E). PROS1 showed moderate staining of cancerous prostate tissues and low staining with normal prostate tissue (Figure 2B & F). Both normal and cancerous prostate tissues stained positive for CPN1 albeit stronger positivity was seen in PCa tissue. PZP was diffusely positive for cancerous prostate tissue and light stromal staining was observed for normal tissue. There was distinct positivity for ACTN1 in PCa while normal prostate tissue stained completely negative. Similar pattern of staining was seen in MYOC which stained positive for PCa and negative for normal tissue. ACPP could not distinguish immunohistochemically between PCa and normal tissue (Figure 2C & H). KLK3 could reasonably distinguish between PCa and normal prostate tissue which stained positive and negative, respectively (Figure 2D & G). There were nonspecific staining patterns for NID1 both for PCa and normal prostate tissue; however, the normal tissue demonstrated more stromal staining. Cancerous prostatic tissue stained positively for CTSZ while there was no signal found for CTSZ in normal prostate tissue. SLAIN1 stained remarkably for PCa tissue in comparison to mild stromal staining of normal prostate tissue (Supplementary Figure 2).

• SRMAtlas biomarkers PTP evaluation

Using this database, all 32 top ranking potential biomarkers with differential expression in the HPA were analyzed, and we selected the best out of three PTPs per potential biomarker. Detected y- and b-ion transitions are automatically ranked by the database and best flying transitions are easy to detect. PTP mass and charges both at the first (Q1) and third (Q3) quadrupole mass analyzers are shown. The preceding amino acid towards the N-terminus (pre-AA) and the one following towards the C-terminus (Fol-AA) are reported. Adjusted suitability score (AdjSS) of transition which is derived from weighting of the predicted and empirical suitability score is also reported. Relative intensities (RI) of peaks in the collision-induced dissociation (CID) spectra and sequence specific retention times (SSRT) for hydrophobicity assessment is provided. Frequency of peptide mapping (N-Map) to protein in target proteome is also available in the results generated. For instance, PTPs for HPR, PROS1, ACPP and KLK3 were found as illustrated in SRMAtlas results (Supplementary Table 3).

• Preliminary targeted proteomics

PRM experiments were used to prevalidate the selected 32 high ranking biomarkers from *in silico* verification using HPA and SRMAtlas. We found differential expression of peptide

Table 1. Pipeline for top ranking potential PCa biomarkers selection for targeted proteomics validation.

Protein name	Gene name	PL	Sequence	Presence in PCa	Presence in NC	Level 1 stringency	Level 2 stringency	Level 3 Stringency	PRM pooled prevalidation
Flavin reductase (NADPH)	*BLVRB*	15	TVAGQDAVIVLLGTR	++	+	Yes	No	No	
Prostatic acid phosphatase	*ACPP*	12	SPIDTFPTDPIK	+	+	No	No	No	
N-acetylmuramoyl-L-alanine amidase	*PGLYRP2*	13	TDCPGDALFDLLR	+++	--	Yes	Yes	Yes	Good
Epididymal secretory protein E1	*NPC2*	16	EVNVSPCPTQPCQLSK	+	++	Yes	No	No	
Glutaredoxin-1	*GLRX*	21	DCIGGCSDLVSLQQSGELLTR	++	+	Yes	No	No	
CD59 glycoprotein	*CD59*	16	TAVNCSSDFDACLITK	-	+	Yes	Yes	No	
Trefoil factor 3	*TFF3*	11	VDCGYPHVTPK	+	-	Yes	Yes	No	
Nonsecretory ribonuclease	*RNASE2*	17	YAQTPANMFYIVACDNR	+	-	Yes	Yes	No	
Prostate-specific antigen	*KLK3*	12	LSEPAELTDAVK	+	-	Yes	Yes	No	
Proactivator polypeptide;Saposin	*PSAP*	11	GCSFLPDPYQK	+	+	No	No	No	
Vitamin K-dependent protein S	*PROS1*	8	HCLVTVEK	++	-	Yes	Yes	Yes	Good
Haptoglobin-related protein	*HPR*	10	VTSIQHWVQK	++	-	Yes	Yes	Yes	Good
Nidogen-1	*NID1*	10	EDLSPSITQR	+	++	Yes	No	No	
Serum paraoxonase/arylesterase 1	*PON1*	16	IFFYDSENPPASEVLR	++	+	Yes	No	No	
Plasma kallikrein	*KLKB1*	14	IAYGTQGSSGYSLR	++	+	Yes	No	No	
Lactotransferrin	*LTF*	12	LADFALLCLDGK	++	+	Yes	No	No	
Inter-alpha-trypsin inhibitor heavy chain H1	*ITIH1*	13	QAVDTAVDGVFIR	+	+	No	no	No	
Carboxypeptidase N catalytic chain	*CPN1*	9	IVQLIQDTR	+++	+	Yes	Yes	Yes	Good
Inter-alpha-trypsin inhibitor heavy chain H3	*ITIH3*	22	STSIVIMLTDGDANVGESRPEK	++	++	No	No	No	
Pregnancy zone protein	*PZP*	10	SLFTDLVAEK	+++	+	Yes	Yes	Yes	Good
Alpha-actinin-1	*ACTN1*	14	FAIQDISVEETSAK	+++	-	Yes	Yes	Yes	Good
Plastin-2	*LCP1*	9	NEALIALLR	+	-	Yes	Yes	No	
Collagen alpha-1(XII) chain	*COL12A1*	12	ITVDPTTDGPTK	++	+	Yes	No	No	
Tenascin	*TNC*	16	CECDDGFTGADCGELK	++	+	Yes	No	No	
Cadherin-11	*CDH11*	12	FFTINPEDGFIK	-	+	Yes	Yes	No	
Cathepsin Z	*CTSZ*	12	NVDGVNYASITR	++	+	Yes	No	No	
Alpha-N-acetylglucosaminidase	*NAGLU*	16	LLGPGPAADFSVSVER	++	+++	Yes	No	No	
Proteasome inhibitor PI31 subunit	*PSMF1*	17	AGLEVLFASAAPAITCR	++	++	No	No	No	
Heat shock protein HSP 90-beta	*HSP90AB1*	15	NPDDITQEEYGEFYK	++	++	No	No	No	
Myocilin	*MYOC*	21	SGEGDTGCGELVWVGEPLTLR	++	-	Yes	Yes	Yes	Good
SLAIN motif-containing protein 1	*SLAIN1*	7	LQEESLR	++	+++	Yes	No	No	
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	*MAN1A1*	18	EEGAPGDPEAALEDNLAR	++	++	No	No	No	
				Total number of Biomarkers		25	13	7	

