Cancer Informatics

ORIGINAL RESEARCH

Comprehensive Assessment and Network Analysis of the Emerging Genetic Susceptibility Landscape of Prostate Cancer

Chindo Hicks[1–3], Lucio Miele[1], Tejaswi Koganti and Srinivasan Vijayakumar[3]

[1]Cancer Institute, University of Mississippi Medical Center, Jackson, MS. [2]Department of Medicine, University of Mississippi Medical Center, Jackson, MS. [3]Department of Radiation Oncology, University of Mississippi Medical Center, Jackson, MS. Corresponding author email: chicks2@umc.edu

Abstract

Background: Recent advances in high-throughput genotyping have made possible identification of genetic variants associated with increased risk of developing prostate cancer using genome-wide associations studies (GWAS). However, the broader context in which the identified genetic variants operate is poorly understood. Here we present a comprehensive assessment, network, and pathway analysis of the emerging genetic susceptibility landscape of prostate cancer.

Methods: We created a comprehensive catalog of genetic variants and associated genes by mining published reports and accompanying websites hosting supplementary data on GWAS. We then performed network and pathway analysis using single nucleotide polymorphism (SNP)-containing genes to identify gene regulatory networks and pathways enriched for genetic variants.

Results: We identified multiple gene networks and pathways enriched for genetic variants including *IGF-1*, androgen biosynthesis and androgen signaling pathways, and the molecular mechanisms of cancer. The results provide putative functional bridges between GWAS findings and gene regulatory networks and biological pathways.

Keywords: prostate cancer GWAS network pathway analysis

Cancer Informatics 2013:12 175–191

doi: 10.4137/CIN.S12128

This article is available from http://www.la-press.com.

Introduction

Prostate cancer is the second most common cause of cancer-related death in men in the United States, accounting for more than 200,000 new cases and 32,000 deaths annually.[1] Recent advances in high-throughput genotyping technologies and reduction in genotyping costs have made possible identification of single nucleotide polymorphisms (SNPs) (herein called genetic variants) associated with an increased risk of developing prostate cancer using genome-wide association studies (GWAS).[2]

GWAS have been very successful in identifying disease loci using single-marker–based association tests that examine the relationship between each SNP marker and the trait of interest in prostate cancer. The ultimate aim of this approach is the identification of genes that are causal for prostate cancer. However, with growth in epidemiological evidence of associations has come an increasing need to collate and summarize the evidence in order to identify credible genetic associations among the large amounts of information. In addition, there is an urgent need to direct focus to functional characterization of the identified genetic variants and identification of aberrant pathways enriched for genetic variants to understand the broader context in which they operate.

These findings are providing valuable clues about the genetic basis of prostate cancer. However, the genetic variants and associated genes reported thus far explain only a small proportion of the phenotypic variation, limiting the potential for early application to predict disease risk. Relatively few SNPs have *P* values sufficiently small to give conclusive evidence of association. In addition, many of the identified SNPs have not been replicated in multiple independent studies. Conversely, there are usually many SNPs with small to moderate associations. However, it is not clear from the published reports whether genes containing genetic variants with large effects and those containing genetic variants with small to moderate effects are functionally related or interact with each other.

Prostate cancer originates from a more complex interplay between a constellation of changes in DNA involving many genes and a broad range of environmental factors. These complex arrays of interacting factors affect entire network states that in turn increase or decrease the risk of developing prostate cancer or affect the disease severity.[3] Therefore, although each single SNP may confer only a small disease risk, their joint actions are likely to have a significant role in the development of prostate cancer. If all effort is directed at identifying only the most significant SNPs, the genetic variants that jointly have significant risk effects, but individually making only a small contribution, could be missed.[4] While some of the genetic variants with small to moderate effects identified using GWAS may be false positives, there are likely many others that contain genuine effects of small to moderate magnitude. The presence of associated SNPs in functionally related genes interacting in gene regulatory networks and biological pathways gives a degree of confidence that associations are potentially genuine even if none of the SNPs individually is highly significant.[5] In addition, the actions of SNP-containing genes may be mediated by other genes not identified by GWAS. Such genes if identified and confirmed could explain the missing variation.

Evidence of association is continuously evolving and much work remains to obtain a complete inventory of the variants at each locus that contribute to prostate cancer risk and to define the molecular mechanisms through which genetic variants operate. However, with the completion of the initial GWAS in prostate cancer, it is timely to evaluate the evidence and the credibility of genetic variants identified thus far to identify potential key drivers of prostate cancer. This undertaking requires evaluation of all available GWAS with evidence of association based on widely accepted criteria for assessment of the cumulative evidence while taking into account the broader context in which the identified genetic variants operate. The guidelines for assessing evidence and credibility of associations, which have become known as the Venice criteria, have been set forth by the Human Genome Epidemiology Networking Group.[6–10] Our group[11] and others[12] recently applied these guidelines to assess association of genetic variants with breast cancer.

In the published literature on GWAS, a catalogue of genetic variants associated with increased risk of developing common human diseases including cancer has been created.[2] But this catalogue as demonstrated in the results section of this study is incomplete and reports only the most highly statistically significant

($P < 10^{-8}$) genetic variants.[2] The emerging picture from large-scale genomic investigations is that prostate cancer originates from more complex interactions between constellations of genes and changes in DNA involving both common and rare variants.[13,14] Integrating GWAS information with biological knowledge using functional network and pathway analysis holds the promise for defining the molecular networks and biological pathways that are involved in susceptibility to prostate cancer and provides insights about the broader context in which genetic variants operate. This study was undertaken in view of the plethora of data from GWAS on prostate cancer, the necessity to assess evidence and the credibility of genetic variants of the emerging genetic susceptibility landscape of prostate cancer, and to understand the broader context in which the identified genetic variants operate. We hypothesized that genetic variants associated with increased risk of developing prostate cancer map to functionally related genes which interact with each other in gene regulatory networks and biological pathways enriched for genetic variants that in turn increase or decrease the risk of developing prostate cancer.

