Published in final edited form as:
Cancer Epidemiol Biomarkers Prev. 2015 January ; 24(1): 255–260. doi: 10.1158/1055-9965.EPI-14-0694-T.

Association of Prostate Cancer Risk Variants with Gene Expression in Normal and Tumor Tissue

Kathryn L. Penney[a,d], **Jennifer A. Sinnott**[a,c,d], **Svitlana Tyekucheva**[c,f], **Travis Gerke**[a], **Irene M. Shui**[a], **Peter Kraft**[a,c], **Howard D. Sesso**[a,e], **Matthew L. Freedman**[g,i], **Massimo Loda**[g,h,i], **Lorelei A. Mucci**[a,d], and **Meir J. Stampfer**[a,b,d]

[a]Department of Epidemiology, Harvard School of Public Health, Boston, MA USA

[b]Department of Nutrition, Harvard School of Public Health, Boston, MA USA

[c]Department of Biostatistics, Harvard School of Public Health, Boston, MA USA

[d]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA USA

[e]Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA USA

[f]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA USA

[g]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA USA

[h]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA USA

[i]The Broad Institute, Cambridge, MA USA

Abstract

Background—Numerous germline genetic variants are associated with prostate cancer risk, but their biological role is not well understood. One possibility is that these variants influence gene expression in prostate tissue. We therefore examined the association of prostate cancer risk variants with the expression of genes nearby and genome-wide.

Methods—We generated mRNA expression data for 20,254 genes with the Affymetrix GeneChip Human Gene 1.0 ST microarray from normal prostate (N=160) and prostate tumor (N=264) tissue from participants of the Physicians' Health Study and Health Professionals Follow-up Study. With linear models, we tested the association of 39 risk variants with nearby genes and all genes, and the association of each variant with canonical pathways using a global test.

Results—In addition to confirming previously reported associations, we detected several new significant ($p<0.05$) associations of variants with the expression of nearby genes including

Contact: Kathryn L. Penney, ScD, Channing Division of Network Medicine, Brigham and Women's Hospital, 181 Longwood Ave, Boston, MA 02115, Phone: 617-525-0860, Fax: 617-525-2008, kpenney@hsph.harvard.edu.

Conflicts of Interest: The authors disclose no potential conflicts of interest.

C2orf43, *ITGA6*, *MLPH*, *CHMP2B*, *BMPR1B*, and *MTL5*. Genome-wide, four genes (*MSMB*, *NUDT11*, *NEFM*, *KLHL33*) were significantly associated after accounting for multiple comparisons for each SNP ($p<2.5x10^{-6}$). Many more genes had a false discovery rate <10%, including *SRD5A1* and *PSCA*, and we observed significant associations with pathways in tumor tissue.

Conclusions—The risk variants were associated with several genes, including promising prostate cancer candidates and lipid metabolism pathways, suggesting mechanisms for their impact on disease. These genes should be further explored in biological and epidemiological studies.

Impact—Determining the biological role of these variants can lead to improved understanding of prostate cancer etiology and identify new targets for chemoprevention.

Keywords

prostate cancer; genetic risk variant; prostate tissue; gene expression; eQTL

Introduction

Numerous germline genetic risk variants have been linked to prostate cancer risk from genome-wide association studies (GWAS) (1–14). With the report from the PRACTICAL consortium in 2013, the number of prostate cancer risk variants is now >70 (15), a major step toward uncovering the genetic etiology of prostate cancer. Family and twin studies demonstrate that prostate cancer is highly heritable (16); these single nucleotide polymorphisms (SNPs) explain an ever increasing portion of this underlying heritability, currently about 30% (15). However, the biological function of these risk SNPs remains largely unknown given that the majority is located outside of protein coding regions. A critical next step in translating knowledge of identified SNPs to the prevention or treatment of prostate cancer is determining their biological mechanisms.