Level 1 stringency: Any difference in expression.
Level 2 stringency: Favor low (-)/high (+) and 'extrememly'.
Level 3 stringency: Greater than 2 units of difference + favor 'extremely'.
-: Low; --: Very low; +: High; ++: Very high; +++: Extremely high; NC: Normal healthy; PL: Protein length.

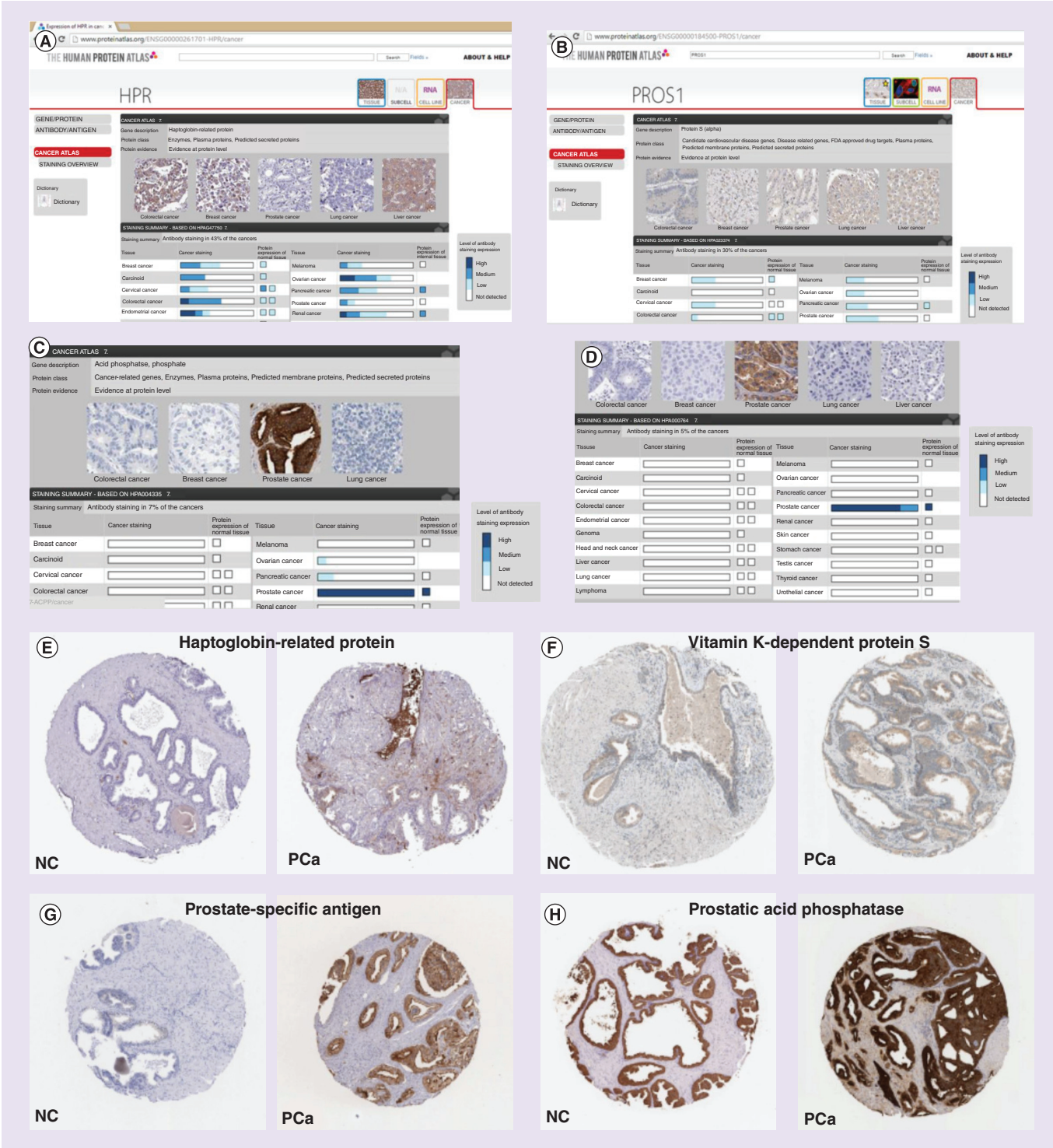


Figure 2. Immunohistochemical confirmation of potential prostate cancer biomarkers using the human protein atlas database. **(A & E)** Haptoglobin-related protein showed high to medium staining in comparison to normal tissue. **(B & F)** Vitamin K-dependent protein S demonstrated a low staining in comparison to normal prostate tissue which demonstrated no staining for this biomarker. **(C & H)** Prostatic acid phosphatase demonstrated heavy staining in both cancerous and normal prostate tissues, showing that it is prostate tissue specific but not PCa specific. **(D & G)** Differential expression between PCa and normal prostate tissue were observed for prostate-specific antigen and no other cancer was shown to express this biomarker.
NC: Normal healthy; PCa: Prostate cancer.

