

# Analysis of Over 140,000 European Descendants Identifies Genetically Predicted Blood Protein Biomarkers Associated with Prostate Cancer Risk



Lang Wu<sup>1,2</sup>, Xiang Shu<sup>2</sup>, Jiandong Bao<sup>2</sup>, Xingyi Guo<sup>2</sup>; the PRACTICAL, CRUK, BPC3, CAPS, PEGASUS Consortia, Zsofia Kote-Jarai<sup>3</sup>, Christopher A. Haiman<sup>4</sup>, Rosalind A. Eeles<sup>3</sup>, and Wei Zheng<sup>2</sup>

## Abstract

Several blood protein biomarkers have been associated with prostate cancer risk. However, most studies assessed only a small number of biomarkers and/or included a small sample size. To identify novel protein biomarkers of prostate cancer risk, we studied 79,194 cases and 61,112 controls of European ancestry, included in the PRACTICAL/ELLIPSE consortia, using genetic instruments of protein quantitative trait loci for 1,478 plasma proteins. A total of 31 proteins were associated with prostate cancer risk including proteins encoded by *GSTP1*, whose methylation level was shown previously to be associated with prostate cancer risk, and *MSMB*, *SPINT2*, *IGF2R*, and *CTSS*, which were previously implicated as potential target genes of prostate cancer risk variants identified in genome-wide association studies. A total of 18 proteins

inversely correlated and 13 positively correlated with prostate cancer risk. For 28 of the identified proteins, gene somatic changes of short indels, splice site, nonsense, or missense mutations were detected in patients with prostate cancer in The Cancer Genome Atlas. Pathway enrichment analysis showed that relevant genes were significantly enriched in cancer-related pathways. In conclusion, this study identifies 31 candidates of protein biomarkers for prostate cancer risk and provides new insights into the biology and genetics of prostate tumorigenesis.

**Significance:** Integration of genomics and proteomics data identifies biomarkers associated with prostate cancer risk.

## Introduction

Prostate cancer is the second most frequently diagnosed malignancy and the fifth leading cause of cancer-related mortality among males worldwide (1). In the United States, there were 164,690 estimated new prostate cancer cases and 29,430 estimated deaths due to prostate cancer in 2018, making it a malignancy with the highest incidence and second highest mortality in males (2). The survival rate is higher when cancer is diagnosed

at a localized stage while it drops substantially when prostate cancer is diagnosed at a metastatic stage (3). Biomarkers are needed for screening and the early detection of prostate cancer. PSA has been used widely for prostate cancer screening (4, 5); however, there are controversies in using PSA screening due to the lack of a clear cut-off point for high sensitivity and specificity (6–8), unclear benefit in reducing mortality in some populations (9–11), and overdiagnosis of prostate cancer (12). Thus, there is a critical need to identify additional screening biomarkers aiming to reduce the mortality of prostate cancer.

Several other protein biomarkers measured in blood have been reported to be potentially associated with prostate cancer risk, such as IGF-1, IGFBP1/2, and IL6 (13–16). However, findings have been inconsistent from previous studies. Most existing studies have assessed only a small number of candidates. With the recent development of proteomics technology, there have been several studies searching the whole proteome to identify novel biomarkers for prostate cancer early detection and diagnosis (17–20). These studies have generated some promising findings. However, these have only included a relatively small number of subjects as it is expensive to profile the proteome in a large population-based study. More importantly, there are multiple limitations that are commonly encountered in conventional epidemiologic studies, including selection bias, potential confounding, and reverse causation. These limitations may explain some of the inconsistent results from previous studies.

To reduce these biases, we used genetic variants associated with blood protein levels as the instruments to assess the associations between genetically predicted protein levels and prostate cancer

<sup>1</sup>Cancer Epidemiology Division, Population Sciences in the Pacific Program, University of Hawaii Cancer Center, University of Hawaii at Manoa, Honolulu, Hawaii. <sup>2</sup>Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee. <sup>3</sup>Division of Genetics and Epidemiology, The Institute of Cancer Research, and The Royal Marsden NHS Foundation Trust, London, United Kingdom. <sup>4</sup>Department of Preventive Medicine, University of Southern California, Los Angeles, California.

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Members from the PRACTICAL, CRUK, BPC3, CAPS, and PEGASUS consortia are provided in the Supplementary Data.

**Corresponding Author:** Wei Zheng, Vanderbilt University Medical Center, 2525 West End Ave, 8th floor, Nashville 37203-1738, TN. Phone: 615-936-0682; Fax: 615-343-0719; E-mail: [wei.zheng@vanderbilt.edu](mailto:wei.zheng@vanderbilt.edu)

Cancer Res 2019;79:4592–8

doi: 10.1158/0008-5472.CAN-18-3997

©2019 American Association for Cancer Research.

risk. Because of the random assortment of alleles transferred from parents to offspring during gamete formation, this approach should be less susceptible to selection bias, reverse causation, and confounding effects. Over the past few years, genome-wide association studies (GWAS) have identified hundreds of protein quantitative loci (pQTL; refs. 21, 22). With a large sample size, many of these genetic variants can serve as strong instrumental variables for evaluating the associations of genetically predicted protein levels with prostate cancer risk. Herein, we report results from the first large study investigating the associations between genetically predicted blood protein levels and prostate cancer risk using genetic instruments. We used the data from 79,194 cases and 61,112 controls of European descent included in GWAS consortia PRACTICAL, CRUK, CAPS, BPC3, and PEGASUS, as described previously (23).