One possibility is that the risk variants are expression quantitative trait loci (eQTL), genetic loci that are associated with mRNA transcript levels. Few large studies have both SNP data and prostate tissue for gene expression studies. We recently showed that a prostate cancer risk SNP on chromosome 10q11, rs10993994, was significantly associated with mRNA expression of two nearby genes (17) in prostate tissue. Men with the risk allele had decreased expression of *MSMB* in both normal prostate and tumor tissue, and increased expression of *NCOA4* in normal prostate tissue only. A similar result for *MSMB* was observed by Lou *et al* (18). A study of 12 prostate cancer risk loci found that four acted as eQTLs. In addition to confirming the *MSMB* and *NCOA4* results, *NUDT11* was associated with rs5945619 and *SLC22A3* was borderline significantly associated with rs9364554 in both tumor and normal tissue; and *HNF1B* was associated with rs4430796 in normal tissue only (19). Harries et al. observed an association between rs6465657 and expression of nearby *LMTK2* (20), while Xu et al. found a proxy for rs12653946 to be strongly associated with the expression of nearby *IRX4* (21).

Published work has primarily focused on the expression of genes near the risk variants. While variants may have larger effects on nearby genes, and there is therefore more power to identify these effects, genetic polymorphisms can influence the expression of genes anywhere in the genome. This can happen either directly or downstream in a pathway, such as through a transcription factor (reviewed in (22)). We therefore examined the association of the risk variants with transcriptome-wide expression data in tumor and normal prostate tissue, performing a *cis* analysis (examining the association of the variants with nearby genes), a *trans* analysis (determining the association of the variants with all genes), and a pathway analysis.

Materials and Methods

Study Participants

Physicians' Health Study (PHS) and Health Professionals Follow-up Study (HPFS)—The men in the study are participants in two prospective studies ongoing for more than 25 years: the PHS and HPFS. PHS began in 1986 as a randomized, double-blind trial of aspirin and β-carotene in the prevention of cardiovascular disease and cancer among 22,071 initially healthy U.S. physicians (23). The HPFS, an ongoing prospective cohort study on the causes of cancer and heart disease in men, consists of 51,529 U.S. health professionals who were aged 40–75 years in 1986 (24). In both studies, men were excluded if they had any serious medical conditions at baseline including all cancers (except non-melanoma skin cancer).

The men in this study were diagnosed with incident, histologically confirmed prostate cancer between 1982 and 2004. Participants are followed through regular questionnaires to collect self-reported data on diet, lifestyle behaviors, medical history, and health outcomes, including prostate cancer. All prostate cancer cases in this study were verified through medical record and pathology review. Through this systematic medical record review we also abstract data on clinical information, including clinical stage and PSA at diagnosis. The Human Subjects Committee at Partners Healthcare and the Harvard School of Public Health approved these studies.

mRNA Expression Profiling—In both cohorts, we sought to retrieve archival formalin-fixed paraffin embedded (FFPE) specimens. The PHS and HPFS Tumor Cohort includes 2,200 men with prostate cancer for whom we have collected archival radical prostatectomy (RP) (95 percent) and trans-urethral resection of the prostate (TURP) (5 percent) specimens. For a subset of the tumor cohort, we undertook whole genome gene expression profiling as part of a study designed to identify expression signatures that can differentiate lethal from indolent prostate cancer. We sampled men from the Tumor Cohort using an extreme case design, which includes 116 men who died of their cancer or developed bony or distant metastases and 292 men who lived at least 8 years after prostate cancer diagnosis and were not diagnosed with metastases through 2012. For a subset of these men, we also profiled adjacent normal tissue. To conduct this profiling in FFPE tissues, whole transcriptome amplification was paired with microarray technologies. Briefly, RNA was extracted using the Biomek FxP automated platform with the Agencourt FormaPure FFPE kit (Beckman

Coulter, Indianapolis, IN). The mRNA was amplified using the WT-Ovation FFPE System V2 (NuGEN, San Carlos, CA), a whole transcriptome amplification system that allows for complete gene expression analysis from archives of FFPE samples known to harbor small and degraded RNA. Using a combination 5′ and random primer, reverse transcription created a cDNA/mRNA hybrid. The mRNA was subsequently fragmented, creating binding sites for DNA polymerase. Isothermal strand-displacement, using a proprietary DNA/RNA chimeric SPIA primer, amplified the cDNA. The cDNA was then fragmented and labeled with a terminal deoxynucleotidyl transferase covalently linked to biotin to prepare for microarray hybridzation. The labeled cDNA was then hybridized to a GeneChip Human Gene 1.0 ST microarray (Affymetrics, Santa Clara, CA).