transitions between PCa and NC for most of the selected 32 peptides **(Supplementary Figure 2)** with a quantifiable distribution of area under curve (AUC) for identified transitions **(Figure 3A & B)**. There was differential expression of PGLYRP2 with a higher expression in PCa in comparison to NC. This difference in expression is more marked in the +2 charge state in comparison to the +3 charge state. There was more than twofold difference in peptide transitions between PCa and NC for PROS1. For both +2 and +3 charge state, b-type ions were higher expressed in PCa in comparison to NC. Similar trend was observed for HPR with more than threefold difference between PCa and NC irrespective of charge states. Differential expression for CPN1 was observed with +2 peptide transitions being higher in PCa and lower in NC. However, more transitions in were found for CPN1 in NC compared with PCa for the +3 state. PZP had more transition signals in the NC as compared with PCa for both charge states and was mostly y ions. ACTN1 was more expressed in NC compared with NC in the +3 charge state. Similarly, detection of transitions for MYOC was higher in NC than in PCa. ACPP had better signals in NC in comparison to PCa. KLK3 demonstrated higher transitions for PCa in comparison to NC. NID1 had a spectacular performance in PRM with over sixfold difference in transition signals for both charge states in NC which was higher than PCa. CTSZ was higher in transition signal for NC compared with PCa for both charge states. SLAIN1, a race-based PCa biomarker demonstrated marked differential expression between Pca and NC. We observed very low coefficient of variation in multiply injected samples in our previous experiments **(Figure 3C & D)**.

• Potential biomarker confirmation

The frequency of detection of these 12 potential biomarkers was confirmed in our shotgun discovery proteomics database. PROS1, HPR, PZP and SLAIN1 were found to be reliable biomarkers of PCa **(Figure 4)**. PROS1, HPR and PZP were only found in PCa and completely absent in BPH and NC, while there is greater than twofold difference in the expression of SLAIN1 in PCa which was higher than NC and BPH. CTSZ, NID1, ACPP and KLK3 were also reasonable biomarkers of PCa with higher expression in NC in comparison to PCA and BPH. NID1 was notable present in NC, minimally present in BPH but completely absent in PCa. ACPP and KLK3 were highly present in NC and BPH in comparison to PCa. CTSZ was highest in NC, present in BPH but absent in PCa. Other biomarkers were indeterminate with either similar expression across groups or similar expression in PC and NC. We further queried these biomarkers against 14 other urinary proteomics database/literature to see the degree of consonance with our findings. The databases are distributed as follows: six PCa, six healthy (NC), two BPH and one multiple condition (MC) urine **(Table 2)**. PGLYRP2 was found in 5 (83.3%) of the NC while absent in all PCa, BPH and MC databases. PROS1 was found in 4 (66.7%) of NC and 1 (16.7%) of PCa, but absent in BPH and MC databases. HPR was found in 3 (50%) of PCa and 1 (16.7%) of NC database; HPR was not found in BPH nor MC databases. CPN1 was found in 5 (83.3%) of NC and 2 (33.3%) of PCa, while absent in BPH and MC. PZP was absent in PCa, BPH and NC databases, albeit present in the MC database. ACTN1 was found in 3 (50%) of the NC and absent in the PCa, BPH and MC databases. MYOC was present in 2 (33.3%) of the NC and absent in PCa, BPH and MC databases. ACPP was found in 5 (83.3%) of the NC and 3 (50%) of the PCa databases; also present in the BPH and MC databases. KLK3 was seen in 5 (87.3%) of the NC and 1 (16.7%) of the PCa databases, while present in BPH but absent in MC databases. NID1 was present in 6 (100%) of the NC and absent (0%) in PCa, BPH and MC. CTSZ was present in 4 (66.7%) of the NC and 1 (16.7%) of the PCA databases; while absent for both MC and BPH. SLAIN1 was not found in any of these 14 searched databases.

Discussion

Performing targeted proteomics validation of biomarkers requires careful planning and rigorous assay optimization [6,29]. Unfortunately, the burgeoning number of potential biomarkers discovered for various diseases, do not correspond to clinically approved or applicable disease biomarker [29,30]. Due to high variations in results and methods, particularly for clinical urinary proteomics, some researchers have recommended standardization of biomarker discovery and validation pipelines [6,31–32]. Careful screening of potential biomarkers is required to validate the glut of candidate biomarkers and ameliorate the shortfall between discovery and validation phases. Considering that a high number