## Materials and Methods

A literature search was performed to identify the GWAS that uncovered genetic variants that were significantly associated with protein levels. After careful evaluation, the study conducted by Sun and colleagues represents the largest and most comprehensive study to date (24). By using the data from two subcohorts of 2,731 and 831 healthy European ancestry participants from the INTERVAL study, Sun and colleagues identified 1,927 genetic associations with 1,478 proteins at a stringent significance level (24). The detailed information of this study has been described elsewhere (24). In brief, an aptamer-based multiplex protein assay (SOMAscan) was used to quantify 3,620 plasma proteins. The robustness of the protein measurements was verified using several methods (24). Genotypes were measured using the Affymetrix Axiom UK Biobank array, which were further imputed using a combined reference panel from 1000 Genomes and UK10K. pQTL analyses were performed within each subcohort, with adjustments for age, sex, duration between blood draw and processing, and the first three principal components. After combining the association results from the two subcohorts via fixed-effects inverse-variance meta-analysis using METAL, the genetic associations between 1,927 variants and 1,478 proteins showed a meta-analysis of  $P < 1.5 \times 10^{-11}$ , and a consistent direction of effect and nominal significance ( $P < 0.05$ ). These pQTLs were used to construct the instrumental variables for assessing associations between protein levels and the risk of developing prostate cancer. When two or more variants located at the same chromosome were identified to be associated with a particular protein, we assessed the correlations of the SNPs using the Pairwise LD function of SNIIPA ([http://snipa.helmholtz-muenchen.de/snipa/index.php?task=pairwise\\_ld](http://snipa.helmholtz-muenchen.de/snipa/index.php?task=pairwise_ld)). For each protein, only SNPs independent of each other, as defined by  $r^2 < 0.1$  (based on 1000 Genomes Project phase III version 5 data focusing on European populations), were used to construct the instruments.

We used the summary statistics data for the association of genetic variants with prostate cancer risk that were generated from 79,194 prostate cancer cases and 61,112 controls of European ancestry in the consortia PRACTICAL, CRUK, CAPS, BPC3, and PEGASUS (23, 25). In brief, 46,939 prostate cancer cases and 27,910 controls were genotyped using OncoArray, which included 570,000 SNPs (<http://epi.grants.cancer.gov/oncoarray/>). Also included were data from several previous prostate cancer GWAS of European ancestry: UK stage 1 and stage 2; CaPS 1 and CaPS 2; BPC3; NCI PEGASUS; and iCOGS. These genotype data were

imputed using the June 2014 release of the 1000 Genomes Project data as a reference. Logistic regression summary statistics were then meta-analyzed using an inverse variance fixed effect approach.

For estimating the association between genetically predicted circulating protein levels and prostate cancer risk, the inverse variance weighted (IVW) method, using summary statistics results, was used (26). The beta coefficient of the association between genetically predicted protein levels and prostate cancer risk was estimated using  $\sum_i \beta_{i,GX} * \beta_{i,GY} * \sigma_{i,GY}^{-2} / (\sum_i \beta_{i,GX}^2 * \sigma_{i,GY}^{-2})$ , and the corresponding SE was estimated using  $1 / (\sum_i \beta_{i,GX}^2 * \sigma_{i,GY}^{-2})^{0.5}$ . Here,  $\beta_{i,GX}$  represents the beta coefficient of the association between *i*th SNP and the protein of interest generated from the pQTL study by Sun and colleagues;  $\beta_{i,GY}$  and  $\sigma_{i,GY}$  represent the beta coefficient and SE, respectively, for the association between *i*th SNP and prostate cancer risk in the prostate cancer GWAS. The association OR, confidence interval (CI), and *P* value were then estimated on the basis of the calculated beta coefficient and SE. A Benjamini–Hochberg FDR of  $< 0.05$  was used to adjust for multiple comparisons. Furthermore, to evaluate whether the identified associations between genetically predicted circulating protein levels and prostate cancer risk were independent of association signals identified in GWAS, we performed conditional analyses, adjusting for the closest risk SNPs identified in previous GWAS or fine-mapping studies. For this analysis, we performed GCTA-COJO analyses (version 1.26.0; refs. 27, 30) to calculate associations of SNPs with prostate cancer risk, after adjusting for the risk SNP of interest. We then reran the IVW analyses using the association estimates generated from conditional analyses.

For each of the genes encoding the proteins that are identified in our study in association with prostate cancer risk, we evaluated genetic variants/mutations/indels in prostate tumor tissues from patients with prostate cancer included in The Cancer Genome Atlas (TCGA). The somatic level genetic changes were analyzed using MuTect (31) and deposited to the TCGA data portal. Data were retrieved in April 2016, through the data portal. The proportion of assessed genes containing such somatic level genetic events tended to be enriched, when compared with the proportion of all protein-coding genes across the genome. Analysis was performed using MedCalc online software.

To further assess whether our identified prostate cancer-associated proteins are enriched in specific pathways, molecular and cellular functions, and networks, we performed an enrichment analysis of the genes encoding identified proteins using Ingenuity Pathway Analysis (IPA) software (32). The detailed methodology of this tool has been described elsewhere (32). In brief, an "enrichment" score (Fisher exact test *P* value) that measures overlap of observed and predicted regulated gene sets was generated for each of the tested gene sets. The most significant pathways and functions with an enrichment *P* value less than 0.05 were reported.

## Data availability

The OncoArray genotype data and relevant covariate information (i.e., ethnicity, country, principal components, etc.) for prostate cancer study are available in dbGAP (Accession no.: phs001391.v1.p1). In total, 47 of the 52 OncoArray studies, encompassing nearly 90% of the individual samples, are available. The previous meta-analysis summary results and genotype data are currently available in dbGAP (Accession no.: phs001081.v1.p1).