Material and Methods

Our first step in this report was to obtain a complete inventory of genetic variants associated with increased risk of developing prostate cancer reported thus far and to assess the epidemiological evidence and evaluate the credibility of the identified genetic variants. Our methods for GWAS data collection were based on the guidelines proposed by the Human Genome Epidemiology Network for systematic review of genetic association studies.[6–10]

We mined SNP data and gene information from the published reports on GWAS and accompanying websites providing supplementary data for prostate cancer. By far, the most frequently used GWAS design to date has been the case-control design, in which allele frequencies in patients with prostate cancer are compared with those in a disease-free comparison group.[15] In this study we adopted this design for screening published GWAS reports and extracting the data. GWAS were eligible to be included if they met the following criteria. First, publications must have been in peer-reviewed journals or online and published in English on or before April 2013. Second, prostate cancer must have been diagnosed by histological examination. Studies were eligible if they were based on unrelated individuals, examined the association between prostate cancer and the polymorphic phenotype, and had a sample size of greater than 500 cases and greater than 500 controls. Only studies published as full-length articles or letters in peer-reviewed journals in English were included in the analysis. The studies must have provided sufficient information such that genotype frequencies for both prostate cancer cases and controls could be determined without ambiguity.

To identify all relevant publications, we used 2 search strategies. First, we queried PubMed with the terms GWAS, GWA, WGAS, WGA, genome-wide, genomewide, whole genome, and all terms plus association or scan in combination with prostate cancer to find all the genome-wide studies published before April 2013. This search yielded 150 publications, which were screened by title, abstract, and full text review to identify studies that met our eligibility criteria. After screening, 100 studies met our eligibility criteria. The exclusion criteria for the 50 studies included studies with insufficient or incomplete information, reviews, studies reporting only intergenic regions, and studies with very small sample sizes. The data were manually extracted from reported GWAS that met our eligibility criteria and websites accompanying those studies providing supplementary data. When a study included multiple ethnic populations we picked the results of the model that adjusted for ethnicity, otherwise each population was considered separately as presented by the originators of the data.

Evidence and credibility of association were assessed using the procedure described by Ioannidis et al,[6] which included the amount of evidence, extent of replication, protection from bias, and a composite assessment of strong, moderate, or weak epidemiological credibility. The search yielded 250 SNPs mapped to 162 protein coding genes derived from a population of over 350,000 cases and over 350,000 controls. In addition, we identified 200 SNPs mapped to intergenic regions, but these were not included in subsequent analysis so these are not presented here. To address publication bias, we catalogued all the available SNPs that showed significant ($P < 0.05$) association with increased risk of developing prostate cancer. Publication bias occurs when only the most

significant SNPs are reported, which introduces a bias toward the most significant SNPs while ignoring those SNPs that do not reach the threshold predetermine by the investigator, the so called winner curse. We reasoned that if we considered only the most significant SNPs, the genetic variants and associated genes that jointly have significant risk effects but individually make only a small contribution would be missed. The SNP, IDs (rs-ID) (i.e. SNP identification number), locations, and gene names were verified using the dbSNP database (http://www.ncbi.nlm.nih.gov/snp/) using chromosome report build 37.7 and the Human Genome Nomenclature (HGNC) database (http://www.genenames.org/). SNPs were matched with gene names using SNP IDs (rs-IDs) information in the database (dbSNP). For SNPs replicated in multiple independent studies, we combined the *P* values to estimate the overall effect size using Fisher's methods as described in our previous study.[5]

Data analysis

Data analysis was conducted to accomplish 2 primary objectives. The first objective was to assess the evidence and credibility of the reported associations from GWAS. The second was to determine whether genes containing SNPs associated with increased risk of developing prostate cancer are functionally related and interact with one another in gene regulatory networks and biological pathways enriched for SNPs. The overarching goal of this analysis was to understand the broader context in which the identified genetic variants operate and to identify the biological mechanisms underlying GWAS findings.

Evidence and credibility of associations were assessed at 3 levels defined as strong, moderate, and weak evidence. Strong evidence was defined as meeting the stringent statistical threshold of $P < 10^{-8}$. This statistical threshold was chosen on the basis of a general consensus for accepting genome-wide association studies as recorded by the National Human Genome research database hosting acceptable associations.[2] Moderate evidence was defined as meeting the statistical significance of $P \sim 10^{-5}$–10^{-7}. We defined weak association as meeting the statistical significance of $P \sim 10^{-2}$–10^{-4}. These statistical threshold levels were chosen to address the publication bias also known as "winner curse" because, as evidenced in this study (see the results section and supplementary Table 1), only a very small proportion of SNPs reach statistically unimpeachable threshold levels ($P < 10^{-8}$) and these explain only a small amount of the variation. The rationale was that if we considered only the most significant SNPs, the genetic variants and associated genes that jointly have significant risk effects but individually make only a small contribution or have low penetrance would be missed. For assessing replication, we relaxed the threshold levels focusing on the range rather than one statistical threshold level. This liberal approach was adopted for several reasons. First, locus heterogeneity, which implies that alleles at different loci cause prostate cancer in different and admixed populations will increase difficult in replication of association of a single marker.[16] Second, sampling errors resulting from differences in sample sizes could affect the results of replication. Third, the statistical model, for example, a model that does not account for population structure or the admixing of the populations could affect the replication results.[16]

To determine whether the SNP-containing genes are functionally related we used gene ontology (GO) analysis.[17] The GO Consortium has developed 3 separate categories to describe the attributes of gene products: molecular function, biological process, and cellular component. Molecular function defines what a gene product does at the biochemical level without specifying where or when the event actually occurs or its broader context, biological process describes the contribution of the gene product to the biological objective, and cellular component refers to where in the cell a gene product functions. Because our goal in this study was to gain biological insights about the broader context in which genetic variants associated with increased risk of developing prostate cancer operate, we considered all 3 GO categories.

To investigate the broader context in which the genetic variants operate and to identify the molecular mechanisms underlying GWAS findings, we used network and pathway analysis and visualization using the Ingenuity Pathway Analysis (IPA) System (http://www.ingenuity.com).[18] The goal was to identify gene regulatory networks and biological pathways that are enriched for genetic variants associated with increased risk of developing prostate cancer. We hypothesized that genes containing SNPs associated with increased risk for developing prostate cancer interact with each other and other genes within biological

pathways enriched for genetic variants. Gene symbols of SNP-containing genes were mapped onto networks and pathways using IPA. The networks and pathways were ranked by score and *P* values, respectively. The score indicated the likelihood of the genes in a network being found together by random chance. Using a 99% confidence interval, scores of ≥3 were considered significant. The *P* values indicated the significant level for correctly assigning a particular genes or sets of genes to the canonical pathway. Additional information, validation of predicted pathways, and identification of other downstream target genes was achieved through the literature and database mining module built in the Ingenuity System, which allowed identification of other functionally related genes not identified by GWAS. The distribution of the overall effect of SNPs in the pathway and replicated SNPs were calculated using the procedure we have previously reported.[5] Genes showing indirect (spurious) interactions were pruned from the networks to ensure the reliability of the identified gene regulatory networks and biological pathways enriched for SNPs.

