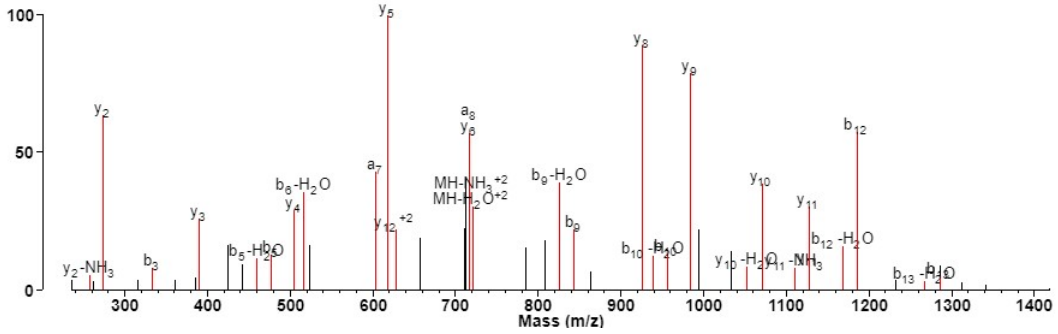
SUPPLEMENTAL FIGURES FOR

Multi-omics biomarker pipeline reveals elevated levels of protein-glutamine gamma-glutamyltransferase 4 in seminal plasma of prostate cancer patients

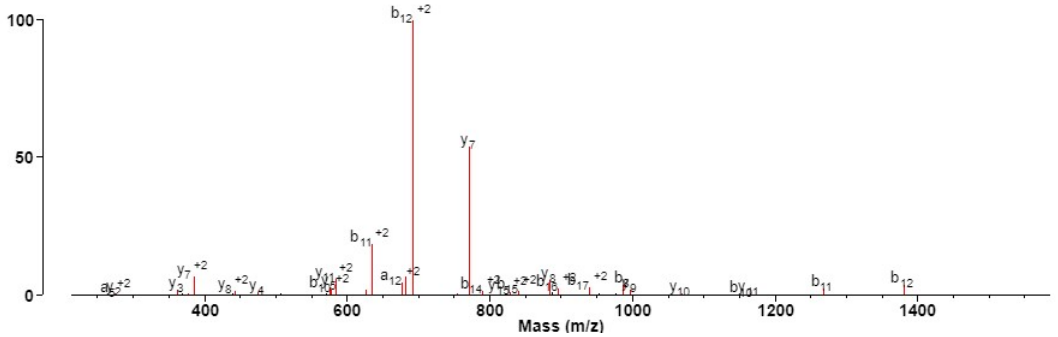
Andrei P. Drabovich*, Punit Saraon, Mikalai Drabovich, Theano D. Karakosta, Apostolos Dimitromanolakis, M. Eric Hyndman, Keith Jarvi and Eleftherios P. Diamandis

*Correspondence to: andrei.drabovich@ualberta.ca

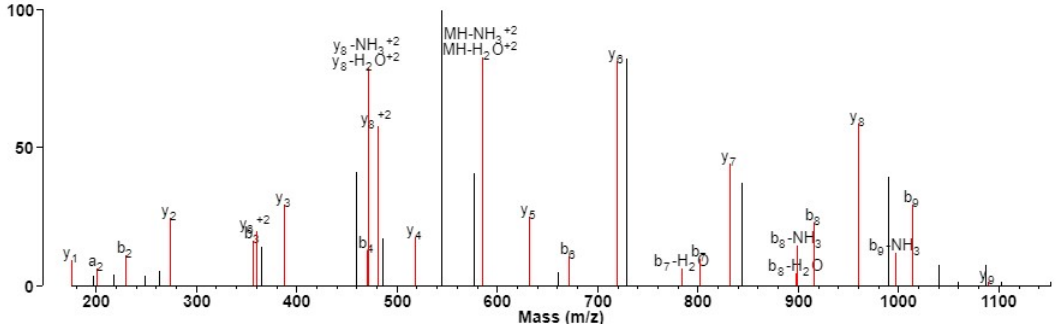
DMBT1_HUMAN (peptide FGQGSGPIVIDDVR^{+2})



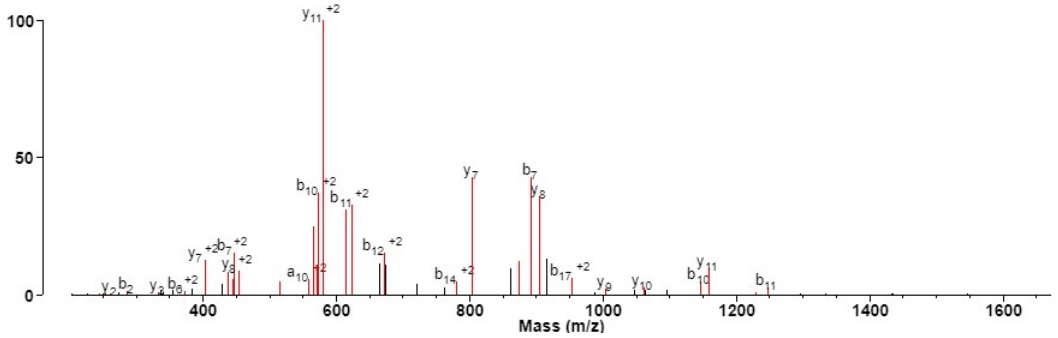
NECT2_HUMAN (peptide VEHESFEEPAIIPVTISVR^{+3})