For the expression profiles generated, we regressed out technical variables including mRNA concentration, age of the block, batch (96-well plate), percent of the probes on the array detectable above the background, log transformed average background signal, and the median of the perfect match probes for each probe intensity of the raw data. The residuals were then shifted to have the original mean expression values and normalized using the RMA method (25, 26). We mapped gene names to Affymetrix transcript cluster IDs using the NetAffx annotations as implemented in Bioconductor annotation package pd.hugene.1.0.st.v1; this resulted in 20,254 unique named genes. Gene expression data are available through Gene Expression Omnibus (GEO) accession number GSE62872.

Risk Variant Genotypes—The SNPs were genotyped on DNA extracted from whole blood as part of the National Cancer Institute funded Breast and Prostate Cancer Cohort Consortium (BPC3) using the TaqMan assay (Applied Biosystems, Foster City, CA) at the Harvard School of Public Health (Boston, MA); details on the SNP selection and genotyping are provided in (27). Prostate cancer participants included in the current study are all of European ancestry. To reduce missing data, we combined data for SNPs in very high linkage disequilibrium. For rs12418451, we used genotypes from either rs12418451 or rs10896438 (r^2=0.96 in HapMap CEU population); for rs2928679, we used genotypes from either rs2928679 or rs13264338 (r^2=0.97); for rs1983891, we used genotypes from either rs1983891 or rs9381080 (r^2=1.00); and for rs11672691, we used genotypes from either rs11672691 or rs11673591 (r^2=1.00). Additionally, eight SNPs from (27) were not genotyped in PHS, and were therefore excluded from this study due to the reduction in sample size. The average SNP and individual call rate were 95.2% and 95.3%, respectively.

Statistical Analysis

There were 264 participants with SNP and tumor tissue expression data; 160 of these cases also had normal prostate tissue expression data. Each SNP (3 genotype categories: 0, 1, or 2 copies of the risk allele) was compared to each gene (continuous expression) with a linear model test for trend. This analysis was performed separately for gene expression from prostate tumor and normal prostate tissue. First, a *cis* analysis was performed, examining the association between each SNP and “nearby” (500 kb up- and downstream) genes. For this more focused analysis, $p<0.05$ was considered statistically significant; despite the possibility for false positives using this liberal approach, these are reported to provide candidates for future studies. Next, a genome-wide analysis was performed, examining the association of

each SNP with all genes. For this analysis, a Bonferroni corrected p-value ($p<2.5x10^{-6}$ = 0.05/20,254 tests) was considered statistically significant, considering each SNP and tissue type as an independent hypothesis. Additionally, all associations with a false discovery rate less than 10% were reported as potentially interesting. Finally, a pathway analysis was performed using a global test model (R package globaltest; (28)). For this analysis, individuals were classified as either not carrying the risk allele, or carrying one or two copies of the risk allele to create a binary outcome variable. The Broad Institute MSigDB KEGG pathway classifications (v4.0) were used (29). Analysis was performed with R version 2.15.0. All p-values reported are two-sided and unadjusted for multiple comparisons. False discovery rate q-values were generated using the R package Qvalue [Alan Dabney < adabney@u.washington.edu>, John D. Storey < jstorey@u.washington.edu> and with assistance from Gregory R. Warnes < gregory_r_warnes@groton.pfizer.com> (2011). qvalue: Q-value estimation for false discovery rate control.]