Results

Evidence and credibility of associations

We have developed a comprehensive catalogue of SNPs and genes associated with increased risk of developing prostate cancer. Our analysis revealed 250 SNPs mapped to 162 protein coding genes. Out of the total genetic variants identified, 62 SNPs mapped to 41 genes and have strong association with prostate cancer ($P < 10^{-8}$) (Table 1). Among the genes containing SNPs with strong associations, 10 genes, including *EEFSEC, HNF1B, JAZF1, KLK3, MSMB, NUDT11, PDLIM5, POU5F1, RFX6, and TERT,* contain multiple genetic variants with strong associations (Table 1). A full catalogue of all the 250 SNPs and the 162 associated genes along with information on the 100 peer-reviewed references from which data were extracted are presented in Supplementary Table 1 provided as supplementary data to this report.

In GWAS analysis, replication of findings in independent data sets is now widely regarded as a prerequisite for convincing evidence of association. Therefore, we used this criterion to assess the credibility of the associations we catalogued in this study. This evaluation revealed 52 SNPs mapped to 40 protein coding genes, including *BIK, BMP5, C2ORF43, CASP3, CNGB3, CTBP2, DAP21P, EEFSEC, EHBP1, FGF10, FOXP4, FREM1, RFX6, HERC2, HNF1B, ITGA6, JAZF1, KLK15, KLK3, KLK5, LMTK2, LOC729852, LOC727677, MLPH, MSMB, MSR1, NCOA4, NKX3-1, NSMCE2, NUD11, PDLIM5, SLC22A3, SLC25A37, TERT, TET2, THADA, TNFSF10, TNRC6B, ZBTB38*, and *ZNF652*, which have been replicated in multiple independent studies (Table 2). The *P* values for the SNPs replicated in multiple independent studies varied markedly ranging from $P = 10^{-47}$ to $P = 0.05$ (Table 2). The number of replications also varied markedly ranging from as low as 2 studies to as many as 36 studies (Table 2). Replication tended to be biased toward SNPs with strong statistical evidence of association in the initial studies (Table 2). Only 7 genes, including *EHBP1, DAP2IP, HNF1B, KLK3, MSMB, NUDT11*, and *PDLIM5*, contained multiple SNPs replicated in multiple independent studies (Table 2). A complete assessment of the SNPs with strong association and SNPs replicated in multiple independent studies revealed 38 SNPs mapped to 17 genes, which have strong associations but have not been replicated in independent studies (Tables 1 and 2). Interestingly, 15 genetic variants with weak to moderate associations have been replicated in multiple independent studies (Table 2). Overall, the majority of the genes (100 genes) contained SNPs with small ($P \sim 10^{-2}$–10^{-4}) to moderate effects ($P \sim 10^{-5}$–10^{-7}) and have not been replicated in multiple independent studies (Supplemental Table 1).

It is difficulty from this study to determine whether the observed between-study heterogeneity and lack of replication in associations reflect genuine functional diversity of the identified loci. This is because the threshold of replication is a matter of considerable debate and many factors can cause lack of replication and between study heterogeneity.[19–21] The reader is referred to an excellent publication on assessment of cumulative evidence on genetic associations[6] and a study on replicating genotype-phenotype associations.[20] In brief, lack of replication or between-study heterogeneity may signal underlying errors and biases, including genotyping errors, phenotypic misclassification, population stratification, and selective reporting biases.[22–25] However, lack of replication observed here for some of the SNPs may represent genuine findings of small magnitude when substan-

Table 1. Genetic variants and genes significantly associates with increased risk of developing prostate cancer (SNP, $P < 10^{-8}$).

Gene name	Chromosome position	SNP ID	SNP *P*-value range
AR	Xq12	rs5919432	1.00E^{-08}
ARL15	5p15.2	rs792017	5.4×10^{-19}
BIK	22q13.31	rs742134	5.6×10^{-9}
C2ORF43	2p24.1	rs13385191	7.5×10^{-8}
CCHCR1	6p21.3	rs130067	3.2×10^{-8}
COL6A3	2q37	rs7584330	3.00E^{-09}
CTBP2	10q26.13	rs4962416	2.7×10^{-8}
CXorf67	Xp11.22	rs1327301	2.00E^{-10}
DPF1	19q13.2	rs8102476	2.00E^{-11}
EEFSEC	3q21.3	rs10934853	3.00E^{-10}
EEFSEC	3q21.3	rs4857841	2.32×10^{-8}
EHBP1	2p15	rs721048	7.7×10^{-9}
FAM84B	8q24.21	rs1016343	4.00E^{-10}
FGF10	5p13-p12	rs2121875	4.0×10^{-8}
FOXP4	6p21.1	rs1983891	7.6×10^{-8}
FSHR	2p21-p16	rs2268363	5.00E^{-8}
GGCX	2p12	rs10187424	3.00E^{-15}
GPRC6A/RFX6	6q22.31	rs339331	1.6×10^{-12}
HNF1B	17q12	rs4430796	1.13×10^{-25}
HNF1B	17q12	rs11649743	1.19×10^{-9}
HNF1B	17q12	rs3744763	1.21×10^{-08}
HNF1B	17q12	rs757210	1.39×10^{-15}
HNF1B	17q12	rs4239217	1.57×10^{-16}
HNF1B	17q12	rs2005705	2.54×10^{-23}
HNF1B	17q12	rs3760511	4.45×10^{-15}
HNF1B	17q12	rs4794758	4.95×10^{-10}
HNF1B	17q12	rs7405696	9.35×10^{-23}
IL16	15q26.3	rs7175701	9.8×10^{-8}
IRX4	5p15.33	rs12653946	3.9×10^{-18}
ITGA6	2q31.1	rs12621278	3.36×10^{-19}
JAZF1	7p15.2-p15.1	rs1080784	2.96×10^{-10}
JAZF1	7p15.2-p15.1	rs10486567	7.05×10^{-14}
KLK3	19q13.41	rs902774	5.00E^{-09}
KLK3	19q13.41	rs6465657	2.00E^{-08}
KLK3	19q13.41	rs17632542	1.6×10^{-24}
KLK3	19q13.41	rs2735839	2.4×10^{-20}
KLK3	19q13.41	rs1058205	2.8×10^{-23}
KRT78	12q13.13	rs651164	2.00E^{-10}
LMTK2	7q21.3	rs2292884	4.00E^{-08}
LOC727677	8q24.21	rs1447295	2.2×10^{-19}
MSMB	10q11.2	rs0993994	8.7×10^{-29}
MSMB	10q11.2	rs7075697	1.46×10^{-9}
MSMB	10q11.2	rs792057	7.2×10^{-13}
MYEOV	11q13.2	rs10896449	8.30×10^{-10}
NUDT11	Xp11.22-p11.1	rs5945619	1.00×10^{-47}
NUDT11	Xp11.22-p11.1	rs5945572	6.17×10^{-11}
PDLIM5	4q22	rs12500426	1.3×10^{-11}
PDLIM5	4q22	rs17021918	4.2×10^{-15}
POU5F1	6p21.31	rs7837688	1.00E^{-25}
POU5F1	6p21.31	rs4242382	3.00E^{-19}
POU5F1	6p21.31	rs4242384	3.00E^{-16}