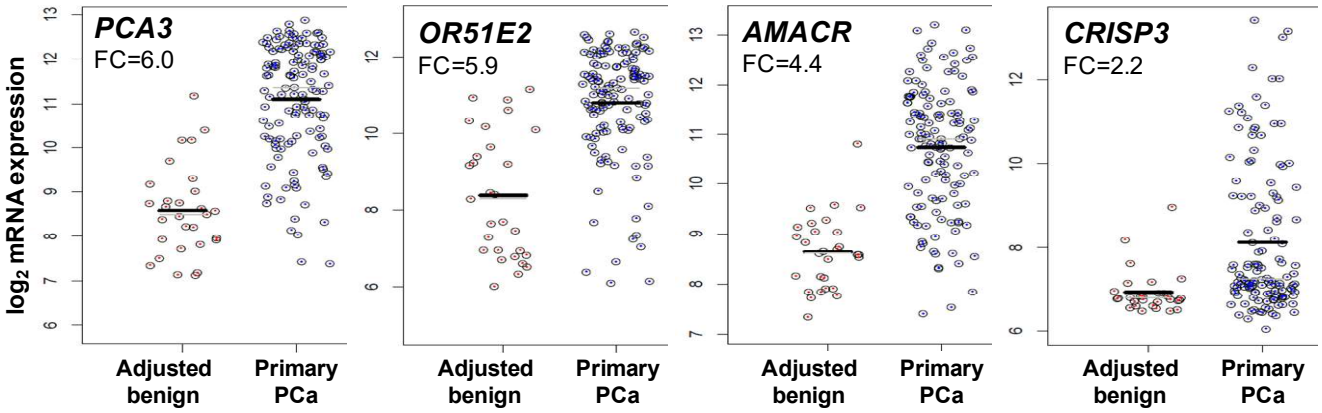
TGFB1_HUMAN (peptide VEQISNMIVR^{+2})



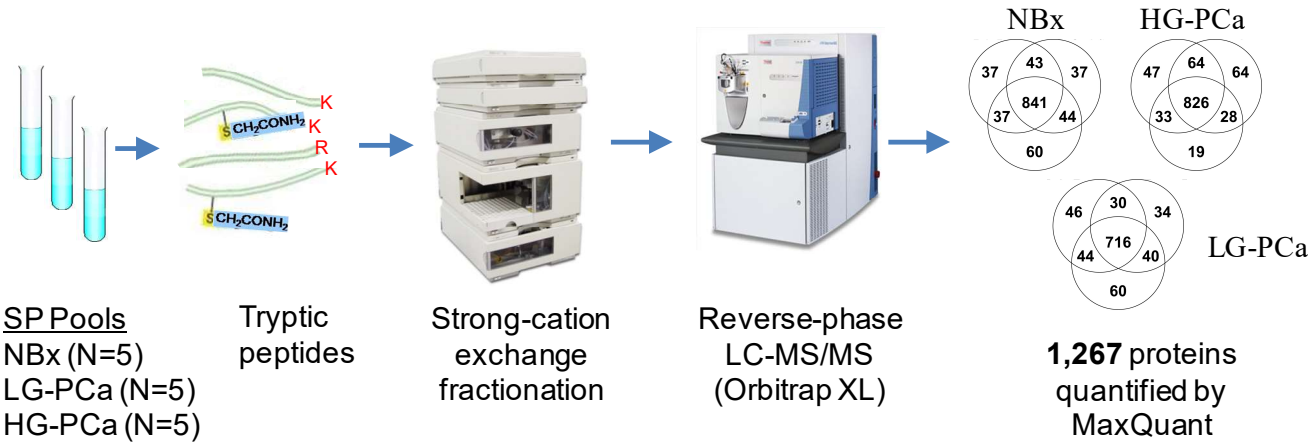
TFF1_HUMAN (peptide ERQNC(Carbamidomethyl)GFPGVTPSQC(Carbamidomethyl)ANK^{+3})



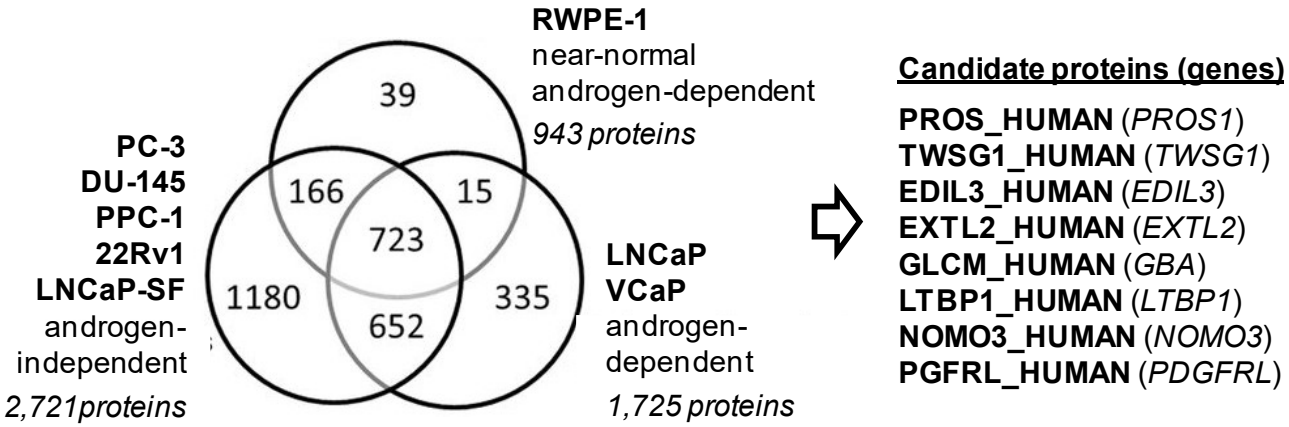
Supplemental Figure S1. Annotated spectra for the single-peptide identification candidate proteins highlighted in Supplemental Tables S3, S5. Annotated spectra for other single-peptide identified proteins can be viewed with MS-Viewer at the following permanent link http://msviewer.ucsf.edu/prospector/cgi-bin/mssearch.cgi?report_title=MS-Viewer&search_key=enahbvtldi&search_name=msviewer