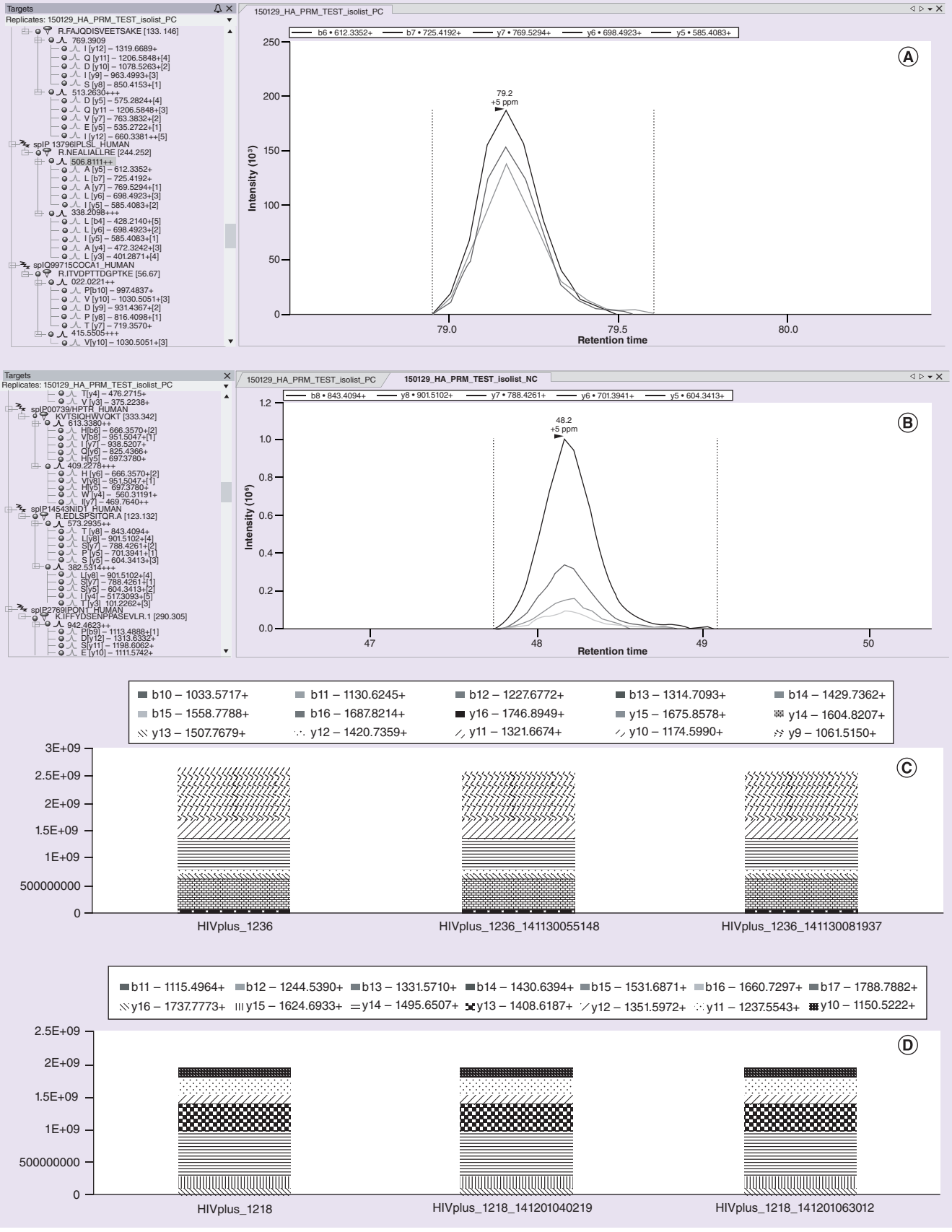

Targets
Replicates: 150129_HA_PRM_TEST_isolist_PC
150129_HA_PRM_TEST_isolist_PC
b6 • 612.3352+
b7 • 725.4192+
y7 • 769.5294+
y6 • 698.4923+
y5 • 585.4083+
A
79.2
+5 ppm
Intensity (10³)
Retention time
150129_HA_PRM_TEST_isolist_NC
b8 • 843.4094+
y8 • 901.5102+
y7 • 788.4261+
y6 • 701.3941+
y5 • 604.3413+
B
48.2
+5 ppm
Intensity (10⁶)
b10 – 1033.5717+
b11 – 1130.6245+
b12 – 1227.6772+
b13 – 1314.7093+
b14 – 1429.7362+
b15 – 1558.7788+
b16 – 1687.8214+
y16 – 1746.8949+
y15 – 1675.8578+
y14 – 1604.8207+
y13 – 1507.7679+
y12 – 1420.7359+
y11 – 1321.6674+
y10 – 1174.5990+
y9 – 1061.5150+
C
HIVplus_1236
HIVplus_1236_141130055148
HIVplus_1236_141130081937
b11 – 1115.4964+
b12 – 1244.5390+
b13 – 1331.5710+
b14 – 1430.6394+
b15 – 1531.6871+
b16 – 1660.7297+
b17 – 1788.7882+
y16 – 1737.7773+
y15 – 1624.6933+
y14 – 1495.6507+
y13 – 1408.6187+
y12 – 1351.5972+
y11 – 1237.5543+
y10 – 1150.5222+
D
HIVplus_1218
HIVplus_1218_141201040219
HIVplus_1218_141201063012

Figure 3. Parallel reaction monitoring for quantitation of potential prostate cancer biomarker target transitions and replicate comparison (see facing page). (A & B) Transition ion was evenly distribution and quantifiable across samples. **(C & D)** There was limited technical variability in peptide quantification reported between similar urinary proteomics experiments using urine from HIV-infected patients.

of randomized clinical samples are required for confidence in validation [29], experimental designs with fewer targeted analytes would more be realistic. Here we used HPA and SRMAtlas to tailor down 82 potential biomarkers we previously discovered for PCa in a South African cohort to 32 candidates. These 32 potential biomarkers were prevalidated with a preliminary PRM experiment to further streamline for large-scale targeted proteomics approaches.