(*Continued*)

Table 1. (*Continued*)

Gene name	Chromosome position	SNP ID	SNP *P*-value range
RFX6	6q22.31	rs12202378	8.8×10^{-8}
SKIL	3q26	rs10936632	$7.00E^{-22}$
SLC22A3	6q25.3	rs9364554	$6.00E^{-10}$
SLC25A37	8p21.2	rs10503733	$8.00E^{-08}$
SQRDL	15q15	rs4775302	$4.00E^{-08}$
TERT	5p15.33	rs2242652	2.7×10^{-24}
TERT	5p15.33	rs2736098	3×10^{-10}
TET2	4q24	rs7679673	6.74×10^{-10}
THADA	2p21	rs1465618	$2.00E^{-08}$
ZBTB38	3q23	rs6763931	$2.00E^{-08}$
ZNF652	17q21.32	rs7210100	$3.00E^{-13}$

tive differences between the discovery and replication studies exist.[6] In addition, lack of replication may also be attributable to different linkage disequilibrium patterns across different populations, population-specific gene-gene epistasis, and/or gene-by-environment interactions.[6] Under these conditions, the heterogeneity in the discovered loci may reflect genuine functional diversity. The small to moderate effect sizes observed in this study may be attributed to several factors including sampling errors resulting from limited or small samples sizes, models used in data analysis, population structure, as well as genuine effects with small magnitude.[19–25] Selection might also be responsible for keeping genetic effect sizes low, as variants of large effect may be selected against and eventually disappear.[6] This follows from the fact that long-term stabilization selection minimizes the production of individuals at the extremes of a trait[6] in part by reducing the additive genetic effects of alleles already present or those arising de novo by mutations[6] to levels beneath the ability of GWAS of reasonable size to detect them. In general, from the perspective of gaining insights into prostate cancer pathogenesis, an effect, regardless of how small, may provide useful information when considered with others.[6]

This study was conducted in an attempt to make a complete inventory and to catalogue all reported GWAS associations with prostate cancer. Over the last several years, the National Institute of Genome Research (NIGR) has created a database that documents SNPs with strong associations,[2] but this database relies on self-reported studies and is incomplete. To determine whether our study provides any new complementary information, we compared the results in this report with those reported in the NIGR database. To make a fair comparison, we first focused on SNPs with strong statistical evidence of association, since that is the metric used by the NIGR. As of this writing, the NIGR database contained 33 genetic variants with $P < 10^{-6}$ mapped to 31 protein coding genes. Using the same statistical threshold, our investigation revealed 217 SNPs mapped to 131 protein coding genes that are not documented in the NIGR database. The disparity between findings reported in our study and those reported in the database can be explained by the fact that our study was based on manual curation and extraction of information, while the NGRI results may be based on self-reporting. Importantly, our analysis focused on evidence and credibility of association as measured by the level of association and reproducibility of that association rather than focusing on the statistical strength of that association. Under such conditions, the observed outcome between this report and the NGRI database should be expected. It is worth noting that our study does not refute the information in the NIGR database but provides complementary and additional information.

Functional relationships of identified SNP-containing genes

To determine whether genes containing SNPs associated with increased risk of developing prostate cancer are functionally related, we performed GO analysis as described in the Methods section. The underlying hypothesis of this investigation is that genetic variants associated with increased risk of developing prostate cancer map to genes that are functionally related and involved in similar biological and cellular processes.

Table 2. Genetic variants and genes associated with increased risk of developing prostate that have been replicated in multiple independent studies.