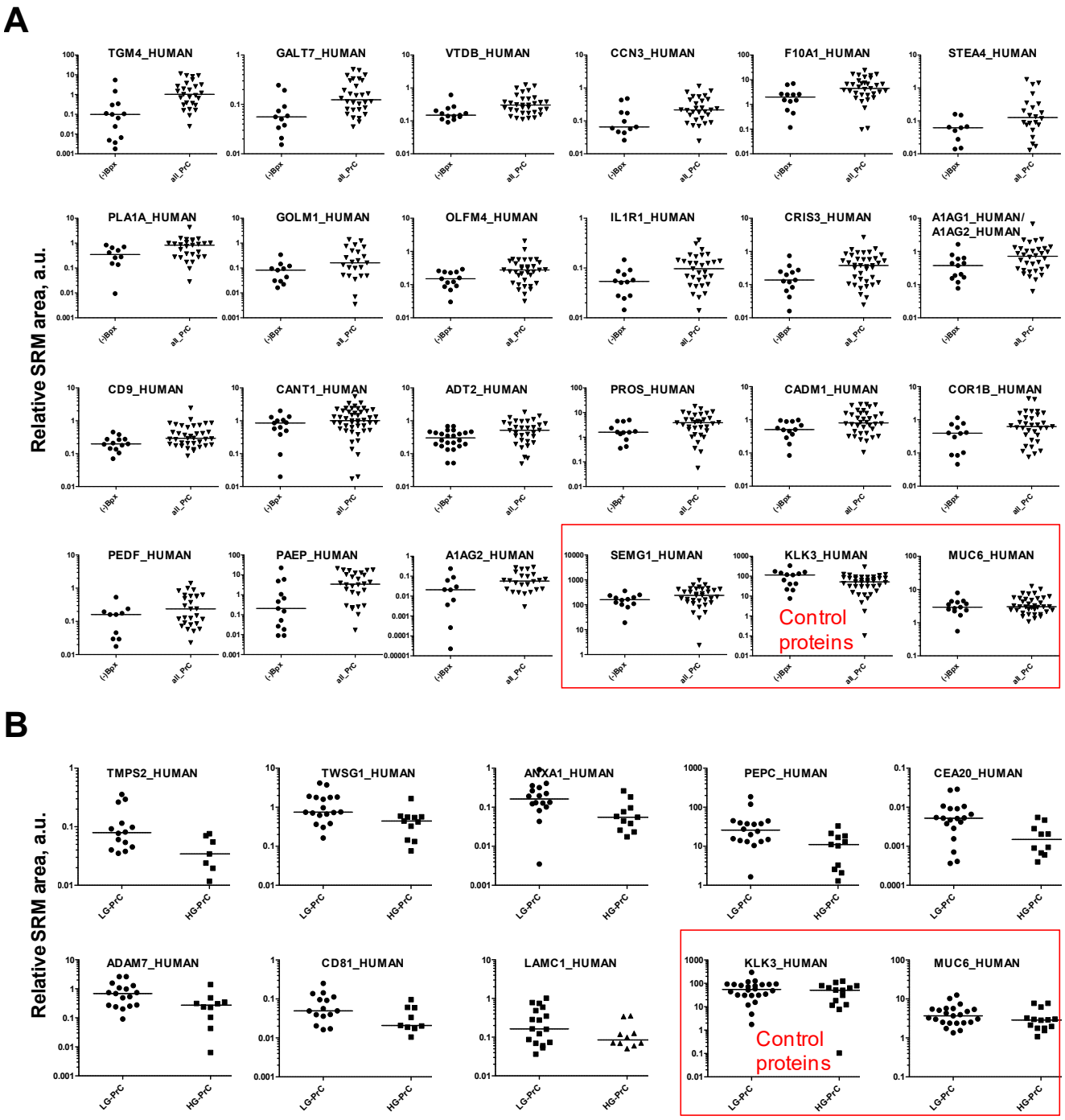
Supplemental Fig. S2. Differential mRNA gene expression of selected candidates, as derived from data available at the Cancer Genomics database. The Memorial Sloan-Kettering Cancer Center Gene expression microarray data (Taylor BS et al. *Cancer Cell* 2010;18:11-22) included 131 primary PCa tissues and 29 adjacent benign prostate tissues, as well clinical data (serum PSA, age and GS).



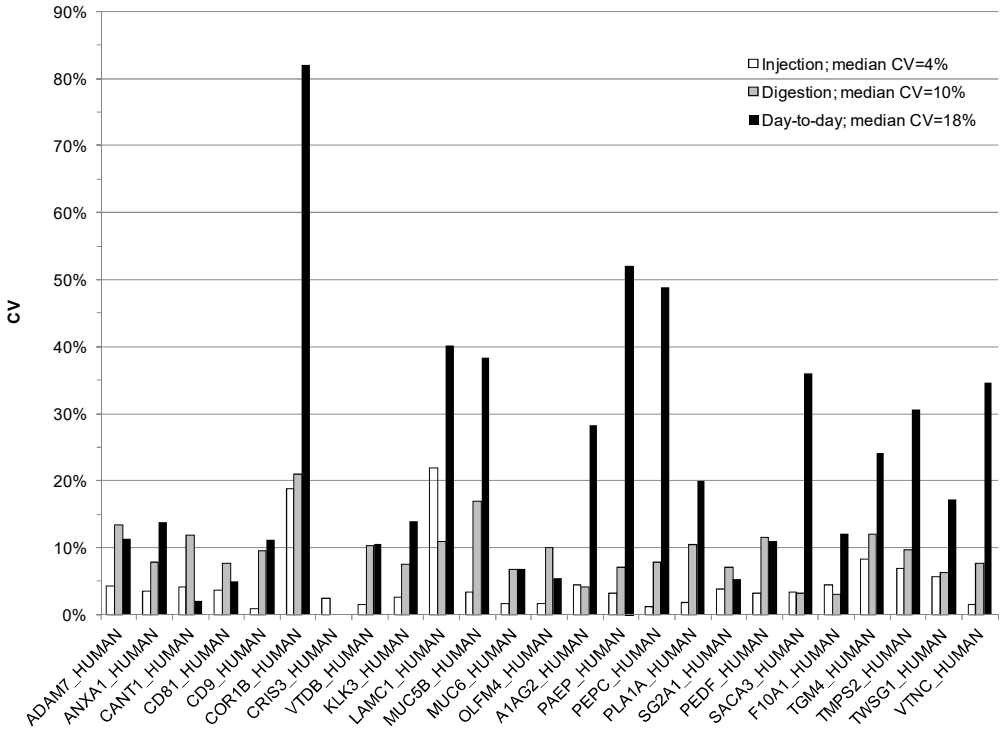
Supplemental Fig. S3. Differential proteomics approach. NBx, negative biopsy; LG, low-grade; HG, high-grade.