## Results

Of the pQTLs for 1,478 proteins assessed in this study, association results for prostate cancer risk were available for pQTLs of 1,469 proteins in the prostate cancer GWAS. For 1,106 of these proteins, only a single pQTL was identified. Two pQTLs were identified for 302 proteins and three or more pQTLs were identified for 71 proteins. Using the IVW method, we identified 31 proteins for which their genetically predicted levels were associated with prostate cancer risk at a FDR < 0.05 (Tables 1 and 2), including 22 encoded by genes located more than 500 Kb away from any reported prostate cancer risk variants identified in GWAS or fine-mapping studies (Table 1). The other nine associated proteins are encoded by genes located at previously reported prostate cancer risk loci (Table 2), including *MSMB*, *SPINT2*, *IGF2R*, and *CTSS*, which were previously implicated as candidate target genes of prostate cancer risk variants identified in GWAS (33–35). Interestingly, we also observed a significant association for glutathione S-transferase Pi, encoded by *GSTP1* (Table 2), whose methylation has been identified as a potential biomarker for prostate cancer (36). In our study, an inverse association between protein level and prostate cancer risk was detected for PSP-94, DcR3, IGF-II receptor, KDELB2, Cathepsin S, ZHX3, ZN175, GPC6, RM33, PIM1, WISP-3, NCF-2, ATF6A, laminin, glutathione S-transferase Pi, GNMT, LRRN1, and SNAB (ORs ranging from 0.69 to 0.97). Conversely, an association between a higher protein level and increased prostate cancer risk was identified for TACT, GRIA4, PDE4D, TIP39, SPINT2, MICB, IL21, ARFP2, RF1ML, TPST1, KLRF1, TM149, and NKp46 (ORs ranging from 1.11 to 1.23).

To determine whether the identified significant associations between genetically predicted protein levels and prostate cancer risk were independent of GWAS-identified association signals, we performed conditional analyses adjusting for the GWAS-identified risk SNPs closest to the genes encoding our identified proteins (Tables 1 and 2; ref. 27). For proteins listed in Table 1, the analysis could not be performed for three proteins due to lack of data, and for all other proteins, the associations remained essentially unchanged in the conditional analysis, suggesting these associations may be independent of GWAS-identified association signals. On the other hand, for proteins whose encoding genes locate at known prostate cancer risk loci, except for *IGF2R*, all other associations were no longer statistically significant when conditioning on GWAS-identified risk SNPs, suggesting these associations may be influenced by GWAS-identified association signals (Table 2).

By analyzing exome-sequencing data of prostate tumor-adjacent normal tissue and tumor tissue obtained from 498 patients with prostate cancer of TCGA, we observed somatic level changes of indels, nonsense mutations, splice site variations, or missense mutations in at least 1 patient for 28 of the 31 genes encoding identified associated proteins (enrichment  $P < 0.0001$  compared with the proportion of all protein-coding genes across the genome; Supplementary Table S1). In addition to the somatic missense mutations detected in 24 genes, indels were detected in four genes (*ARFP2*, *LRRN1*, *ZN175*, and *PDE4DIP*), splice site variations were detected in four genes (*IGF2R*, *IL21*, *MICB*, and *PTH2R*), and a nonsense mutation was detected in *KLRF1* (Supplementary Table S1). Although the majority of these somatic changes occurred in only 1 patient, a missense mutation in *PTH2* occurred in 9 patients (1.8%; Supplementary Table S1).

On the basis of the IPA analysis, several cancer-related functions were enriched for the genes encoding the associated proteins identified in this study (Supplementary Table S2). The top canonical pathways identified included STAT3 pathway ( $P = 4.54 \times 10^{-3}$ ), glutathione redox reactions I ( $P = 0.027$ ), glutathione-mediated detoxification ( $P = 0.030$ ), endoplasmic reticulum stress pathway ( $P = 0.031$ ), and tRNA splicing ( $P = 0.044$ ).

## Discussion

This is the first large-scale study to evaluate the associations of genetically predicted protein levels with prostate cancer risk using GWAS-identified pQTLs as instruments. We identified 31 proteins that demonstrated a statistically significant association with prostate cancer risk after FDR correction, including 22 whose encoding genes were located more than 500 Kb away from any reported prostate cancer risk variants. Our study provides novel information to improve the understanding of genetics and etiology for prostate cancer, and generates a list of promising proteins as potential biomarkers for early detection of prostate cancer, the most common malignancy among men in most countries around the world.

In this article, we used data from large GWASs involving 79,194 prostate cancer cases and 61,112 controls. The purpose and approach of this analysis are different from those of the study of Schumacher and colleagues (23). In the GWAS, investigators evaluated each genetic variant across the genome one at a time, aiming to identify novel susceptibility variants showing an association with prostate cancer risk (23). This work aimed to use genetically predicted protein expression levels as the testing unit to identify prostate cancer-associated proteins. We used a protein-based approach that aggregates the effects of several SNPs into one testing unit whenever possible. The analysis unit for our study is proteins, while the analysis unit in GWAS by Schumacher and colleagues (23) is genetic variants.

Previous research suggests that PSA, IGF-1, IGFBP1/2, and IL6 measured in blood may be associated with prostate cancer risk. For PSA, IGFBP1/2, and IL6, there was no corresponding pQTL identified in the study conducted by Sun and colleagues (24), thus they were not investigated in this study. For IGF-1, by using its pQTL rs74480769 as instrument, we did not observe a significant association with prostate cancer risk (OR = 0.98; 95% CI, 0.90–1.07;  $P = 0.70$ ). The inconsistent finding of IGF-1 with previous studies could be due to either a weak instrument used in this study or potential confounded estimates of associations in previous studies using a conventional epidemiologic design. Indeed, the significant positive association of IGF-1 was observed in the Health Professionals Follow-up Study (15), but not in the Prostate Cancer Prevention Trial (14). Further research would be needed to better understand the relationship between these proteins and prostate cancer.