In silico database searching is becoming an indispensable adjunct to high throughput systems biology, albeit there are limitations to its use [33]. Using the SRM database, we verified that over half (51.1%) of our selected PTPs have been assayed previously using targeted SRM. Another subset (47.7%) of selected PTPs from our shotgun proteomics evidence peptides have not been previously reported and can be added to the SRMAtlas. Up to 80% coverage has been reported when comparing *in silico* with experimental methods [34], showing the prospects of its application in the biomarker pipeline. Both prostate epithelial and stromal staining was put into consideration in the screening process because of the high predicted membrane/ECM localization of potential biomarkers. Due to wide coverage (11,200 unique proteins) and tissue-specific differential immunoproteomics expression profile of normal versus cancerous tissue; HPA has been

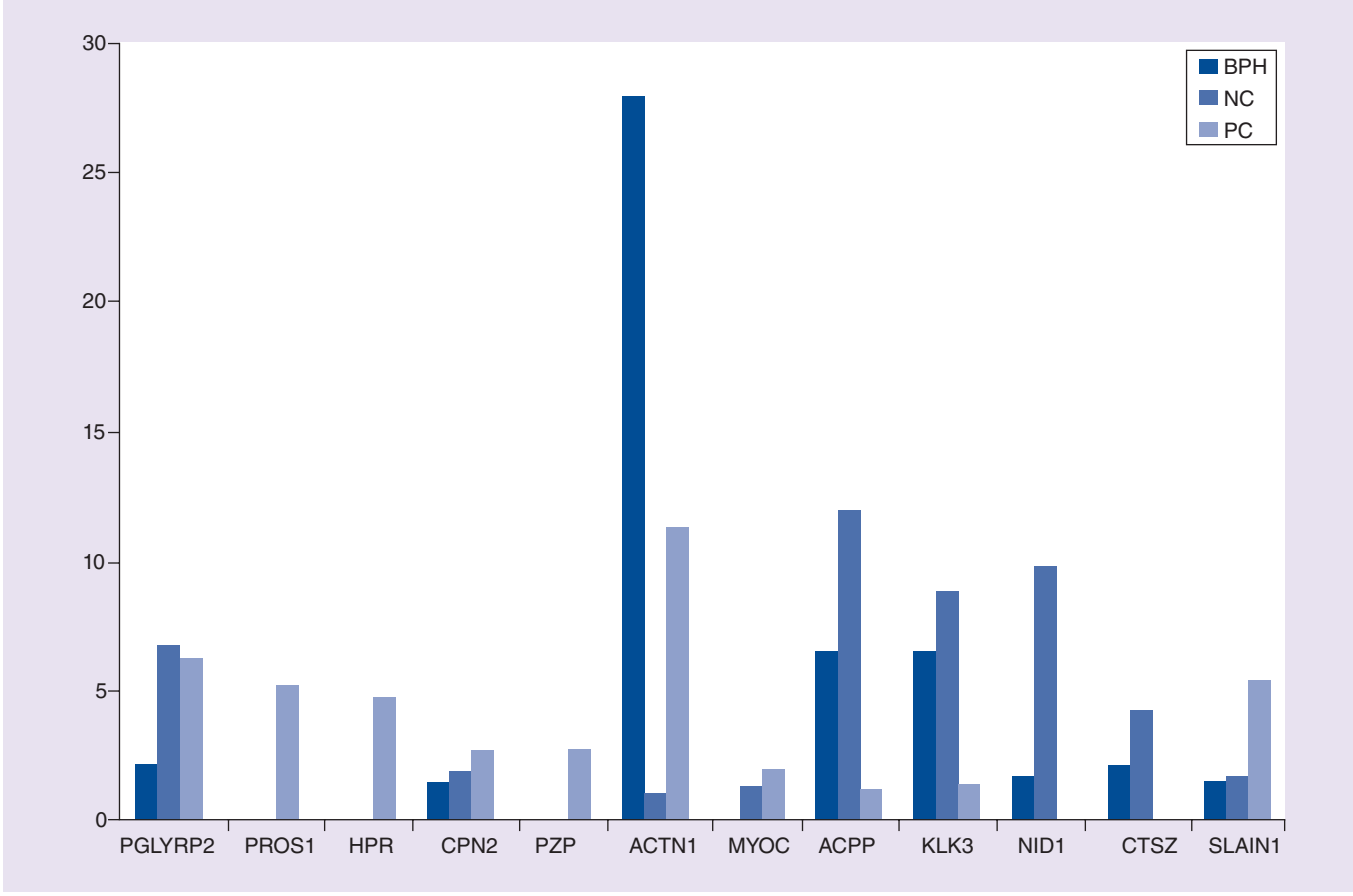


Figure 4. Discovery Shotgun database confirmation of potential prostate cancer biomarker. PROS1, HRP, PZP, NID1 and SLAIN1 were found to be good biomarkers of PCa by checking potential biomarker frequencies in our previous shotgun proteomics database. BPH: Benign prostatic hyperplasia; NC: Normal healthy; PCa: Prostate cancer.

Table 2. Confirmation of top ranking 12 biomarkers in 14 publicly available urinary proteomics databases.