Supplemental Fig. S4. Differential secretomics approach based on our previous data (Saraon P *et al. J Biol Chem* 2012, 287, 34019-31). Eight protein found in the secretome of androgen-independent versus androgen-dependent/near normal cell lines were selected.



Supplemental Fig. S5. Quantification of candidate biomarkers in the qualification phase by SRM. (**A**) All PCa (N=38) versus negative biopsy ((-)Bpx, N=13) SP samples. For this comparison, 21 of 76 proteins were found differentially expressed (P<0.05). (B) Low-grade (LG-PCa, N=24) versus high-grade (HG-PCa, n=14) PCa SP samples. For this comparison, 8 of 76 proteins were found differentially expressed (P<0.05). SEMG1, KLK3 and MUC6 were used as control proteins.



Supplemental Fig. S6. Reproducibility of protein analysis by SRM in the verification phase. Three SP samples with different total protein concentration (35, 67 and 102 mg/mL) were digested in triplicates (variability of digestion) and measured by SRM in technical triplicates (variability of injection). The full protocol was repeated on a different day (day-to-day variability, analytical duplicates). Medians for all technical, analytical and day-to-day replicates across three samples were calculated for each protein.

A

QUALIFICATION	1	2	3	4	5	6	7	8	9	10	11	12
A	NBx	HG-PCa	NBx	HG-PCa	LG-PCa	LG-PCa	LG-PCa	NBx	HG-PCa	LG-PCa	NBx	HG-PCa
B	HG-PCa	NBx	HG-PCa	HG-PCa	LG-PCa	NBx	HG-PCa	NBx	LG-PCa	HG-PCa	LG-PCa	HG-PCa
C	HG-PCa	NBx	LG-PCa	LG-PCa	LG-PCa	NBx	LG-PCa	LG-PCa	HG-PCa	LG-PCa	NBx	LG-PCa
D	NBx	NBx	LG-PCa	NBx	LG-PCa	NBx	LG-PCa	LG-PCa	LG-PCa	LG-PCa	LG-PCa	LG-PCa
E	LG-PCa	NBx	NBx	NBx	LG-PCa	NBx	NBx	LG-PCa	LG-PCa	HG-PCa	LG-PCa	LG-PCa
F	HG-PCa	HG-PCa										
G												
H										Blank	Blank	Blank

Negative biopsy (N=13)
Low-grade PCa (N=24)
High-grade PCa (N=14)

Samples not used in data analysis
Negative biopsy: SRM acquisition failed (N=6)
LG-PCa: SRM acquisition failed (N=4)
HG-PCa: SRM acquisition failed (N=1)

B

VERIFICATION: Plate 1	1	2	3	4	5	6	7	8	9	10	11	12
A	HG-PCa	NBx	HG-PCa	LG-PCa		IG-PCa	LG-PCa	NBx	NBx	LG-PCa		LG-PCa
B	NBx	IG-PCa		NBx	LG-PCa	IG-PCa	IG-PCa	LG-PCa	IG-PCa	NBx	LG-PCa	LG-PCa
C	LG-PCa		IG-PCa	HG-PCa		LG-PCa	IG-PCa	NBx	LG-PCa	LG-PCa	LG-PCa	LG-PCa
D			LG-PCa	NBx	IG-PCa	IG-PCa	HG-PCa	HG-PCa				NBx
E	0.04 pmol/mL	0.4	4	40	400	4000	6000	SP pool				
F												
G												
H										Blank	Blank	Blank

VERIFICATION: Plate 2	1	2	3	4	5	6	7	8	9	10	11	12
A		LG-PCa	IG-PCa	IG-PCa			IG-PCa	LG-PCa	NBx	IG-PCa		IG-PCa
B	LG-PCa	HG-PCa	LG-PCa	LG-PCa	NBx	LG-PCa	NBx	LG-PCa		LG-PCa		NBx
C		HG-PCa	NBx	IG-PCa	NBx	NBx	LG-PCa	LG-PCa		LG-PCa	HG-PCa	NBx
D	LG-PCa	HG-PCa	NBx	LG-PCa	NBx	LG-PCa	LG-PCa	LG-PCa	NBx	HG-PCa	NBx	
E	0.04 pmol/mL	0.4	4	40	400	4000	6000	SP pool				
F												
G												
H										Blank	Blank	Blank

VERIFICATION: Plate 3	1	2	3	4	5	6	7	8	9	10	11	12
A	NBx	LG-PCa	NBx	LG-PCa	NBx		IG-PCa	LG-PCa	NBx	IG-PCa	LG-PCa	IG-PCa
B	LG-PCa	LG-PCa		LG-PCa	LG-PCa	NBx			IG-PCa	LG-PCa	IG-PCa	NBx
C	NBx	IG-PCa	NBx	IG-PCa	IG-PCa	LG-PCa	IG-PCa	LG-PCa	NBx	NBx	LG-PCa	NBx
D	NBx	LG-PCa		IG-PCa	LG-PCa	LG-PCa	LG-PCa	NBx	LG-PCa	NBx	LG-PCa	LG-PCa
E	0.04 pmol/mL	0.4	4	40	400	4000	6000	SP pool				
F												
G												
H										Blank	Blank	Blank

VERIFICATION: Plate 4	1	2	3	4	5	6	7	8	9	10	11	12
A	NBx	IG-PCa	IG-PCa	LG-PCa		NBx	IG-PCa	NBx	LG-PCa	LG-PCa	NBx	
B	NBx	NBx	IG-PCa	HG-PCa	NBx		NBx	LG-PCa	LG-PCa	NBx	LG-PCa	LG-PCa
C	LG-PCa	HG-PCa	NBx	LG-PCa	LG-PCa	LG-PCa	LG-PCa	LG-PCa	LG-PCa		LG-PCa	NBx
D	NBx	NBx		LG-PCa	LG-PCa	LG-PCa	NBx	IG-PCa	IG-PCa	NBx	LG-PCa	LG-PCa
E	0.04 pmol/mL	0.4	4	40	400	4000	6000	SP pool				
F												
G												
H										Blank	Blank	Blank