In this large study, we identified 22 associated proteins, of which, the encoding genes are located at genomic loci not mapped by any of the previous GWAS. The statistical power in our study is larger than GWAS because (i) the number of comparisons is smaller in our study than GWAS and thus we could use a less stringent statistical significance threshold rather than  $5 \times 10^{-8}$  in GWAS and (ii) the predicted protein levels are continuous variables, which improves statistical power. It is worth noting that nine of the proteins identified in this study are encoded by genes located at the GWAS-identified loci. For many of the identified proteins,

**Table 1.** Twenty-two novel protein-prostate cancer associations for proteins whose encoding genes are located at genomic loci at least 500 kb away from any GWAS-identified prostate cancer risk variants

Protein	Protein full name	Protein-encoding gene	Region	Index SNP(s) <sup>a</sup>	Distance of gene to the index SNP (kb)	Instrument variants	Type of pQTL	OR (95% CI) <sup>b</sup>	P	FDR P value <sup>c</sup>	P value after adjusting for risk SNP <sup>d</sup>
ATF6A	Cyclic AMP-dependent transcription factor ATF-6 alpha	<i>ATF6</i>	1q23.3	rs4845695	6,824	rs8111, rs61738953	trans, trans	0.90 (0.86-0.95)	1.31 × 10 <sup>-4</sup>	9.18 × 10 <sup>-3</sup>	1.31 × 10 <sup>-4</sup>
NCF-2	Neutrophil cytosol factor 2	<i>NCF2</i>	1q25.3	rs199774366	20,932	rs4632248, rs28929474	trans, trans	0.95 (0.92-0.97)	9.93 × 10 <sup>-5</sup>	7.29 × 10 <sup>-3</sup>	NA <sup>e</sup>
Laminin	Laminin	<i>LAMC1</i>	1q25.3	rs199774366	21,377	rs62199218, rs4129858	trans, cis	0.93 (0.89-0.97)	4.16 × 10 <sup>-4</sup>	0.03	NA <sup>e</sup>
RM33	39S ribosomal protein L33_mitochondrial	<i>MRPL33</i>	2p23.2	rs13385191	7,106	rs28929474	trans	0.93 (0.90-0.96)	9.61 × 10 <sup>-6</sup>	7.43 × 10 <sup>-4</sup>	9.42 × 10 <sup>-6</sup>
LRRNI	Leucine-rich repeat neuronal protein 1	<i>LRRNI</i>	3p26.2	rs2660753	83,221	rs4293558, rs6801789	trans, cis	0.97 (0.95-0.99)	7.21 × 10 <sup>-4</sup>	0.04	7.21 × 10 <sup>-4</sup>
TACT	T-cell surface protein tactile	<i>CD96</i>	3q13.13-3q13.2	rs7611694	1,891	rs3132451	trans	1.22 (1.16-1.29)	1.02 × 10 <sup>-12</sup>	3.75 × 10 <sup>-10</sup>	1.02 × 10 <sup>-12</sup>
IL21	IL21	<i>IL21</i>	4q27	rs34480284	17,469	rs12368181, rs3129897	trans, trans	1.11 (1.06-1.16)	7.77 × 10 <sup>-6</sup>	7.43 × 10 <sup>-4</sup>	NA <sup>e</sup>
PDE4D	cAMP-specific 3,5-cyclic phosphodiesterase 4D	<i>PDE4D</i>	5q11.2-5q12.1	rs1482679	13,879	rs3132451	trans	1.17 (1.12-1.22)	1.02 × 10 <sup>-12</sup>	3.75 × 10 <sup>-10</sup>	1.02 × 10 <sup>-12</sup>
GNMT	Glycine N-methyltransferase	<i>GNMT</i>	6p21.1	rs4717148	763	rs57736976	cis	0.93 (0.89-0.97)	6.80 × 10 <sup>-4</sup>	0.04	2.78 × 10 <sup>-4</sup>
PI3I	Serine/threonine-protein kinase pim-1	<i>PI3I</i>	6p21.2	rs9469899	2,345	rs28929474	trans	0.88 (0.83-0.93)	9.61 × 10 <sup>-6</sup>	7.43 × 10 <sup>-4</sup>	9.42 × 10 <sup>-6</sup>
WISP-3	WNT1-inducible signaling pathway protein 3	<i>WISP3</i>	6q21	rs2273669	3,090	rs28929474	trans	0.83 (0.77-0.90)	9.61 × 10 <sup>-6</sup>	7.43 × 10 <sup>-4</sup>	9.42 × 10 <sup>-6</sup>
TPST1	Protein-tyrosine sulfotransferase 1	<i>TPST1</i>	7q11.21	rs56232506	18,233	rs513829	cis	1.14 (1.06-1.22)	5.23 × 10 <sup>-4</sup>	0.03	5.43 × 10 <sup>-4</sup>
ARFP2	Arfaftin-2	<i>ARFP2</i>	11p15.4	rs61890184	1,045	rs28929474	trans	1.23 (1.12-1.35)	9.61 × 10 <sup>-6</sup>	7.43 × 10 <sup>-4</sup>	9.42 × 10 <sup>-6</sup>
GRIA4	Glutamate receptor 4	<i>GRIA4</i>	11q22.3	rs1800057	2,291	rs3132451	trans	1.17 (1.12-1.22)	1.02 × 10 <sup>-12</sup>	3.75 × 10 <sup>-10</sup>	1.02 × 10 <sup>-12</sup>
KLRF1	Killer cell lectin-like receptor subfamily F member 1	<i>KLRF1</i>	12p13.31	rs2066827	2,873	rs11708955, rs62143194	trans, trans	1.13 (1.05-1.20)	5.74 × 10 <sup>-4</sup>	0.03	5.74 × 10 <sup>-4</sup>
GPC6	Glypican-6	<i>GPC6</i>	13q31.3-13q32.1	rs9600079	20,151	rs28929474	trans	0.81 (0.73-0.89)	9.61 × 10 <sup>-6</sup>	7.43 × 10 <sup>-4</sup>	9.42 × 10 <sup>-6</sup>
TM149	IGF-like family receptor 1	<i>IGFLR1</i>	19q13.12	rs8102476	2,502	rs12459634	cis	1.06 (1.02-1.09)	7.31 × 10 <sup>-4</sup>	0.04	4.68 × 10 <sup>-3</sup>
TIP39	Tuberoinfundibular peptide of 39 residues	<i>PTH2</i>	19q13.33	rs2659124	1,428	rs375375234	trans	1.22 (1.13-1.32)	3.06 × 10 <sup>-7</sup>	4.99 × 10 <sup>-5</sup>	2.96 × 10 <sup>-7</sup>
ZNF75	Zinc finger protein 175	<i>ZNF75</i>	19q13.41	rs2735839	710	rs28929474	trans	0.91 (0.87-0.95)	9.61 × 10 <sup>-6</sup>	7.43 × 10 <sup>-4</sup>	9.42 × 10 <sup>-6</sup>
NKp46	Natural cytotoxicity triggering receptor 1	<i>NCR1</i>	19q13.42	rs103294	620	rs2278428	cis	1.16 (1.06-1.26)	9.91 × 10 <sup>-4</sup>	0.05	9.65 × 10 <sup>-4</sup>
SNAB	Beta-soluble NSF attachment protein	<i>NAPB</i>	20p11.21	rs11480453	7,945	rs4293558, rs7658970	trans, trans	0.91 (0.86-0.96)	9.77 × 10 <sup>-4</sup>	0.05	9.77 × 10 <sup>-4</sup>
ZHX3	Zinc fingers and homeoboxes protein 3	<i>ZHX3</i>	20q12	rs11480453	8,460	rs1694123	trans	0.79 (0.71-0.88)	9.38 × 10 <sup>-6</sup>	7.43 × 10 <sup>-4</sup>	9.67 × 10 <sup>-6</sup>