Results

A description of the study participants is provided in Table 1. Information on the 39 risk SNPs and the frequencies in this population are in Supplementary Table 1. To examine *cis* relationships where a larger effect size and therefore more statistical power is expected, we specifically looked at the association of the risk SNPs with the expression of genes within a 1 Mb window (500 kb up- and downstream) around the SNP. In this focused analysis, we considered $p<0.05$ to be statistically significant. Several SNPs were associated with the expression of nearby genes in normal and tumor (Table 2). We observed new associations of SNPs with the genes in which they are located. rs13385191 was significantly associated with *C2orf43* in both tumor and normal, rs12621278 with *ITGA6* in tumor, and rs2292884 with *MLPH* in tumor (and borderline significant in normal). Additional novel associations with genes very nearby, where the SNP could be in a regulatory region or in linkage disequilibrium with another SNP within the gene, were also observed: rs7629490 with *CHMP2B* in tumor, rs17021918 with *BMPR1B* in tumor (and borderline in normal), rs4242382 with *POU5F1B* in tumor, rs7127900 with *ASCL2* in tumor, rs902774 with *KRT79* in normal, rs10896449 with *MTL5* in tumor and normal, and several others. Some significant associations observed in tumor that are not significant in normal tissue could be due to the larger number of tumor samples, so we also performed the *cis* analysis on the subset of the tumor that had normal tissue data available. The associations remained significant, or close to significant, even with the smaller sample size for the vast majority (Table 2). Results for all genes within the 1 Mb windows are reported in Supplementary Table 2 indicating additional associations that were just over the statistically significant threshold, but involve clear prostate cancer candidate genes (e.g. *KLK2* (p=0.07) and *KLK3* (p=0.07) with rs2735839 on chromosome 19 in tumor only). Box plots showing the gene expression levels by genotype for significant associations are presented in Supplementary Data.

We additionally replicated previously reported associations. rs10993994 was associated with *MSMB* ($p=3.9x10^{-7}$), both located on chromosome 10q11, and rs5945619 with *NUDT11* ($p=4.6x10^{-11}$), both on chromosome Xp11. These associations were also nominally associated in normal tissue (p=0.0017 and 0.0004, respectively), consistent with previous

findings (19). rs10993994 was associated with *NCOA4* in normal tissue with $p=5.7x10^{-5}$, but was null in tumor (p=0.64), consistent with a previous finding (19). We also confirmed other previous findings (20, 21): rs12653946 with *IRX4* (in tumor and normal), rs8102476 with *PPP1R14A* (in tumor and normal), rs6465657 with *LMTK2* (in tumor only), and rs5945619 with *NUDT10* (in tumor only). However, we did not confirm an association of rs4430796 with *HNF1B* in normal tissue (p=0.66) (19) or associations with *FAM83F* (p=0.29), *YIF1B* (p=0.37), *FAM98C* (p=0.17), *FOXP4* (p=0.19), *or TFEB* (p=0.78) in tumor tissue (21).

We performed a genome-wide ("*trans*") analysis of the risk SNPs across all transcripts. After Bonferroni correction for the 20,254 tests performed for each SNP, only five results remained statistically significant ($p<2.5x10^{-6}$). These include the *MSMB* and *NUDT11* associations in tumor, mentioned above, as well as *RBPMS2* (15q22) with rs11672691 (19q13; $p=2.22x10^{-6}$ in tumor). There were two associations in normal tissue: rs1859962 (8q24) with *NEFM* (8p21; $p=2.22x10^{-6}$) and rs1571801 (9q33) with *KLHL33* (14q11; $p=1.97x10^{-6}$). Associations using a more liberal threshold (false discovery rate (FDR) q<0.1) are presented in Supplementary Table 3. Of note, rs339331 and rs11672691 were each associated with tumor expression of dozens of genes at FDR q<0.1

While a risk SNP may not be strongly associated with individual genes, it may influence an entire pathway indirectly by impacting a transcription factor or other regulator of gene processes. After accounting for the 186 pathways tested, nine pathways were significantly ($p<2.7x10^{-4}$) associated with rs1512268 in tumor, most likely due to the overlap of several lipid-related genes within these pathways. These results are presented in Table 3, with all pathways significant with a less stringent threshold (p<0.001) in Supplementary Table 4.

Discussion

Prostate cancer is one of the most heritable malignancies. Numerous germline genetic risk variants have been linked to risk; however, their function is often difficult to discern because many lie in intergenic and intronic regions. eQTL studies of these risk loci have been performed, but have primarily focused on genes close to the SNP. We evaluated the association of the risk loci with genes and pathways across the genome in tumor and normal tissue.