VERIFICATION: Plate 5	1	2	3	4	5	6	7	8	9	10	11	12
A	LG-PCa	LG-PCa	LG-PCa	NBx	NBx	HG-PCa	HG-PCa		NBx	NBx	NBx	LG-PCa
B		LG-PCa		IG-PCa	HG-PCa	NBx		LG-PCa	LG-PCa		IG-PCa	LG-PCa
C	NBx			HG-PCa	NBx	HG-PCa	LG-PCa	LG-PCa		LG-PCa	LG-PCa	LG-PCa
D	LG-PCa	NBx	HG-PCa	LG-PCa	NBx		LG-PCa	LG-PCa	LG-PCa	NBx		
E	0.04 pmol/mL	0.4	4	40	400	4000	6000	SP pool				
F												
G												
H										Blank	Blank	Blank

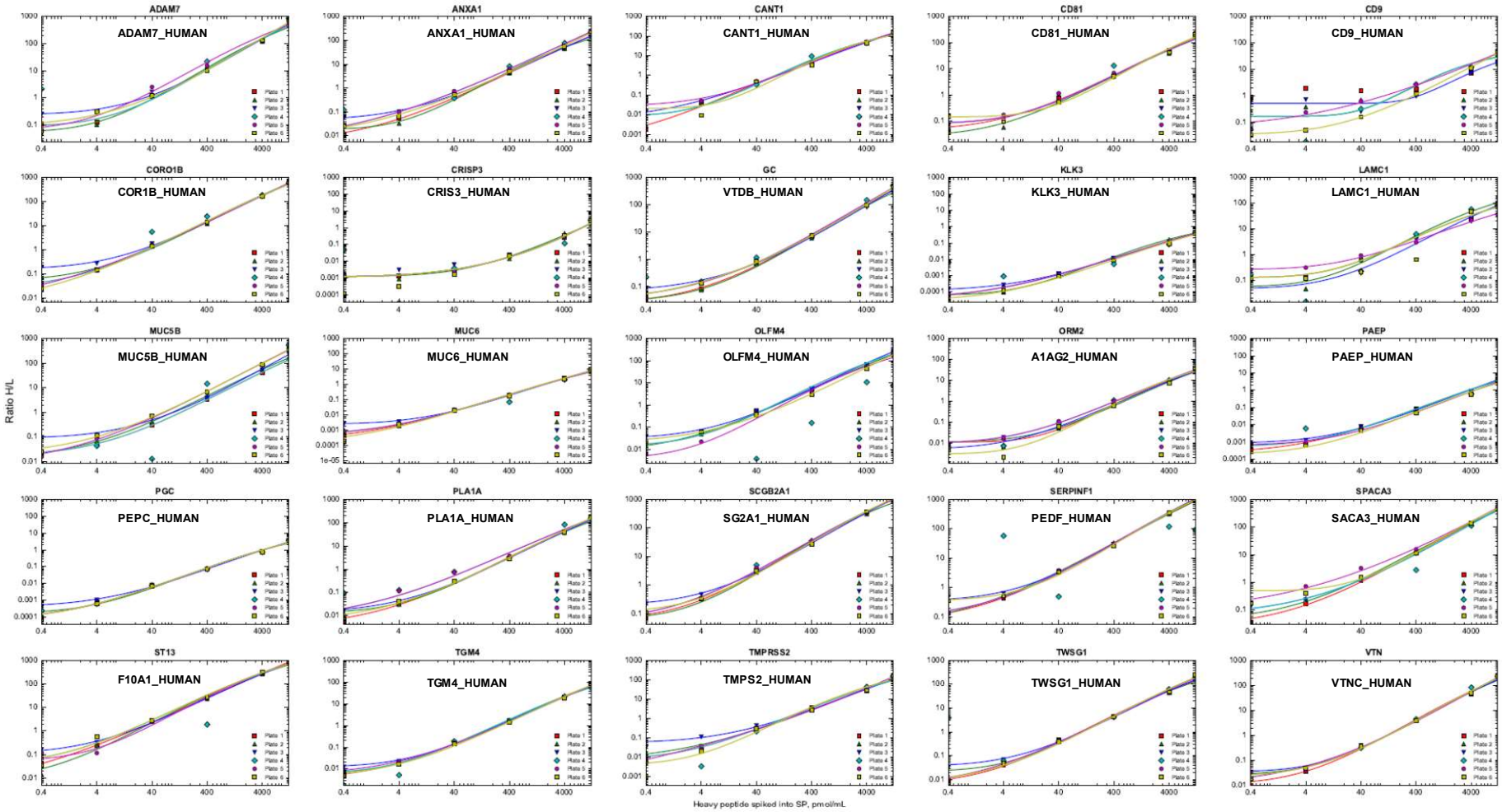
VERIFICATION: Plate 6	1	2	3	4	5	6	7	8	9	10	11	12
A	NBx	NBx		NBx	IG-PCa	LG-PCa	IG-PCa		HG-PCa	NBx	HG-PCa	LG-PCa
B												
C	NBx	NBx	LG-PCa	LG-PCa	LG-PCa	IG-PCa	LG-PCa		IG-PCa	IG-PCa		
D												
E	0.04 pmol/mL	0.4	4	40	400	4000	6000	SP pool				
F												
G												
H										Blank	Blank	Blank

Negative biopsy (N=67)
Low-grade PCa (N=94)
Intermediate-grade PCa (N=38)
High-grade PCa (N=20)