NOTE: NA<sup>e</sup>, the adjacent risk variant is not available in the 1000 Genomes Project data.

<sup>a</sup>Closest risk variant identified in previous GWAS or fine-mapping studies for prostate cancer risk.

<sup>b</sup>OR and CI per one SD increase in genetically predicted protein.

<sup>c</sup>FDR P value, FDR-adjusted P value; associations with a FDR P ≤ 0.05 considered statistically significant.

<sup>d</sup>Using COJO method (27).

**Table 2.** Nine novel protein–prostate cancer associations for proteins whose encoding genes are located at genomic loci within 500 kb of previous GWAS-identified prostate cancer risk variants

Protein	Protein name	Region	Index SNP(s) <sup>a</sup>	Distance of gene to the index SNP (kb)	Instrument variants	Type of pQTL	OR (95% CI) <sup>b</sup>	P	FDR P value <sup>c</sup>	P value after adjusting for risk SNPs <sup>d</sup>
Cathepsin S	Cathepsin S	1q21.3	rs17599629	44	rs41271951	cis	0.91 (0.88–0.95)	$2.73 \times 10^{-7}$	$4.99 \times 10^{-5}$	0.16
MICB	MHC class I polypeptide-related sequence B	6p21.33	rs2596546	133	rs3134900	cis	1.09 (1.05–1.12)	$2.07 \times 10^{-6}$	$2.76 \times 10^{-4}$	0.03
RFML	Peptide chain release factor 1-like, mitochondrial	6q25.2	rs3968480	109	rs503366	cis	1.18 (1.08–1.29)	$4.67 \times 10^{-4}$	0.03	0.21
IGF-II receptor	Cation-independent mannose-6-phosphate receptor	6q25.3	rs651164	47	rs629849	cis	0.92 (0.90–0.94)	$3.98 \times 10^{-10}$	$9.73 \times 10^{-8}$	$9.95 \times 10^{-11}$
PSP-94	Beta-microseminoprotein	10q11.22	rs10993994	0.002	rs541781976, rs10993994	trans, cis	0.81 (0.80–0.82)	$3.60 \times 10^{-155}$	$5.29 \times 10^{-152}$	NA <sup>e</sup>
Glutathione S-transferase Pi	Glutathione S-transferase P	11q13.2	rs12785905	399	rs1695, rs62143206	cis, trans	0.94 (0.91–0.97)	$5.91 \times 10^{-4}$	0.03	$3.12 \times 10^{-3}$
KDEL2	KDEL motif-containing protein 2	11q22.3	rs1800057	199	rs7491261	cis	0.89 (0.86–0.93)	$1.83 \times 10^{-8}$	$3.85 \times 10^{-6}$	0.42
SPINT2	Kunitz-type protease inhibitor 2	19q13.2	rs8102476	0	rs17354995	cis	1.05 (1.03–1.06)	$1.31 \times 10^{-6}$	$1.92 \times 10^{-4}$	0.07
DcR3	Tumor necrosis factor receptor superfamily member 6B	20q13.33	rs6062509	33	rs62217798	cis	0.69 (0.62–0.77)	$1.98 \times 10^{-11}$	$5.81 \times 10^{-9}$	0.05

NOTE: NA<sup>e</sup>, the adjacent risk variant is the corresponding pQTL.