For the *cis*- based analysis, in addition to previously reported associations, several novel associations were observed. Many of the associated genes are interesting candidates, including several that are transcription factors (*DLX2*, *IRX4*, *ASCL2*, *SP7*, *DMRTC2*). Other novel associations include rs17021981 with *BMPR1B*, a bone morphogenic protein. Since the primary metastatic site of prostate cancer is bone this gene could be relevant to progression as well. Though their expression did not vary in LNCaP cell lines following the addition of androgens, Hazelett et al. note that *BMPR1B*, as well as *IGF2R* and *CHMP2B* (both associated with nearby risk SNPs in this analysis in tumor), are androgen regulated genes (30). The association of rs12621278 with the expression of *C2orf43* was previously reported in liver cancer. This gene is associated with defective apolipoprotein B-100, which leads to hypercholesteremia (31). This may suggest a plausible mechanism for further

investigation since statin use has been inversely associated with prostate cancer (32). Although rs1512268 was not associated with any single genes, it was the only SNP significantly associated with pathways in tumor tissue. Many of these pathways were related to lipid metabolism, which may lend additional support to the hypothesis that lipids are involved with cancer development (33). *POU5F1B* has previously been associated with gonadoblastoma. While we observed rs4242382 to be significantly associated with *POU5F1B* in tumor and borderline significant in normal tissue, Breyer et al. recently found other risk SNPs in the 8q24 region to be associated with *POU5F1B* expression only in normal tissue (34). *MTL5*, found to be reduced in those with the risk allele in both tumor and normal, is a metallothionien-like protein that may be involved with cell growth and differentiation, as well as spermatogenesis. Several other genes, while not quite reaching the $p<0.05$ threshold, are also interesting candidates, including *KLK2* and *KLK3* mentioned above, since *KLK3* encodes the PSA gene, and *NKX3.1* was associated with rs2928679 in normal tissue only with borderline significance ($p=0.06$).

When examining all genes, we observed several highly significant associations. Some of these associations suggest excellent candidate genes for prostate cancer, particularly *SRD5A1* (steroid 5-alpha-reductase-1), the target of 5-alpha-reductase inhibitors, such as finasteride, which was associated with rs339331 in normal tissue (FDR q-value=0.045). *PSCA* (prostate stem cell antigen) was associated with rs1859962 in normal tissue (FDR q-value=0.099); this gene is a cell surface marker that has been found to be upregulated in prostate and other cancers.

We observe different associations for some SNPs in normal and tumor tissue, as others have reported previously. Results in normal may suggest earlier effects of the SNPs and involvement with tumor initiation; these associations could be lost in tumor due to the development of mutations, or the dysregulation of another gene, microRNA, or lncRNA, which then has a larger influence on the expression of these genes. Significant results in tumor only may point toward genes driving continued carcinogenesis; these associations could be masked in normal tissue because the gene expression is tightly regulated by other mechanisms that are lost during tumorigenesis.

Confirming many previously reported associations gives us confidence in our data and our findings. The lack of replication of some previously reported associations could be for several reasons, including limited power, expression technology with differing probe location or splice variants measured, or that the original report was a false positive. Our study does have some limitations that could lead to false negatives. While this study is large for a gene expression profiling study, the statistical power to detect the small effects anticipated is relatively low. Also, the risk SNPs could affect mRNA expression levels in a transitory way, and we are only able to capture one time point after cancer has already developed. Additionally, using a less stringent p-value cutoff of 0.05 in the *cis* analysis may have led to the reporting of some false positive associations; however, this provides a list of candidates to be confirmed in additional studies. Also, approximately 100 prostate cancer risk SNPs now have been identified; future analysis should not only attempt to confirm the results reported here but also expand to include the more recently identified SNPs.

The genes and pathways we identified that are associated with the risk SNPs can improve the biological understanding of prostate cancer development. These genes may additionally help explain the mechanism of epidemiological results and provide candidates for new treatment or prevention strategies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial Support:

The Physicians' Health Study was supported by grants CA34944, CA40360, CA097193, HL26490 and HL34595. The Health Professionals Follow-up Study was supported by grants CA133891, CA141298, P01CA055075. This study and the following authors were supported by 5U01CA098233 (P. Kraft), CA136578 (L.A. Mucci), CA141298 (M.J. Stampfer), CA131945 (M. Loda), and P50CA090381 (L.A. Mucci and S. Tyekucheva). J.A. Sinnott received support from T32CA09001. K.L. Penney and L.A. Mucci are supported by Prostate Cancer Foundation Young Investigator Awards. I.M. Shui was supported by a Department of Defense Prostate Cancer Research Fellowship. K.L. Penney was supported by the A. David Mazzone Research Awards Program.