Gene name	Chromosome position	SNP ID	Number of repetitions	SNP *P*-value
BIK	22q13.31	rs5759167	3	1.30×10^{-12}–3.01×10^{-3}
BMP5	6p12.1	rs3734444	3	3.0×10^{-2}–4.0×10^{-2}
C2ORF43	2p24.1	rs13385191	9	7.5×10^{-8}–1.0×10^{-2}
CASP3	4q34	rs4862396	2	4.0×10^{-2}
CNGB3	8q21.3	rs4961199	2	2.79×10^{-2}–1.0×10^{-2}
CTBP2	10q26.13	rs4962416	10	2.7×10^{-8}–3.0×10^{-3}
DAP2IP	9q33.1-q33.3	rs1571801	3	2.84×10^{-5}–3.0×10^{-3}
DAP2IP	9p21	rs1571801	4	2.84×10^{-5}–3.0×10^{-2}
EEFSEC	3q21.3	rs4857841	4	2.3×10^{-8}–3.30×10^{-3}
EHBP1	2p15	rs721048	8	7.7×10^{-9}–4.0×10^{-2}
EHBP1	2p15	rs2710646	2	2.5×10^{-3}–2.0×10^{-3}
FGF10	5p13-p12	rs2121875	3	4.0×10^{-8}
FOXP4	6p21.1	rs1983891	9	7.6×10^{-8}–2.0×10^{-2}
FREM1	9p22.3	rs1552895	2	2.0×10^{-3}
GPRC6A/RFX6	6q22.31	rs339331	7	1.6×10^{-12}–2.0×10^{-3}
HERC2	15q13	rs6497287	2	5.20×10^{-5}–4.0×10^{-3}
HNF1B	17q12	rs4430796	27	1.13×10^{-25}–1.0×10^{-2}
HNF1B	17q12	rs3760511	2	4.45×10^{-15}–8.8×10^{-4}
HNF1B	17q12	rs11649743	8	1.2×10^{-9}–2.58×10^{-2}
HNF1B	17q12	rs7501939	7	3×10^{-18}–1.0×10^{-3}
ITGA6	2q31.1	rs12621278	7	9×10^{-23}–2.0×10^{-2}
JAZF1	7p15.2-p15.1	rs10486567	23	7.05×10^{-14}–4.0×10^{-2}
KLK15	19q13.4	rs2659056	6	2.7×10^{-4}–1.0×10^{-2}
KLK3	19q13.41	rs17632542	2	1.6×10^{-24}–3×10^{-10}
KLK3	19q13.41	rs1058205	3	2.8×10^{-23}–4.0×10^{-2}
KLK3	19q13.41	rs266849	3	1.4×10^{-14}
KLK5	19q13.33	rs268908	2	0.0001–1.0×10^{-3}
LMTK2	7q22.1	rs6465657	10	1.1×10^{-9}–2.0×10^{-2}
LOC727677	8q24.21	rs1447295	18	2.2×10^{-19}–1.0×10^{-2}
LOC729852	7p21.3	rs2348763	2	1.0×10^{-2}
MLPH	2q37.2	rs2292884	3	4.00×10^{-08}
MSMB	10q11.2	rs0993994	36	8.7×10^{-29}–1.0×10^{-2}
MSMB	10q11.2	rs7920517	3	1.0×10^{-3}–1.0×10^{-2}
MSR1	8p22	rs351572	2	9.0×10^{-3}–2.0×10^{-2}
NCOA4	10q11.2	rs7350420	2	5.6×10^{-3}–7.0×10^{-3}
NKX3-1	8p21.2	rs1512268	4	5.52×10^{-7}–7.0×10^{-3}
NSMCE2	8q24.13	rs7008482	4	5.0×10^{-4}–4.0×10^{-2}
NUDT11	Xp11.22-p11.1	rs5945619	9	1.00×10^{-47}–4.10^{-2}
NUDT11	Xp11.22-p11.1	rs5945572	8	6.17×10^{-11}–5.0×10^{-2}
PDLIM5	4q22	rs17021918	5	4.2×10^{-15}–7.30×10^{-2}
PDLIM5	4q22	rs12500426	4	1.3×10^{-11}–2.0×10^{-3}
PDLIM5	4q22	rs17021918	7	4.0×10^{-15}–3.3×10^{-5}
RFX6	6q22.31	rs339331	4	3.1×10^{-6}–4.43×10^{-5}
SLC22A3	6q25.3	rs9364554	12	9.3×10^{-7}–2.0×10^{-3}
SLC25A37	8p21.2	rs2928679	2	2.64×10^{-1}–3.0×10^{-2}
TERT	5p15.33	rs2242652	3	2.7×10^{-24}
TET2	4q24	rs7679673	3	1.2×10^{-2}–6.74×10^{-10}
THADA	2p21	rs1465618	7	1.6×10^{-8}–2.0×10^{-2}
TNFSF10	3q26	rs3774315	2	7.34×10^{-5}–2.0×10^{-3}
TNRC6B	22q13	rs9623117	2	5×10^{-7}–1.22×10^{-3}
ZBTB38	3q23	rs6763931	2	2.0×10^{-8}
ZNF652	17q21.32	rs7210100	2	3.4×10^{-13}

The rationale is that the presence of SNPs in genes of similar biological functions and involved in the same biological processes and cellular components gives a degree of confidence that the associations could potentially be genuine even if none of the SNPs individually has a very strong statistical association or is not replicated in multiple independent studies.

GO analysis revealed that the genes containing SNPs associated with increased risk of developing prostate cancer are functionally related and are involved in similar biological and cellular processes. A comprehensive list of 155 SNP-containing genes, the molecular functions, and the biological and cellular processes in which they are involved is presented in Supplementary Table 2, provided as supplementary data to this report. The difference between 162 protein coding genes mentioned earlier in this report and the 155 reported in Supplementary Table 2 is due to lack of probe annotation for the 7 genes. Interestingly, genes containing SNPs with large effects and SNPs replicated in multiple independent studies were found to be functionally related with genes containing SNPs with small to moderate effects. This is a significant finding given that relatively fewer genes contain SNPs with large effects and SNPs replicated in multiple independent studies. Although the results of GO analysis cannot explain how genetic variants regulate gene function, they provide insights into molecular functions, the biological processes, and cellular components in which the genetic variants and genes associated with increased of developing prostate cancer are involved.

Network and pathway analysis

To gain insights about the broader context in which genetic variants operate, we performed network and pathway analysis and visualization using the Ingenuity IPA System as described in the Methods section. Our working hypothesis was that genes containing genetic variants with strong and moderate to weak associations interact with one another and with their downstream targets in gene regulatory networks and biological pathways enriched for SNPs. Three specific objectives were of interest in network and pathway analysis: (1) to identify gene regulatory networks and biological pathways enriched for genetic variants, (2) to determine whether genes containing genetic variants with strong associations and genetic variants replicated in multiple independent studies interact with genes containing genetic variants with weak to moderate associations, and (3) to identify novel genes not identified by GWAS. Our initial network analysis produced 9 networks with similar but overlapping functions with scores ranging from 16 to 38. We merged the first 5 networks and the last 4 networks into 2 separate but consolidated networks using the merge and build modules as implemented in the Ingenuity IPA System. The results of network analysis are presented in Figures 1 and 2. In the networks, nodes represent genes and edges represent interactions and functional relationships. Genes containing SNPs with strong associations and SNPs replicated in multiple independent studies are marked in red font whereas the genes containing SNPs with small to moderate effects are marked in blue font to distinguish the 2 sets of genetic variants and associated genes. Also presented in the networks in black font are the novel genes which have not been reported in GWAS.