<sup>a</sup>Closest risk variant(s) identified in previous GWAS or fine-mapping studies for prostate cancer risk.

<sup>b</sup>OR and CI per one SD increase in genetically predicted protein.

<sup>c</sup>FDR P value, FDR-adjusted P value; associations with a FDR  $P \leq 0.05$  considered statistically significant.

<sup>d</sup>Using COJO method (27).

the genetic instrument includes trans pQTL(s) beyond only cis pQTL(s) (Tables 1 and 2), thus explaining why the corresponding protein-coding genes are not always at known susceptibility loci. *In vitro/in vivo* studies and human studies have suggested that some of these novel genes may play an important role in prostate tumorigenesis. For example, an interchromosomal interaction between a known prostate cancer risk locus, 8q24, and *CD96* was observed by the use of a chromosome conformation capture-based multi-target sequencing technology (37). *GPC6* was found to be recurrently altered across tumors of patients with advanced and lethal prostate cancer (38). *PDE4D* was shown to function as a proliferation-promoting factor in prostate cancer and was over-expressed in human prostate carcinoma (39); its inhibition had been shown to decrease prostate cancer cell growth (40). *ATF6*, which is related to the unfolded protein response, was observed to be downregulated in high-grade prostatic intraepithelial neoplasia compared with normal prostate samples (41).

Of the nine associated proteins, of which, the encoding genes are located at GWAS-identified prostate cancer risk loci, several have also been found to potentially play functional roles in prostate cancer development. For example, the decreased *GSTP1* expression was observed to accompany human prostatic carcinogenesis (42). It is highly expressed in benign prostate glands while tends to not express in prostate cancer glands (43). *MSMB* encodes MSP for prostatic secretory protein of 94 amino acids, which is secreted by the prostate and functions as a suppressor of tumor growth and metastasis (44). Besides the study of Sun and colleagues (24), several other studies also support the potential of MSP as a serum marker for the early detection of high-grade prostate cancer (45, 46). The decreased expression of *IGF2R* was thought to be partly responsible for the increased growth of LNCaP human prostate cancer cells (47). In a mouse model, the mRNA of *IGF2R* was significantly decreased in metastatic prostate lesions and androgen-independent prostate cancer (48). By analyzing patient samples, it was identified that the loss of the heterozygosity of *IGF2R* was an early event in the development of prostate cancer (49). In *in vivo* and human studies, it was suggested that the shedding of MICB might contribute to the impairment of natural killer cell antitumor immunity in prostate cancer formation (50, 51). These previous studies provide support for a potential role of these genes in prostate carcinogenesis.

The sample size for the main association analysis of our study was large, providing high statistical power to detect the protein–prostate cancer associations. Also, the design of using genetic instruments reduces biases, such as selection bias and potential confounding, and eliminates potential influence due to reverse causation. On the other hand, there are several potential limitations of our study. The possibility of pleiotropy effect cannot be excluded. For example, rs28929474, which was the instrument for proteins ZN175, ARFP2, *GPC6*, *RM33*, *PIM1*, and *WISP-3*, as well as one of the two variants constituting an instrument for *NCF-2*, was also reported to be associated with several other traits, including glycoprotein acetyls (52–54). Similarly, rs429358, which was included in the instruments of *LRRN1* and *SNAB*, was associated with cerebral amyloid deposition and red cell distribution width (55, 56); rs62143206, which was included in the instrument of glutathione S-transferase Pi, was also associated with the monocyte percentage of white cells and the granulocyte percentage of myeloid white cells (55). Further studies will be needed to validate our identified protein–prostate cancer associations. Second, our analysis was constrained by the pQTLs

identified in previous GWAS of circulating protein levels, and thus we were unable to evaluate some important protein biomarkers for prostate cancer as discussed previously. We anticipated that additional protein biomarkers could be identified using newly identified pQTLs in the future. Furthermore, this work generates a list of promising protein candidates that show an association with prostate cancer, which can be investigated further in future studies that directly measure levels of these proteins. Identification of circulating protein biomarkers should be useful for prostate cancer risk assessment.

In conclusion, in a large-scale study assessing associations between genetically predicted circulating protein levels and prostate cancer, we identified multiple novel proteins showing a significant association. Further investigation of these proteins will provide additional insight into the biology and genetics of prostate cancer and facilitate the development of appropriate biomarker panels for the early detection of prostate cancer.

### Disclosure of Potential Conflicts of Interest

R.A. Eeles has received speakers bureau honoraria from GUASCO to San Francisco in Jan 2016, Janssen Nov 2017, and ASCO, Chicago, June 2018. No potential conflicts of interest were disclosed by the other authors.

### Authors' Contributions

**Conception and design:** L. Wu, Z. Kote-Jarai, W. Zheng  
**Development of methodology:** L. Wu, X. Guo, W. Zheng  
**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** the PRACTICAL CRUK, BPC3, CAPS, PEGASUS consortia, Z. Kote-Jarai, C.A. Haiman

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** L. Wu, J. Bao, X. Guo, the PRACTICAL CRUK, BPC3, CAPS, PEGASUS consortia, W. Zheng  
**Writing, review, and/or revision of the manuscript:** L. Wu, X. Shu, the PRACTICAL CRUK, BPC3, CAPS, PEGASUS consortia, Z. Kote-Jarai, R.A. Eeles, W. Zheng  
**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** the PRACTICAL CRUK, BPC3, CAPS, PEGASUS consortia, W. Zheng  
**Study supervision:** W. Zheng

### Acknowledgments

The authors thank Jirong Long and Wanqing Wen of the Vanderbilt University School of Medicine for their help for this study. The authors also would like to thank all of the individuals for their participation in the parent studies and all the researchers, clinicians, technicians, and administrative staff for their contribution to the studies. The data analyses were conducted using the Advanced Computing Center for Research and Education at Vanderbilt University. This project at Vanderbilt University Medical Center was supported in part by funds from the Anne Potter Wilson endowment. L. Wu was supported by NCI K99 CA218892 and the Vanderbilt Molecular and Genetic Epidemiology of Cancer training program (US NCI grant R25 CA160056 to X. Shu).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 19, 2018; revised April 21, 2019; accepted July 17, 2019; published first July 23, 2019.