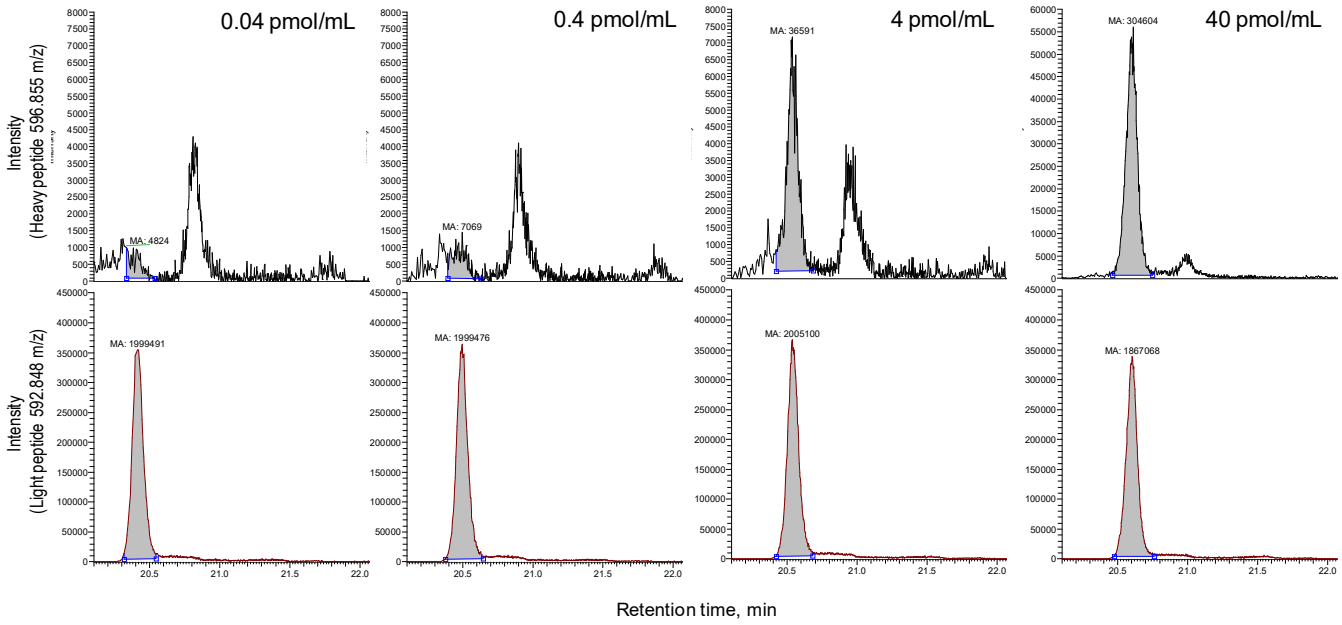
Samples not used in data analysis
Atypical small acinar proliferation (N=2)
High-grade prostatic intraepithelial neoplasia (N=3)
Prostate Inflamation (N=13)
Healthy Fertile men (N=3)
Healthy Fertile men Pre-Vasectomy (N=5)
Healthy Fertile men Post-Vasectomy (N=5)
Diagnosis not confirmed (N=6)
Negative biopsy: SRM acquisition failed (N=7)
PCa: SRM acquisition failed (N=17)

Calibration curve SP samples, pmol/mL

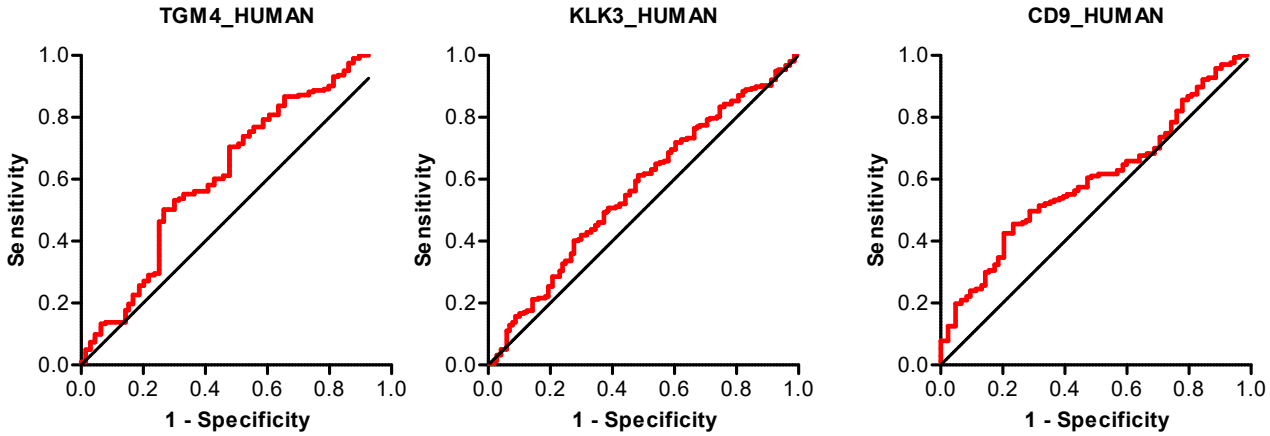
Supplemental Figure S7. Plate setup in the qualification phase by SRM (**A**) and verification phase by SRM (**B**). Color scheme defines groups of samples. To randomize SP samples, the Excel function RAND() was applied, and the list was sorted "low-to-high".