Databases	Condition	Sample	PGLYRP2	PROS1	HRP	CPN1	PZP	ACTN1	MYOC	ACPP	KLK3	NID1	CTSZ	SLAIN1
Youhe Gao *et al.* (2011)	Multiple condition	Urine	(-)	(-)	(-)	(-)	(+)	(-)	(-)	(+)	(-)	(-)	(-)	(-)
Hong-Lin Cheng *et al.* (2012)	Benign prostatic hyperplasia	Urine	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(+)	(+)	(-)	(-)	(-)
Ihor Batruch *et al.* (2011)	Healthy	Urine	(+)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(+)	(-)	(-)
Xuejiao Liu *et al.* (2012)	Healthy	Urine	(+)	(+)	(-)	(+)	(-)	(-)	(-)	(+)	(+)	(+)	(+)	(-)
Qing-Run Li *et al.* (2010)	Healthy	Urine	(+)	(+)	(-)	(+)	(-)	(-)	(-)	(+)	(+)	(+)	(-)	(-)
HUPO Max Plank (2014)	Healthy	Urine	(-)	(+)	(-)	(+)	(-)	(+)	(+)	(+)	(+)	(+)	(+)	(-)
Jun Adachi *et al.* (2006)	Healthy	Urine	(+)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(+)	(+)	(+)	(-)
Akilesh Pandey *et al.* (2011)	Healthy	Urine	(+)	(-)	(+)	(+)	(-)	(+)	(+)	(+)	(+)	(+)	(+)	(-)
Dan Theodorescu *et al.* (2008)	PCa vs normal (PCa)	Urine	(-)	(-)	(+)	(+)	(-)	(-)	(-)	(+)	(+)	(-)	(-)	(-)
Katarina Davalieva *et al.* (2015)	PCa vs normal (PCa)	Urine	(-)	(-)	(+)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
Cordelia Geisler *et al.* (2015)	PCa vs normal (PCa)	Urine	(-)	(-)	(-)	(+)	(-)	(-)	(-)	(+)	(-)	(-)	(+)	(-)
Chunhui Li *et al.* (2015)	PCa vs normal (PCa)	Urine	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
Andrew Percy *et al.* (2015)	PCa vs normal (PCa)	Urine	(-)	(+)	(+)	(-)	(-)	(-)	(-)	(+)	(-)	(-)	(-)	(-)
Tujin Shi *et al.* (2014)	PCa vs normal (PCa)	Urine	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)

PCa: Prostate cancer.

branded as an important tool for histopathologic evaluation [35] and biomarkers discovery [36]. HPA has also contributed huge proteogenomic resources to the Human Proteome Project [37,38]. An important caveat to immunoproteomics verification of potential biomarker is sequence specificity and avidity of the antibody if a specified variant is to be targeted [39,40].

Due to concurrent availability of all product ion and minimal interference over the full scan, PRM has been said to possess highly dependable spectral identification. This is done even without prior knowledge or preselection of target transition. However, a duty cycle of close to 100% in the triple quadrupole instrument with electron multiplier-based detection offers a better sensitivity compared with the sensitivity of the Orbitrap's image current detection [41]. This calls to attention the tradeoff between the high mass accuracy and sensitivity between SRM and PRM experiments [24,33,41], PRMs are generally preferable in the screening mode of targeted proteomics experiments while SRM may be required for precise/absolute quantification of analytes across samples [41]. The stochastic nature of ion selection in a PRM setup may be responsible for occasional poor PTP results as well as the fact that we have used pooled samples.

Retrospective assessment of the frequency of identification of top ranking 12 potential biomarkers following PRM revealed differential expression of most of the biomarkers **(Figure 4)**. ACPP and KLK3 which are both putative biomarkers of prostate cancer were also detected in the normal urine, BPH as well as PCa. These biomarkers may be more appropriately markers of prostate disease rather than cancer-specific. Race-based SLAIN1 was found to be relatively more expressed in PCa compared with NC and BPH, albeit no data existed for this biomarker in the SRMAtlas. This may possibly be accounted for by the paucity of targeted urinary proteomics data from African cohorts. Notably HRP, PROS1, PZP, NID1 and SLAIN1 were good biomarkers of PCa. Further confirmation based on 14 other databases and literature showed variable reliability of these 12 top ranking biomarkers with many of these biomarkers found in healthy urinary proteomics databases as compared with PCa. Remarkably, NID1 was found only in the healthy urine samples. In order to evaluate the connection between these prevalidated biomarkers and the relevant putative existing PCa pathways such as *ETS* and *PTEN*. Therefore, we

performed functional network association analysis (data not shown) using GeneMANIA [42] to create a network for the top 12 biomarkers in addition to *ETS* and *PTEN* and found a network with predicted 79.19% co-expression, 17.26% physical interaction, 2.73% co-localization and 0.83% of shared protein domains between these biomarkers. Despite the relatively low level of predicted physical interaction between these biomarkers and *ETS*/*PTEN* pathways, the high level of predicted co-expression between them warrants further study.

Conclusion

It has become more evident that a combination of *in silico* and experimental approaches for high throughput systems biology approaches like proteomics, lipidomics and metabolomics is essential. This study was carried out on pooled samples drawn from a small population of PCa patients and hence the 12 prevalidated biomarkers would require further validation in a larger cohort of individual PCa patient samples using SRM or PRM. Despite its immense benefits, an important limitation of database-dependent analysis is the confidence measure of the data quality and the occasional restricted access to independently filter or screen these data [43]. We have demonstrated that an intermediate prevalidation step such as a combination of *in silico* HPA, SRMAtlas and preliminary targeted proteomics using parallel reaction monitoring was highly beneficial in potential PCa biomarker screening process. We have taken advantage of newly emerging high throughput cutting edge technology and existing databases to

EXECUTIVE SUMMARY

- Prostate cancer is a leading cause of death in elderly males globally and particularly in Africa.
- Men of African descent have been known to suffer aggressive phenotypes of the disease.
- Prostate-specific antigen (PSA), an important biomarker of prostate cancer currently available is known to be fraught with false positives an overtreatment, particularly in the lower reference ranges.
- There is an increase in the number of potential biomarkers generated from discovery experiment, albeit only a few of these biomarkers make it to clinical utility.
- Despite the emergence of novel high-throughput cutting edge technologies for large-scale discovery of candidate disease biomarkers, there remains a bioinformatics gap for clinical applicability of these biomarkers.