### References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin* 2018;68:7–30.
- Gaudreau PO, Stagg J, Soulieres D, Saad F. The present and future of biomarkers in prostate cancer: proteomics, genomics, and immunology advancements. *Biomark Cancer* 2016;8:15–33.
- Catalona WJ, Smith DS, Ratliff TL, Basler JW. Detection of organ-confined prostate cancer is increased through prostate-specific antigen-based screening. *JAMA* 1993;270:948–54.
- Antenor JA, Han M, Roehl KA, Nadler RB, Catalona WJ. Relationship between initial prostate specific antigen level and subsequent prostate cancer detection in a longitudinal screening study. *J Urol* 2004;172:90–3.
- Thompson IM, Ankerst DP, Chi C, Lucia MS, Goodman PJ, Crowley JJ, et al. Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. *JAMA* 2005;294:66–70.
- Parekh DJ, Ankerst DP, Troyer D, Srivastava S, Thompson IM. Biomarkers for prostate cancer detection. *J Urol* 2007;178:2252–9.
- Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. *N Engl J Med* 2004;350:2239–46.
- Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Zappa M, Nelen V, 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. *Lancet* 2014;384:2027–35.
- Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* 2009;360:1320–8.
- Andriole GL, Crawford ED, Grubb RL III, Buys SS, Chia D, Church TR, et al. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* 2009;360:1310–9.
- Draisma G, Etzioni R, Tsodikov A, Mariotto A, Wever E, Gulati R, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J Natl Cancer Inst* 2009;101:374–83.
- Nguyen DP, Li J, Tewari AK. Inflammation and prostate cancer: the role of interleukin 6 (IL-6). *BJU Int* 2014;113:986–92.
- Neuhouser ML, Platz EA, Till C, Tangen CM, Goodman PJ, Kristal A, et al. Insulin-like growth factors and insulin-like growth factor-binding proteins and prostate cancer risk: results from the prostate cancer prevention trial. *Cancer Prev Res* 2013;6:91–9.
- Cao Y, Nimptsch K, Shui IM, Platz EA, Wu K, Pollak MN, et al. Prediagnostic plasma IGFBP-1, IGF-1 and risk of prostate cancer. *Int J Cancer* 2015;136:2418–26.
- Ercole CJ, Lange PH, Mathisen M, Chiou RK, Reddy PK, Vessella RL. Prostatic specific antigen and prostatic acid phosphatase in the monitoring and staging of patients with prostatic cancer. *J Urol* 1987;138:1181–4.
- Tanase CP, Codrini E, Popescu ID, Mihai S, Enciu AM, Necula LG, et al. Prostate cancer proteomics: current trends and future perspectives for biomarker discovery. *Oncotarget* 2017;8:18497–512.
- McLerran D, Grizzle WE, Feng Z, Thompson IM, Bigbee WL, Cazares LH, et al. SELDI-TOF MS whole serum proteomic profiling with IMAC surface does not reliably detect prostate cancer. *Clin Chem* 2008;54:53–60.
- Byrne JC, Downes MR, O'Donoghue N, O'Keane C, O'Neill A, Fan Y, et al. 2D-DIGE as a strategy to identify serum markers for the progression of prostate cancer. *J Proteome Res* 2009;8:942–57.
- Qingyi Z, Lin Y, Junhong W, Jian S, Weizhou H, Long M, et al. Unfavorable prognostic value of human PEDF decreased in high-grade prostatic intraepithelial neoplasia: a differential proteomics approach. *Cancer Invest* 2009;27:794–801.
- Enroth S, Johansson A, Enroth SB, Gyllensten U. Strong effects of genetic and lifestyle factors on biomarker variation and use of personalized cut-offs. *Nat Commun* 2014;5:4684.
- Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun* 2017;8:14357.

23. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* 2018;50:928–36.
24. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. *Nature* 2018;558:73–9.
25. Wu L, Wang J, Cai Q, Cavazos TB, Emami NC, Long J, et al. Identification of novel susceptibility loci and genes for prostate cancer risk: a transcriptome-wide association study in over 140,000 European descendants. *Cancer Res* 2019;79:3192–204.
26. Shu X, Bao J, Wu L, Long J, Shu XO, Guo X, et al. Evaluation of associations between genetically predicted circulating protein biomarkers and breast cancer risk. *Int J Cancer* 2019 Jul 2 [Epub ahead of print].
27. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of Anthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012;44:369–75.
28. Yang Y, Wu L, Shu X, Lu Y, Shu XO, Cai Q, et al. Genetic data from nearly 63,000 women of European descent predicts DNA methylation biomarkers and epithelial ovarian cancer risk. *Cancer Res* 2019;79:505–17.
29. Lu Y, Beeghly-Fadiel A, Wu L, Guo X, Li B, Schildkraut JM, et al. A transcriptome-wide association study among 97,898 women to identify candidate susceptibility genes for epithelial ovarian cancer risk. *Cancer Res* 2018;78:5419–30.
30. Wu L, Shi W, Long J, Guo X, Michailidou K, Beesley J, et al. A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. *Nat Genet* 2018;50:968–78.
31. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013;31:213–9.
32. Kramer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 2014;30:523–30.
33. Thibodeau SN, French AJ, McDonnell SK, Chevillet J, Middha S, Tillmans L, et al. Identification of candidate genes for prostate cancer-risk SNPs utilizing a normal prostate tissue eQTL data set. *Nat Commun* 2015;6:8653.
34. Penney KL, Sinnott JA, Tyekucheva S, Gerke T, Shui IM, Kraft P, et al. Association of prostate cancer risk variants with gene expression in normal and tumor tissue. *Cancer Epidemiol Biomarkers Prev* 2015;24:255–60.
35. Chang BL, Cramer SD, Wiklund F, Isaacs SD, Stevens VL, Sun J, et al. Fine mapping association study and functional analysis implicate a SNP in MSMB at 10q11 as a causal variant for prostate cancer risk. *Hum Mol Genet* 2009;18:1368–75.
36. Phe V, Cussenot O, Roupert M. Methylated genes as potential biomarkers in prostate cancer. *BJU Int* 2010;105:1364–70.
37. Du M, Yuan T, Schilter KF, Dittmar RL, Mackinnon A, Huang X, et al. Prostate cancer risk locus at 8q24 as a regulatory hub by physical interactions with multiple genomic loci across the genome. *Hum Mol Genet* 2015;24:154–66.
38. Kumar A, White TA, MacKenzie AP, Clegg N, Lee C, Dumpit RF, et al. Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc Natl Acad Sci U S A* 2011;108:17087–92.
39. Rahrman EP, Collier LS, Knutson TP, Doyal ME, Kuslak SL, Green LE, et al. Identification of PDE4D as a proliferation promoting factor in prostate cancer using a Sleeping Beauty transposon-based somatic mutagenesis screen. *Cancer Res* 2009;69:4388–97.
40. Powers GL, Hammer KD, Domenech M, Frantskevich K, Malinowski RL, Bushman W, et al. Phosphodiesterase 4D inhibitors limit prostate cancer growth potential. *Mol Cancer Res* 2015;13:149–60.
41. So AY, de la Fuente E, Walter P, Shuman M, Bernales S. The unfolded protein response during prostate cancer development. *Cancer Metastasis Rev* 2009;28:219–23.
42. Lee WH, Morton RA, Epstein JI, Brooks JD, Campbell PA, Bova GS, et al. Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. *Proc Natl Acad Sci U S A* 1994;91:11733–7.
43. Martignano F, Gurioli G, Salvi S, Calistri D, Costantini M, Gunelli R, et al. GSTP1 methylation and protein expression in prostate cancer: diagnostic implications. *Dis Markers* 2016;2016:4358292.
44. Beke L, Nuytten M, Van Eynde A, Beullens M, Bollen M. The gene encoding the prostatic tumor suppressor PSP94 is a target for repression by the Polycomb group protein EZH2. *Oncogene* 2007;26:4590–5.
45. Nam RK, Reeves JR, Toi A, Dulude H, Trachtenberg J, Emami M, et al. A novel serum marker, total prostate secretory protein of 94 amino acids, improves prostate cancer detection and helps identify high grade cancers at diagnosis. *J Urol* 2006;175:1291–7.
46. Reeves JR, Dulude H, Panchal C, Daigneault L, Ramnani DM. Prognostic value of prostate secretory protein of 94 amino acids and its binding protein after radical prostatectomy. *Clin Cancer Res* 2006;12:6018–22.
47. Schaffer BS, Lin MF, Byrd JC, Park JH, MacDonald RC. Opposing roles for the insulin-like growth factor (IGF)-II and mannose 6-phosphate (Man-6-P) binding activities of the IGF-II/Man-6-P receptor in the growth of prostate cancer cells. *Endocrinology* 2003;144:955–66.
48. Kaplan PJ, Mohan S, Cohen P, Foster BA, Greenberg NM. The insulin-like growth factor axis and prostate cancer: lessons from the transgenic adenocarcinoma of mouse prostate (TRAMP) model. *Cancer Res* 1999;59:2203–9.
49. Hu CK, McCall S, Madden J, Huang H, Clough R, Jirtle RL, et al. Loss of heterozygosity of M6P/IGF2R gene is an early event in the development of prostate cancer. *Prostate Cancer Prostatic Dis* 2006;9:62–7.
50. Liu G, Lu S, Wang X, Page ST, Higano CS, Plymate SR, et al. Perturbation of NK cell peripheral homeostasis accelerates prostate carcinoma metastasis. *J Clin Invest* 2013;123:4410–22.
51. Wu JD, Atteridge CL, Wang X, Seya T, Plymate SR. Obstructing shedding of the immunostimulatory MHC class I chain-related gene B prevents tumor formation. *Clin Cancer Res* 2009;15:632–40.
52. Kettunen J, Demirkan A, Wurtz P, Draisma HH, Haller T, Rawal R, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 2016;7:11122.
53. Lutz SM, Cho MH, Young K, Hersh CP, Castaldi PJ, McDonald ML, et al. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet* 2015;16:138.
54. Merkel PA, Xie G, Monach PA, Ji X, Ciavatta DJ, Byun J, et al. Identification of functional and expression polymorphisms associated with risk for antineutrophil cytoplasmic autoantibody-associated vasculitis. *Arthritis Rheumatol* 2017;69:1054–66.
55. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* 2016;167:1415–29.
56. Li QS, Parrado AR, Samtani MN, Narayan VA, Alzheimer's Disease Neuroimaging Initiative. Variations in the FRA10AC1 fragile site and 15q21 are associated with cerebrospinal fluid Aβ1-42 level. *PLoS One* 2015;10:e0134000.