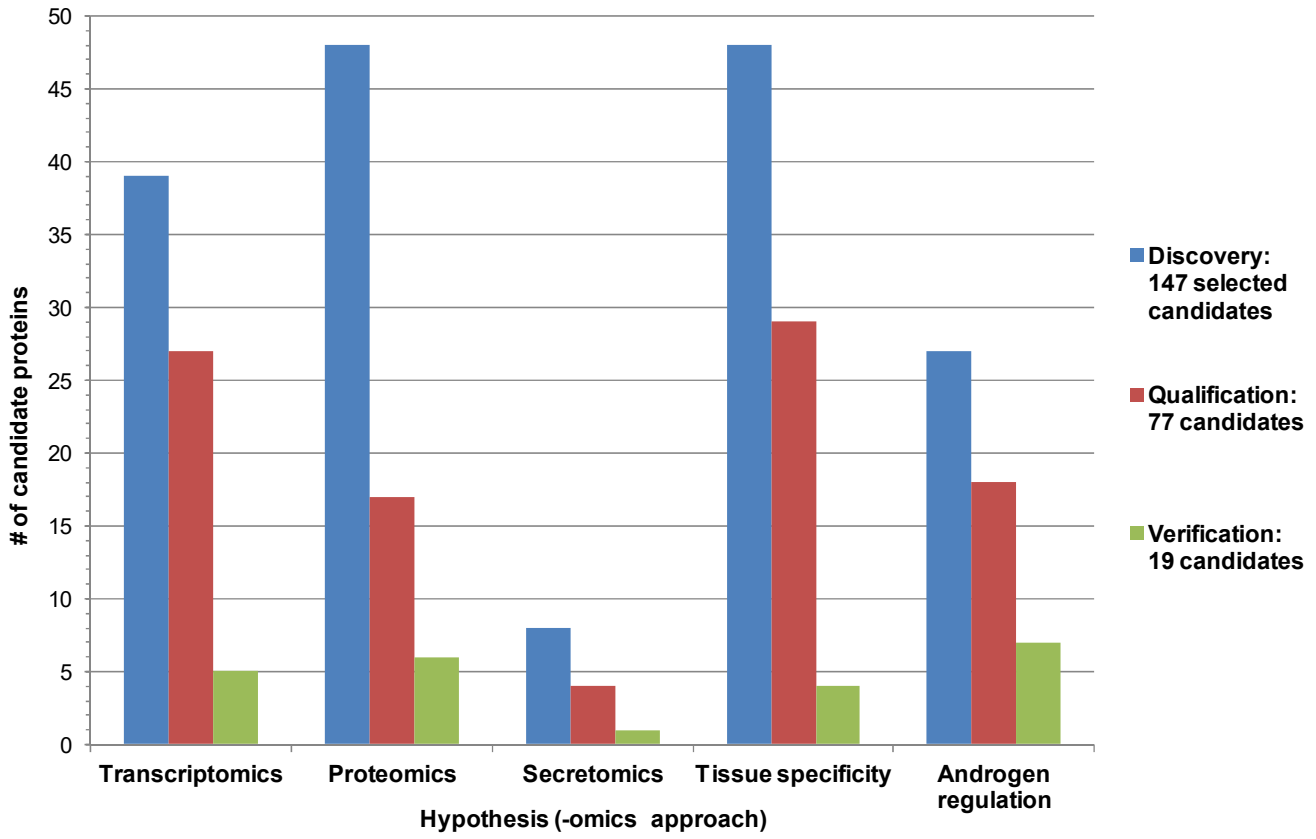
Supplemental Fig. S8. Calibration curves used to quantify 25 proteins (19 candidates and 9 controls) in 219 SP samples randomized and distributed between six 96-well plates.



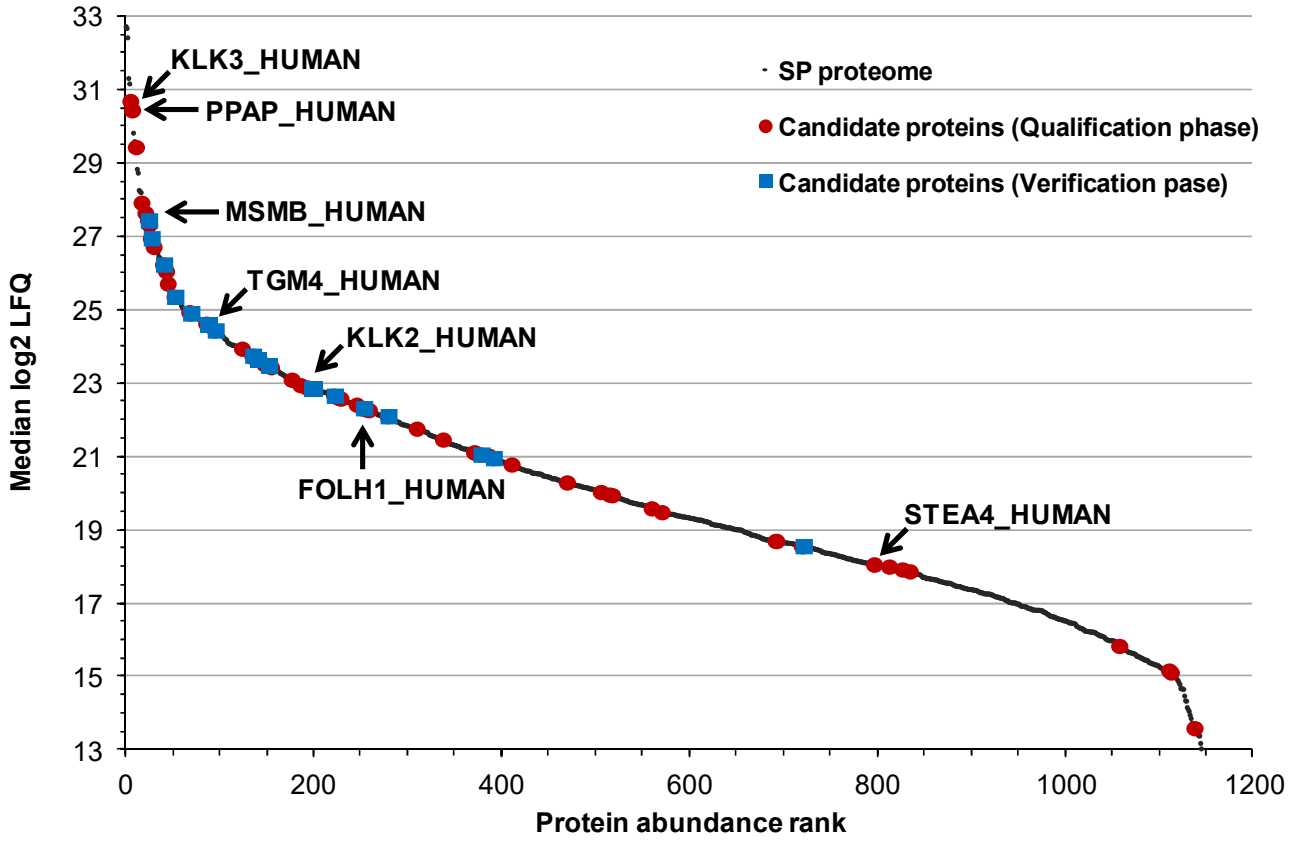
Supplemental Figure S9. Extracted ion chromatograms for a representative calibration curve for TGM4 peptide NTLAIPLTDVK. Dilution series of quantified heavy peptide standards were spiked into aliquots of an SP pool containing endogenous light proteins, and digested. Heavy-to-light peptide ratios were used to plot calibration curves (see **Supplemental Figure S8**). Concentration 0.04 pmol/mL was used as a Limit of Blank, while concentration 0.4 pmol/mL (31 ng/mL) was chosen as a Limit of Detection (see **Supplemental Table S14** for details). MA, manual peak area.