Materials & methods

- Using freely available online data repositories, we have employed an intermediate *in silico* verification to streamline 82 previously discovered potential biomarkers of prostate cancer from our urinary shotgun proteomics experiments.
- We further prevalidated 32 verified biomarkers in prostate cancer and normal sample using parallel reaction monitoring (PRM) targeted proteomics in a QExactive™ Hybrid Quadrupole-Orbitrap mass spectrometer in line with Dionex UltiMate® 3500 RSnano UPLC system.

Results

- Based on stringency criteria, 32, 25, 13 and 7 top ranking biomarkers were discovered.
- We observed good transition signals in 12 biomarkers, *viz* the top ranking seven biomarkers in addition to five other verified biomarkers (including two putative prostate cancer biomarkers).
- Frequency of occurrence of these 12 prevalidated potential biomarkers was reconfirmed in our previous discovery urinary shotgun proteomics data and observed variations in expression of these biomarkers between prostate cancer, benign prostatic hyperplasia and normal healthy samples.
- We also observed variation in top 12 biomarker abundance between prostate cancer and controls in another 14 publicly available urinary proteomics databases.

Conclusion

- This study was carried out on a small study population and hence needs further multiplatform validation in a larger cohort of prostate cancer patients.
- The penultimate prevalidation workflow presented herein could help improve clinical translation of potential biomarkers from hypotheses generation discovery proteomics experiments.

potentially address the putative clinical implementation bottlenecks of targeted potential biomarker validation. This approach can provide reliable biomarkers for large scale, multiplatform, hypothesis testing targeted proteomics experiment.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: http://www.futuremedicine.com/doi/suppl/10.2217/fon.15.296

Financial & competing interests disclosure

This work was funded by the International Centre for Genetic Engineering and Biotechnology (ICGEB). HAA was recipient of the ICGEB predoctoral fellowship. JMB is grateful to the National Research Foundation for a South African Research Chair Initiative grant. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

References

Papers of special note have been highlighted as:
• of interest; •• of considerable interest

1 Parkin DM, Bray F, Ferlay J, Jemal A. Cancer in Africa 2012. *Cancer Epidemiol. Biomarkers Prev.* 23(6), 953–966 (2014).

2 Filella X, Foj L. Emerging biomarkers in the detection and prognosis of prostate cancer. *Clin. Chem. Lab. Med.* 53(7), 963–973 (2015).

3 Schalken J, Dijkstra S, Baskin-Bey E, Van Oort I. Potential utility of cancer-specific biomarkers for assessing response to hormonal treatments in metastatic prostate cancer. *Ther. Adv. Urol.* 6(6), 245–252 (2014).

4 Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* 1(11), 845–867 (2002).

5 Gutman S, Kessler LG. The US Food and Drug Administration perspective on cancer biomarker development. *Nat. Rev. Cancer* 6(7), 565–571 (2006).

• **This paper reviews the bottlenecks to cancer biomarker development observed by the US FDA and suggested possible ways to improve the biomarker pipeline.**

6 Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* 24(8), 971–983 (2006).

•• **A very good paper that underscores one of the major research problems which is the poor targeted clinical proteomics validation of discovered potential disease biomarkers.**

7 Taube SE, Clark GM, Dancey JE, Mcshane LM, Sigman CC, Gutman SI. A perspective on challenges and issues in biomarker development and drug and biomarker codevelopment. *J. Natl Cancer Inst.* 101(21), 1453–1463 (2009).

8 Chen C, Mcgarvey PB, Huang H, Wu CH. Protein bioinformatics infrastructure for the integration and analysis of multiple high-throughput 'omics' data. *Adv. Bioinformatics* 2010, 423589 (2010).

• **This paper emphasizes the importance of computational biology and bioinformatics in integrating data generated from multiple high-throughput omics-based approaches and drawing translational/clinical inferences from them.**

9 Zhang GL, Deluca DS, Brusic V. Database resources for proteomics-based analysis of cancer. *Methods Mol. Biol.* 723, 349–364 (2011).

10 Ahmed J, Meinel T, Dunkel M *et al.* Cancer resource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge. *Nucleic Acids Res.* 39, D960–D967 (2011).

11 Vizcaino JA, Foster JM, Martens L. Proteomics data repositories: providing a safe haven for your data and acting as a springboard for further research. *J. Proteomics* 73(11), 2136–2146 (2010).

12 Colangelo CM, Shifman M, Cheung KH *et al.* YPED: an integrated bioinformatics suite and database for mass spectrometry-based proteomics research. *Genomics Proteomics Bioinformatics* 13(1), 25–35 (2015).

13 Gaudet P, Michel PA, Zahn-Zabal M *et al.* The neXtProt knowledgebase on human proteins: current status. *Nucleic Acids Res.* 43, D764–D770 (2015).

14 Uhlen M, Bjorling E, Agaton C *et al.* A Human Protein Atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell Proteomics* 4(12), 1920–1932 (2005).