We are grateful to the participants and staff of the Physicians' Health Study and Health Professionals Follow-Up Study for their valuable contributions, as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. In addition, this study was approved by the Connecticut Department of Public Health (DPH) Human Investigations Committee. Certain data used in this publication were obtained from the DPH. The authors assume full responsibility for analyses and interpretation of these data. We acknowledge Natalie Dupre, Elizabeth Nuttall, Michael Pitt, Sam Peisch, and Rosina Lis for their significant involvement with study design and implementation. Thank you to Edward Fox (in memoriam) at the DFCI Microarray Core.

References

1. Freedman ML, Haiman CA, Patterson N, et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc Natl Acad Sci U S A. 2006; 103:14068–73. [PubMed: 16945910]
2. Amundadottir LT, Sulem P, Gudmundsson J, et al. A common variant associated with prostate cancer in European and African populations. Nat Genet. 2006; 38:652–8. [PubMed: 16682969]
3. Haiman CA, Patterson N, Freedman ML, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet. 2007; 39:638–44. [PubMed: 17401364]
4. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet. 2007; 39:631–7. [PubMed: 17401366]
5. Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet. 2007; 39:645–9. [PubMed: 17401363]
6. Gudmundsson J, Sulem P, Steinthorsdottir V, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet. 2007; 39:977–83. [PubMed: 17603485]
7. Eeles RA, Kote-Jarai Z, Giles GG, et al. Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet. 2008; 40:316–21. [PubMed: 18264097]
8. Thomas G, Jacobs KB, Yeager M, et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet. 2008; 40:310–5. [PubMed: 18264096]
9. Gudmundsson J, Sulem P, Rafnar T, et al. Common sequence variants on 2p15 and Xp11. 22 confer susceptibility to prostate cancer. Nat Genet. 2008; 40:281–3. [PubMed: 18264098]
10. Yeager M, Chatterjee N, Ciampa J, et al. Identification of a new prostate cancer susceptibility locus on chromosome 8q24. Nat Genet. 2009; 41:1055–7. [PubMed: 19767755]

11. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. Nat Genet. 2009; 41:1122–6. [PubMed: 19767754]
12. Eeles RA, Kote-Jarai Z, Al Olama AA, et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. Nat Genet. 2009; 41:1116–21. [PubMed: 19767753]
13. Al Olama AA, Kote-Jarai Z, Giles GG, et al. Multiple loci on 8q24 associated with prostate cancer susceptibility. Nat Genet. 2009; 41:1058–60. [PubMed: 19767752]
14. Sun J, Zheng SL, Wiklund F, et al. Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. Nat Genet. 2008; 40:1153–5. [PubMed: 18758462]
15. Eeles RA, Olama AA, Benlloch S, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat Genet. 2013; 45:385–91. 91e1–2. [PubMed: 23535732]
16. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000; 343:78–85. [PubMed: 10891514]
17. Pomerantz MM, Shrestha Y, Flavin RJ, et al. Analysis of the 10q11 cancer risk locus implicates MSMB and NCOA4 in human prostate tumorigenesis. PLoS Genet. 2010; 6:e1001204. [PubMed: 21085629]
18. Lou H, Yeager M, Li H, et al. Fine mapping and functional analysis of a common variant in MSMB on chromosome 10q11. 2 associated with prostate cancer susceptibility. Proc Natl Acad Sci U S A. 2009; 106:7933–8. [PubMed: 19383797]
19. Grisanzio C, Werner L, Takeda D, et al. Genetic and functional analyses implicate the NUDT11, HNF1B, and SLC22A3 genes in prostate cancer pathogenesis. Proc Natl Acad Sci U S A. 2012; 109:11252–7. [PubMed: 22730461]
20. Harries LW, Perry JR, McCullagh P, Crundwell M. Alterations in LMTK2, MSMB and HNF1B gene expression are associated with the development of prostate cancer. BMC Cancer. 2010; 10:315. [PubMed: 20569440]
21. Xu X, Hussain WM, Vijai J, et al. Variants at IRX4 as prostate cancer expression quantitative trait loci. Eur J Hum Genet. 2013
22. Gilad Y, Rifkin SA, Pritchard JK. Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet. 2008; 24:408–15. [PubMed: 18597885]
23. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. N Engl J Med. 1989; 321:129–35. [PubMed: 2664509]
24. Giovannucci E, Pollak M, Liu Y, et al. Nutritional predictors of insulin-like growth factor I and their relationships to cancer in men. Cancer Epidemiol Biomarkers Prev. 2003; 12:84–9. [PubMed: 12582016]
25. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 2003; 31:e15. [PubMed: 12582260]
26. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4:249–64. [PubMed: 12925520]
27. Shui IM, Lindstrom S, Kibel AS, et al. Prostate Cancer (PCa) Risk Variants and Risk of Fatal PCa in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium. Eur Urol. 2014
28. Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. Bioinformatics. 2004; 20:93–9. [PubMed: 14693814]
29. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545–50. [PubMed: 16199517]
30. Hazelett DJ, Rhie SK, Gaddis M, et al. Comprehensive functional annotation of 77 prostate cancer risk loci. PLoS Genet. 2014; 10:e1004102. [PubMed: 24497837]
31. Innocenti F, Cooper GM, Stanaway IB, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet. 2011; 7:e1002078. [PubMed: 21637794]