Network analysis revealed complex gene regulatory networks enriched for SNPs associated with increased risk of developing prostate cancer, confirming our hypothesis. Genes containing SNPs with strong statistical associations and genes containing SNPs replicated in multiple independent studies were found to interact with each other (Figs. 1 and 2). However, most of the interactions were through other genes suggesting that the actions of genes containing genetic variants with strong associations and genetic variants replicated in multiple independent studies may be mediated through other genes. Interestingly, genes containing genetic variants with strong associations and genetic variants replicated in multiple independent studies were found to interact with genes containing SNPs with weak to moderate associations (Figs. 1 and 2). Even more intriguing was the discovery that genes containing genetic variants with strong associations and genetic variants replicated in multiple independent studies were found to interact with genes not identified by GWAS. These are very significant findings in that they demonstrate that genes that contain genetic variants with small to moderate effects and genes that do not harbor genetic variants reported in GWAS, but that could potentially play key roles in parts of the biological networks by mediating or regulating SNP-containing genes, may be systematically missed by focusing on

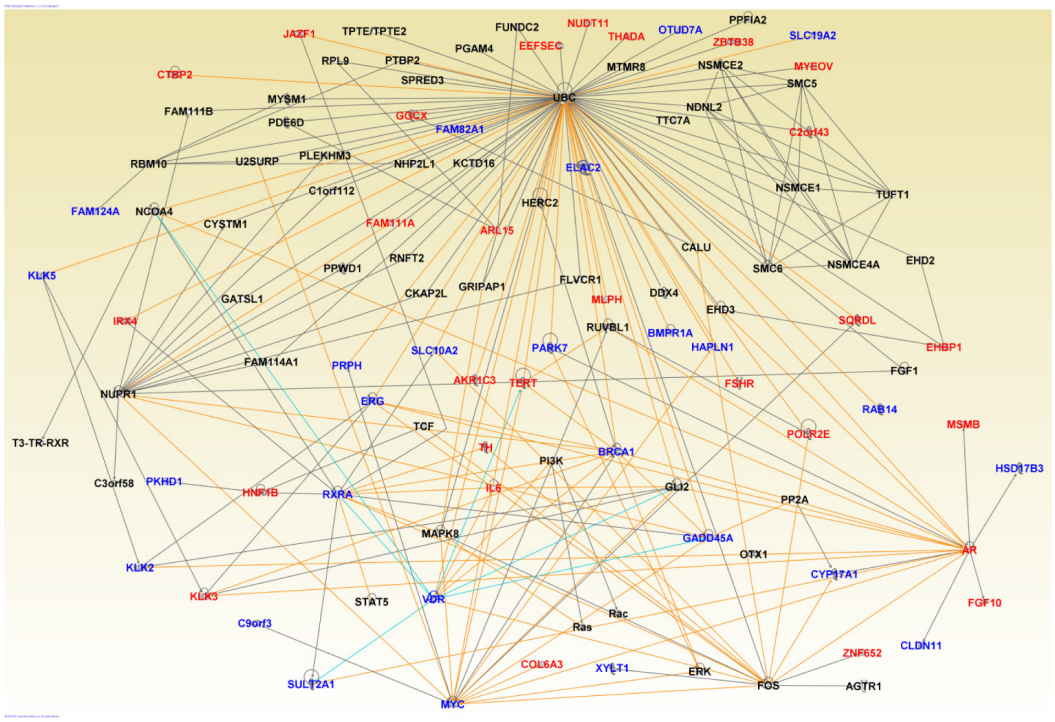

Figure 1. Gene regulatory networks of SNP-containing genes from GWAS reports and novel genes identified in this study not reported in GWAS. The genes containing SNPs with large effects and SNPs replicated in multiple independent studies are shown in red font, the genes containing SNPs with weak to moderate effects are shown in blue, whereas novel genes found in this study but not reported in GWAS reports are shown in black font. The genes are represented in nodes and edges (solid likes) represent interactions and functional relationships. Please note that the network is a merger of the 5 top networks which had the highest scores.

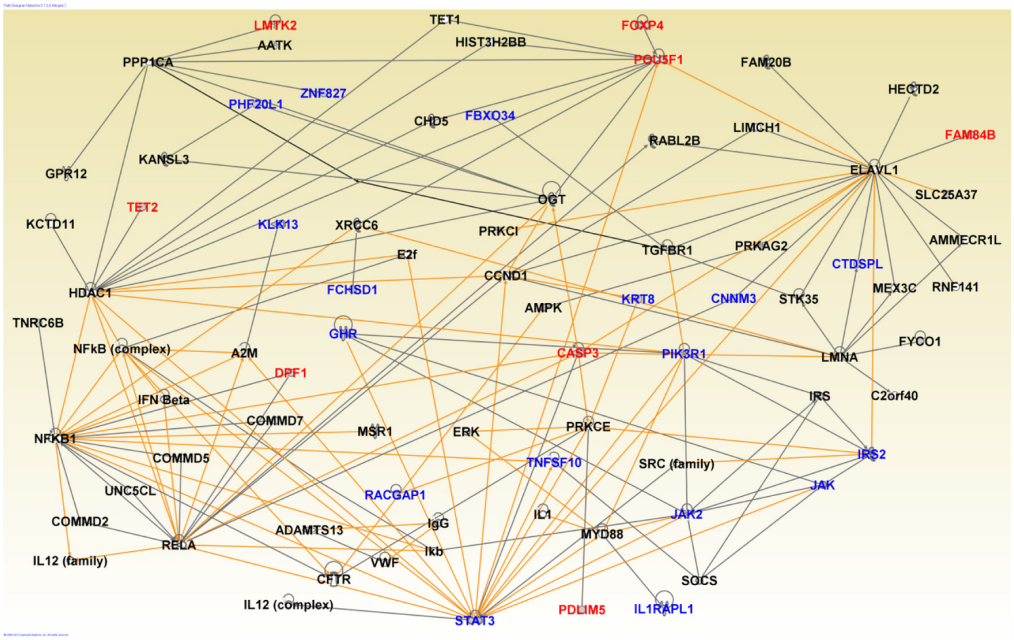

Figure 2. Gene regulatory networks of SNP-containing genes from GWAS reports and novel genes identified in this study not reported in GWAS. The genes containing SNPs with large effects and SNPs replicated in multiple independent studies are shown in red font, the genes containing SNPs with weak to moderate effects are shown in blue, whereas novel genes found in this study but not reported in GWAS reports are shown in black font. The genes are represented in nodes and edges (solid likes) represent interactions and functional relationships. Please note that the network is a merger of the second 4 top networks which had the highest scores.

single SNP GWAS analysis alone. Given that genetic variants with strong associations and genetic variants replicated in multiple independent studies explain only a small fraction of the variation, the novel genes could partially explain the missing variation not explained by GWAS.