• **Describes in great details the potential biomarker applications of information generated in the Human Protein Atlas database which we used in this study.**

15 Nilsson P, Paavilainen L, Larsson K *et al.* Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics* 5(17), 4327–4337 (2005).

16 Desiere F, Deutsch EW, King NL *et al.* The PeptideAtlas project. *Nucleic Acids Res.* 34, D655–D658 (2006).

17 Deutsch EW, Lam H, Aebersold R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* 9(5), 429–434 (2008).

18 Deutsch EW. The PeptideAtlas project. *Methods Mol. Biol.* 604, 285–296 (2010).

• **Describes the development of the PeptideAtlas database upon which the SRMAtlas repository we used in this paper was developed**

19 Farrah T, Deutsch EW, Kreisberg R *et al.* PASSEL: the PeptideAtlas SRMexperiment library. *Proteomics* 12(8), 1170–1175 (2012).

20 Rost H, Malmstrom L, Aebersold R. A computational tool to detect and avoid redundancy in selected reaction monitoring. *Mol. Cell. Proteomics* 11(8), 540–549 (2012).

21 Huttenhain R, Surinova S, Ossola R *et al.* 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(4), 1005–1016 (2013).

22 Kusebauch U, Deutsch EW, Campbell DS, Sun Z, Farrah T, Moritz RL. Using PeptideAtlas, SRMAtlas, and PASSEL: comprehensive resources for discovery and targeted proteomics. *Curr. Protoc. Bioinformatics* 46, 13.25.11–13.25.28 (2014).

23 Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molec. Syst. Biol.* 4, 222 (2008).

24 Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics* 11(11), 1475–1488 (2012).

25 Doerr A. Targeting with PRM. *Nat. Methods* 9(10), 950 (2012).

26 Ronsein GE, Pamir N, Von Haller PD *et al.* Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *J. Proteomics* 113, 388–399 (2015).

27 Adeola HA, Soares NC, Paccez JD, Kaestner L, Blackburn JM, Zerbini LF. Discovery of novel candidate urinary protein biomarkers for prostate cancer in a multiethnic cohort of South African patients via label-free mass spectrometry. *Proteomics Clin. Appl.* 9(5–6), 597–609 (2015).

•• **The 82 potential biomarkers which we verified and prevalidated in the current study were discovered in this previous prostate cancer urinary biomarker discovery shotgun proteomics experiment.**

28 Maclean B, Tomazela DM, Shulman N *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26(7), 966–968 (2010).

29 Ye X, Blonder J, Veenstra TD. Targeted proteomics for validation of biomarkers in clinical samples. *Brief Funct. Genomic Proteomics* 8(2), 126–135 (2009).

30 Solier C, Langen H. Antibody-based proteomics and biomarker research – current status and limitations. *Proteomics* 14(6), 774–783 (2014).

31 Mischak H, Kolch W, Aivaliotis M *et al.* Comprehensive human urine standards for comparability and standardization in clinical proteome analysis. *Proteomics Clin. Appl.* 4(4), 464–478 (2010).

• **This paper discussed the challenges of clinical proteomics and the need to set up a standard protocol for conducting targeted clinical proteomics for validating potential biomarkers.**

32 Decramer S, Gonzalez De Peredo A, Breuil B *et al.* Urine in clinical proteomics. *Mol. Cell. Proteomics* 7(10), 1850–1862 (2008).

33 Sacan A, Ekins S, Kortagere S. Applications and limitations of *in silico* models in drug discovery. *Methods Mol. Biol.* 910, 87–124 (2012).

34 Nahnsen S, Kohlbacher O. *In silico* design of targeted SRM-based experiments. *BMC Bioinformatics* 13(Suppl. 16), S8 (2012).

35 Ponten F, Jirstrom K, Uhlen M. The Human Protein Atlas – a tool for pathology. *J. Pathol.* 216(4), 387–393 (2008).

36 Ponten F, Schwenk JM, Asplund A, Edqvist PH. The Human Protein Atlas as a proteomic resource for biomarker discovery. *J. Intern. Med.* 270(5), 428–446 (2011).

37 Berglund L, Bjorling E, Oksvold P *et al.* A genecentric Human Protein Atlas for expression profiles based on antibodies. *Mol. Cell. Proteomics* 7(10), 2019–2027 (2008).

38 Fagerberg L, Hallstrom BM, Oksvold P *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* 13(2), 397–406 (2014).

39 Rhodes KJ, Trimmer JS. Antibodies as valuable neuroscience research tools versus reagents of mass distraction. *J. Neurosci.* 26(31), 8017–8020 (2006).

40 Praekelt U, Kopp PM, Rehm K *et al.* New isoform-specific monoclonal antibodies reveal different sub-cellular localisations for TALIN1 and TALIN2. *Eur. J. Cell Biol.* 91(3), 180–191 (2012).

41 Gallien S, Bourmaud A, Kim SY, Domon B. Technical considerations for large-scale parallel reaction monitoring analysis. *J. Proteomics* 100, 147–159 (2014).

• **An important paper that describes intricately the technical quality control aspects of using parallel reaction monitoring for targeted hypothesis testing proteomics experiment on a large scale.**

42 Warde-Farley D, Donaldson SL, Comes O *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38, W214–W220 (2010).

43 Schramm SJ, Jayaswal V, Goel A *et al.* Molecular interaction networks for the analysis of human disease: utility, limitations, and considerations. *Proteomics* 13(23–24), 3393–3405 (2013).