32. Pelton K, Freeman MR, Solomon KR. Cholesterol and prostate cancer. Curr Opin Pharmacol. 2012; 12:751–9. [PubMed: 22824430]

33. Zadra G, Photopoulos C, Loda M. The fat side of prostate cancer. Biochim Biophys Acta. 2013; 1831:1518–32. [PubMed: 23562839]

34. Breyer JP, Dorset DC, Clark TA, et al. An Expressed Retrogene of the Master Embryonic Stem Cell Gene POU5F1 Is Associated with Prostate Cancer Susceptibility. Am J Hum Genet. 2014; 94:395–404. [PubMed: 24581739]

Table 1

Clinical characteristics of men with prostate cancer in the Physicians' Health Study and Health Professionals Follow-up Study

	TUMOR (n=264)	NORMAL (n=160)
Lethal, n (%)	41 (15.5)	24 (15.0)
Gleason score*, n (%)		
5–6	39 (14.8)	28 (17.5)
7	167 (63.3)	100 (62.5)
8–10	58 (22.0)	32 (20.0)
Pathologic stage, n (%)**		
T2	164 (62.1)	107 (66.9)
T3	81 (30.7)	43 (26.9)
T4/N1/M1	17 (6.4)	9 (5.6)
missing	2 (0.8)	1 (0.6)
PSA at diagnosis, n (%)		
0–4	28 (10.6)	18 (11.3)
4–10	136 (51.5)	83 (51.9)
10–20	49 (18.6)	28 (17.5)
>20	24 (9.1)	15 (9.4)
Pre-PSA era (before 1992)	27 (10.2)	16 (10.0)

* Gleason score is from radical prostatectomy for 256 tumor and 154 normal, and from TURP for the remaining cases

** RP pathologic stage for most, but clinical stage at diagnosis for TURP cases

Table 2

Significant (p<0.05) associations of risk SNPs with the expression of nearby (±500 kb) genes in tumor or normal prostate tissue