To further understand the broader context in which the genetic variants operate and to identify the biological pathways enriched for SNPs, we mapped the SNP-containing genes onto the canonical pathways using the Ingenuity IPA System. The pathways were ranked according to *P* value. The top 15 most highly significant biological pathways enriched for SNPs are presented in Figure 3. Pathway analysis revealed multigene pathways. The top canonical pathways included Thrombopoieting signaling pathway (1.46E-06), the *IGF-1* signaling (4E-05), the role of *JAK2* in hormone-like cytokine signaling (6.74E-05), the growth hormone signaling (8.53E-05), and the androgen biosynthesis pathway (9.24E-05). Additional biological pathways enriched for genetic variants included the prolactin signaling pathway, the VDR/RXR activation, the molecular mechanisms of cancer, the mouse embryonic stem cell pluripotency, the renin-angiotensin signaling pathway, the role of *JAK1* and *JAK3* in cytokine signaling, the androgen signaling pathway, and the RAR activation pathway (Fig. 3).

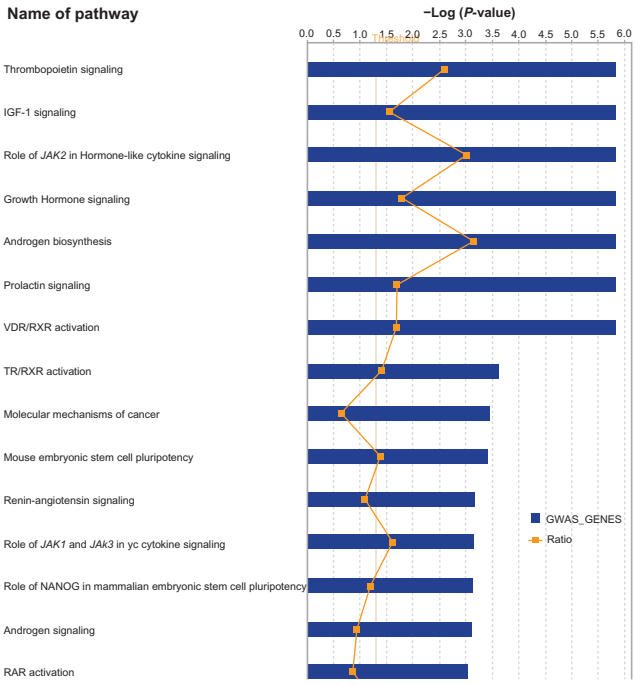

Figure 3. Top 15 most highly significant biological pathways enriched for SNPs and genes associated with increased risk of developing prostate cancer derived from GWAS. Also mapped to the pathways are novel genes not reported in GWAS reports. The y-axis shows the names of the pathway colored in blue bars. The numbers on the top of the x-axis denote the log *P* value indicating the significance level of the pathway enriched for SNPs and associated genes. The thin yellow line indicates the threshold level for declaring significance. The orange line denotes the ratio of SNP containing genes to the molecules in the pathways.

To understand the role and the significance of the networks and biological pathways enriched for SNPs and associated genes as potential clinically actionable biomarkers, we examined the literature on prostate cancer. The goal was to determine whether the networks and biological pathways enriched for genetic variants contain or are key drivers of prostate cancer and have the potential to be clinically actionable biomarkers. Interestingly, almost all the pathways have been implicated in prostate cancer pathogenesis. Here we provide a summary of this exploratory analysis.

Among the genes containing SNPs associated with increased risk of developing prostate cancer found to be interacting in gene regulatory networks and biological pathways included the genes *AR, KLK3, NKX3.1, NCOA4, STAT3, JAK2, HSD17B3, AKR1C3*, and *SULT2A1*. These genes contain genetic variants that have been associated with the androgen signaling in men from the populations of European ancestry.[26] Network and pathway analysis also revealed the genes *PARK7, SLC19A2, PLEKHM3, IKZF2, GHR, JAZF1, DERLI, PHF20L1, MSMB, GPATCH2L, OTUD7A, XYLT1*, and *TNRC6B*. These genes contain AR-binding SNPs associated with prostate cancer risk.[27] The androgen signaling pathways play key roles in the pathogenesis and treatment of prostate cancer. The SNPs mapped to the genes *CYP17A1, HSD17B4, JAK2, NCOA4*, and *SULT2A1* have been associated with aggressiveness in prostate cancer.[26] In particular, the SNP-containing genes *KLK2, KLK3*, and *NKX3.1* are well-characterized androgen regulated genes.[28] The most well studied of these androgen regulated genes is the prostate-specific antigen (PSA/KLK3), which is a widely used clinical marker for detection and monitoring of prostate cancer progression.[29] Expression of PSA is prostate-specific and regulated by androgens, and increased PSA levels may be an indication of prostate abnormalities.[29] The *KLK2* has very close structural homology to *PSA*/*KLK3*. Like *KLK3*, *KLK2* is

androgen regulated and may have utility as a prostate cancer biomarker in conjunction with *KLK3*, which is expressed at low levels compared with *KLK2* in poorly differentiated tumors.[29]

The identification of pathways enriched for genetic variants is of particular interest because directed therapies targeting these pathways could be developed. For example, since the androgen signaling and biosynthesis pathways play key roles in prostate cancer, they could be targeted rather than individual SNPs. The androgen receptor (*AR*) is required for prostate cancer growth in all stages, including the relapsed, androgen-independent tumors in the presence of very low levels of androgens.[28] When prostate cancers progress following androgen depletion therapy, there are currently few treatment options, with only 1, docetaxel, that has been shown to prolong life.[30] Moreover, recent work has shown that castration-resistant prostate cancers continue to depend on *AR* signaling which is activated despite low serum androgen levels.[30] For example, *STAT3* also found in this study interacts with *AR* to enhance *AR* activity.[31] The practical implication is that directed therapies could be developed targeting the androgen-receptor signaling axis.[32]

In light of the interactions between the *AR* and other genes containing genetic variants associated with increased risk of developing prostate cancer, it is conceivable that adaptation of prostate cancer cells to androgen deprivation may involve both mutations and amplification of *AR*.[33] Alterations of *AR* functions could be mediated by protein kinase signaling pathways activated by peptide hormones and local growth factors that are known to promote proliferation and survival of prostate cancer cells either directly or through stimulation of *AR* action.[33] One such local growth factor-initiated protein kinase signaling pathway identified in this study is the prolactin signaling pathway (Fig. 3).