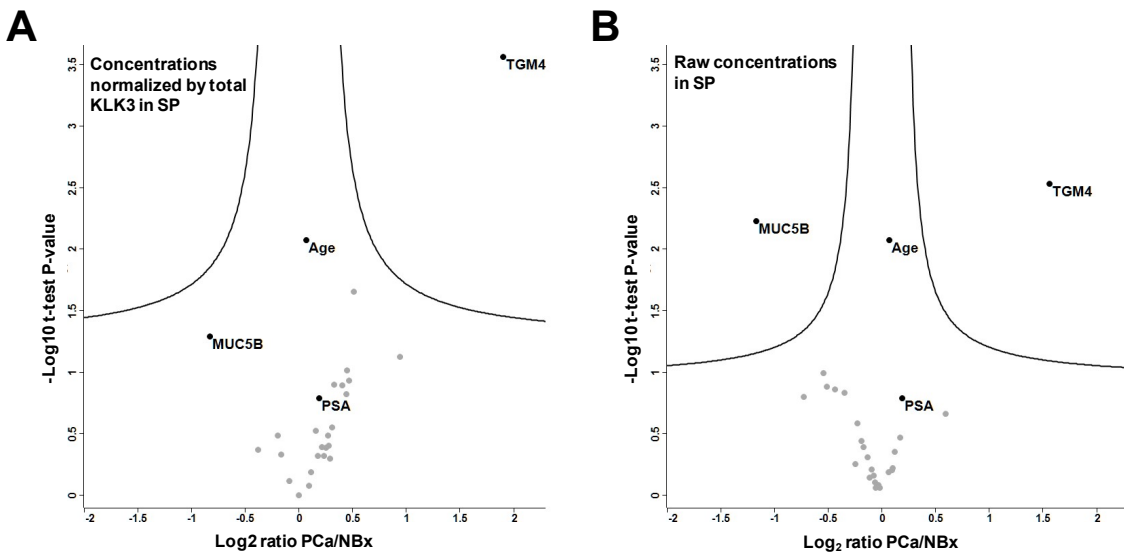
Supplemental Figure S10. ROC curves for selected proteins (Figure 3 and Table 1) for differentiation between negative biopsy (N=67) and PCa (N=152) groups based on SRM analysis in SP.



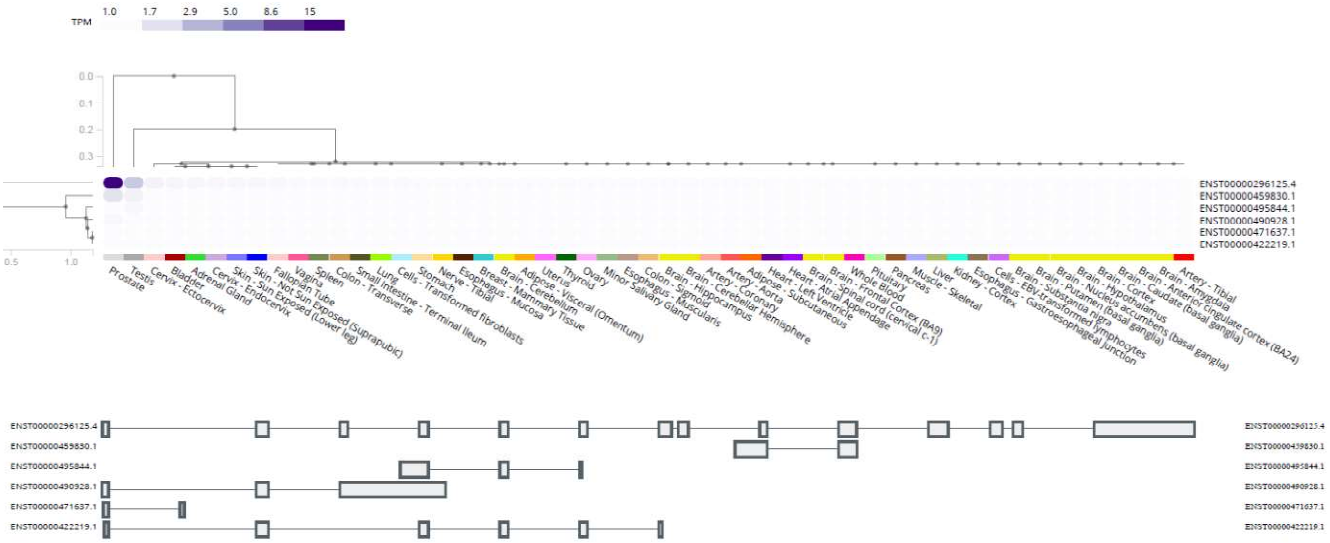
Supplemental Figure S11. Distribution of candidate proteins with respect to their initial selection in the discovery phase, as well as qualification and verification phases. Overlaps included: TGM4_HUMAN – 4 approaches (transcriptomics, proteomics, tissue specificity and androgen regulation); AMPN_HUMAN – 3 approaches; and LYSC_HUMAN, PI15_HUMAN, PPA5_HUMAN and IBP5_HUMAN – 2 approaches.



Supplemental Figure S12. Distribution of candidate proteins with respect to the dynamic range of SP proteome. Selected prostate-specific proteins are shown.



Supplemental Figure S13. Impact of normalization by total KLK3 in SP. Normalization by total KLK3 in SP (**A**) improved TGM4 performance (t-test P-value 0.0003 versus 0.003) versus unnormalized concentrations (**B**). Curves denote boundaries with 5% FDR, and performance of blood serum PSA and age are presented for comparison. Normalization had a noticeable impact on MUC5B protein, a mucin secreted by the Cowper's glands. Since moderate increase in performance may not justify the need to measure an additional protein in a clinical test, normalization by total KLK3 was not further considered in data analysis.



Supplemental Figure S14. Expression of *TGM4* transcripts in human tissues according to GTeX data (www.gtexportal.org/home/gene/TGM4). Only the canonical isoform ENST00000296125 which emcodes the only protein isoform of TGM4 (P49221-1; 684 aa; 77,145 Da) was highly expressed in the prostate tissue.