SNP	chromosome	position	gene	distance from SNP (kb)	p (tumor)*	p (normal)*	p (tumor - restricted)**
rs13385191	2p24	20888264	C2orf43	0	0.029 (lower)	0.016 (lower)	0.029
rs12621278	2q31	173311552	DLX2	−344.1	0.035 (higher)	0.439	0.006
			ITGA6	0	0.022 (higher)	0.295	0.111
rs2292884	2q37	238443225	MLPH	0	0.020 (lower)	0.051 (lower)	0.009
			LRRFIP1	93	0.475	0.030 (higher)	0.537
rs7629490	3p11	87241496	CHMP2B	34.9	0.01 (higher)	0.958	0.008
rs17021918	4q22	95562876	BMPR1B	116.3	0.034 (lower)	0.056	0.012
rs12653946	5p15	1895828	IRX4	−8.5	0.001 (lower)	0.019 (lower)	0.009
rs1983891	6p21	41536426	TREM2	−405.5	0.738	0.012 (higher)	0.750
			MED20	336.7	0.016 (higher)	0.636	0.046
rs339331	6q22	117210051	RSPH4A	−255.9	0.029 (lower)	0.976	0.029
rs9364554	6q25	160833663	IGF2R	−306.1	0.048 (lower)	0.501	0.041
rs10486567	7p15	27976562	HIBADH	−273.9	0.028 (lower)	0.130	0.172
			TAX1BP1	−107.2	0.443	0.037 (lower)	0.933
rs6465657	7q21	97816326	LMTK2	0	0.021 (higher)	0.607	0.132
rs4242382	8q24	128517572	POU5F1B	−88.1	0.010 (higher)	0.083	0.048
rs1571801	9q33	124427372	RAB14	−463	0.510	0.023 (lower)	0.704
			TTLL11	156.8	0.889	0.046 (higher)	0.744
			NDUFA8	479	0.407	0.045 (lower)	0.088
rs10993994	10q11	51549495	MSMB	0.1	3.94E-07 (lower)	0.002 (lower)	4.29E-05
			NCOA4	15.6	0.640	5.70E-05 (higher)	0.649
			ASAH2	397.5	0.046 (lower)	0.205	0.398

SNP	chromosome	position	gene	distance from SNP (kb)	p (tumor)*	p (normal)*	p (tumor - restricted)**
rs7127900	11p15	2233573	ASCL2	56.1	2.84E-04 (higher)	0.793	0.003
			KCNQ1	232.6	0.025 (higher)	0.559	0.225
rs12418451	11q13	68935418	IGHMBP2	−227.3	0.779	0.006 (higher)	0.849
rs10896449	11q13	68994666	MTL5	−475.7	0.026 (lower)	0.037 (lower)	0.116
rs902774	12q13	53273903	KRT6B	−428	0.034 (higher)	0.744	0.124
			KRT79	−45.8	0.780	0.019 (lower)	0.641
			ESPL1	388.2	0.473	0.043 (lower)	0.763
			SP7	446.5	0.017 (higher)	0.027 (lower)	0.145
rs8102476	19q13	38735612	PPP1R14A	6.3	0.004 (higher)	0.038 (higher)	0.014
			SPINT2	19.5	0.012 (lower)	0.344	0.014
			GGN	139.3	0.237	0.012 (higher)	0.309
rs11672691	19q13	41985586	CYP2F1	−351.3	0.177	0.040 (higher)	0.560
			AXL	−217.9	0.004 (lower)	0.915	0.004
			DMRTC2	363.4	0.044 (lower)	0.565	0.023
rs2735839	19q13	51364622	C19orf63	−378	0.016 (higher)	0.386	0.069
rs5945619	Xp11	51241671	NUDT10	−161.3	0.040 (lower)	0.331	0.076
			NUDT11	−2.2	4.60E-11 (higher)	3.98E-04 (higher)	3.24E-09

* lower or higher in parentheses indicates value in those carrying 2 copies of the risk allele compared to those carrying 0 copies of the risk allele for significant ($p<0.05$) associations

** tumor restricted refers to using only the tumors that have paired normal samples (N=159)

Table 3

Pathways significantly associated with risk SNPs in tumor tissue

	PATHWAY	p
rs1512268	KEGG_VEGF_SIGNALING_PATHWAY	5.27E-05
	KEGG_LINOLEIC_ACID_METABOLISM	5.42E-05
	KEGG_ALPHA_LINOLENIC_ACID_METABOLISM	6.09E-05
	KEGG_ARACHIDONIC_ACID_METABOLISM	6.42E-05
	KEGG_ETHER_LIPID_METABOLISM	7.71E-05
	KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY	1.13E-04
	KEGG_LONG_TERM_DEPRESSION	1.40E-04
	KEGG_GNRH_SIGNALING_PATHWAY	1.59E-04
	KEGG_GLYCEROPHOSPHOLIPID_METABOLISM	1.72E-04