The insulin growth factor (IGF) signaling pathway identified in this study has been implicated in prostate cancer.[34] Research has associated circulating *IGF-1* with prostate cancer risk, and studies have elucidated the implication of the IGF network in the early stages of prostate cancer.[34] Most notably, it has been reported that *IGF-1* induces ligand-independent activation of the *AR* and enhances expression of the *KLK2*, a homology and functionally related gene to *PSA*/*KLK3* gene, the main diagnostic marker in prostate cancer. In addition, progression to androgen independence has been linked to deregulation of the *IGF-1*-*IGF-1*-receptor axis.[34] The genes like *STAT3* and *JAK2* identified in this study are involved in the JAK and STAT pathways reported in Figure 3 and have been associated with prostate cancer risk.[26] The prolactin signaling pathway plays an important role in prostate cancer. For example, autocrine prolactin promotes cancer cells growth via janus kinase-2-signal transducer and activator of transcription-5a/b signaling pathway.[35] Overall, the clinical significance of the SNP-containing genes, the gene regulatory networks, and biological pathways reported in this study lies in the fact that they represent potential biomarkers and important therapeutic focal points.

Although the AR signaling axis featured prominently in both network and pathway analyses in this study, a number of other pathways enriched for genetic variants were also identified. The interactions between the AR and other pathways suggest cross-talk between AR and other signaling pathways, most notably the JAK, STAT, IGF, and PRL pathways. The ubiquitin axis may be involved in the development and progression of prostate cancer. These results are consistent with literature reports on prostate cancer.[28,34] These findings taken together demonstrate that genetic variants associated with increased risk of developing prostate cancer are likely to affect entire network states and biological pathways that in turn increase or decrease the risk of developing prostate cancer or amplify the severity of the disease. The practical significance of the results from network and pathway analysis is that identification of potential causal pathways provides new opportunities for identification of potential therapeutic targets within the identified pathways.

Discussion

We have developed a comprehensive catalogue of genetic variants and genes associated with increased risk of developing prostate cancer. Additionally, we performed network and pathway analysis to identify gene regulatory networks and biological pathways enriched for genetic variants. The significance of the results in this study can be summarized as follows.

First, the integration of GWAS information and network and pathways analysis provides putative

functional bridges between GWAS findings and biological pathways relevant to prostate cancer, thus providing insights about the broader biological context in which genetic variants and associated genes operate.[3] This serves as a powerful approach to identifying the biological mechanisms underlying GWAS findings. The identified genes and biological pathways could be prioritized for targeted sequencing as potential clinically actionable biomarkers. Because GWAS does not necessarily identify causal variants or genes,[3] and the fact that many of the identified genetic variants have not been replicated and explain only a small proportion of the variation, network and pathways analyses provide a powerful and complementary approach to holistically unravel the complex genetic susceptibility landscape of prostate cancer in order to gain insights about the genomic mechanisms underlying the disease.

Our analysis suggests that while, so far, most efforts at replication have concentrated on the genetic variants with strong evidence of association,[2] efficient identification of additional susceptibility loci with small to moderate effects might benefit from the integration of statistical evidence with assessment of functional candidacy achievable through network and pathway analysis.[5] The results of network analysis suggest that prostate cancer susceptibility effects are likely mediated through a constellation of genes containing genetic variants with both small and large effects interacting with one another. The involvement of multigene pathways found in this study is consistent with literature reports.[28–33]

Second, network and pathway analysis revealed that genes containing genetic variants strongly associated with increased risk of developing prostate cancer are functionally related and interact with one another and with genes containing genetic variants with small to moderate associations. This is an important finding in that prostate cancer is a complex disease originating from joint actions of many genes and many pathways.[3] The results of network and pathway analysis in this study demonstrate clearly that if we focus on only those genes containing genetic variants with strong associations and those containing genetic variants replicated in multiple independent studies, genes and genetic variants that jointly have significant effects or play key roles in prostate cancer but individually have a small effect will be missed. This is an important finding because rare variants cannot be captured using GWAS analysis.

Third, the comprehensive catalogue developed in this study demonstrates that GWAS has uncovered many loci that are associated with prostate cancer. But 2 fundamental limitations have hampered our ability to translate the GWAS results into clinically actionable biomarkers. First, the identified genetic variants and associated genes generally explain very little of the disease risk and the variation. Second, the functions of the majority of the genetic variants, specifically those in the intergenic regions and noncoding regions of the genes remain largely unknown. Network and pathway analyses provide a unified approach for understanding the broader biological context in which a given potential causal genetic variant and associated gene for prostate cancer operate. This is a necessary step in identifying targets for the development of novel therapeutic strategies and early interventions. Most notably, as demonstrated in this study, network and pathway analyses revealed novel genes that have not been reported in GWAS. This is a significant finding in that it identifies genes that could potentially explain the missing variation not explained by GWAS. Importantly, many of these novel genes are likely to mediate the actions of the SNP-containing genes.

As discussed in the Results section of this study, understanding the biological context in which SNP-containing genes operate is a necessary step in identifying potential drug targets.[30] This might involve identification of therapeutic targets within potential causal pathways such as the androgen signaling pathway that could lead to the development of novel and more effective therapies and prevention strategies. Identification of potential causal pathways should also bolster our efforts to identify biomarkers, allowing for improved prostate cancer prediction and monitoring of disease progression and treatment responses. Most notably, as evidenced in this study and other studies in which we have previously reported on breast cancer,[5,11] even genes containing genetic variants with small to moderate effects (provided they are confirmed as genuine through replication and functional studies) can offer significant new translational opportunities through the identification of novel modifiable pathways.

Beyond identification of functionally related genes, gene regulatory networks, and biological pathways to gain insights about the biological context in which genetic variants operate, the integrative analysis approach presented here has another application. That is, it could be used to identify candidate genes and pathways to prioritize for targeted sequencing. Gene and pathway prioritization aims at identifying the most promising genes and pathways that could be used as clinically actionable biomarkers or potential targets for the development of novel therapeutic and early intervention strategies. As demonstrated in this study, using network and pathway analyses, genes can be prioritized on the basis of putative links to other genes that contain genetic variants with strong associations or to other genes that have been implicated in prostate cancer or the process of interest, notably resistance to drug treatment or chemical castration.[33] This broadening of applications is beginning to take hold with targeted and clinically directed sequencing approaches.

In the published reports on GWAS, meta-analysis has been used to increase the sample size and the power to identify genetic variants with strong associations.[36] Using traditional GWAS analysis of meta-analysis to identify loci that are associated with prostate cancer is an important and laudable step for dissecting the genetic susceptibility landscape of prostate cancer. However, the view emerging from this and many other genomic studies on prostate cancer is that prostate cancer is an emergent property of gene regulatory networks and pathways whose states are affected by genetic alterations and complex interactions between genetic and environmental factors.[3] To understand the effects of any 1 gene containing genetic variants associated with prostate cancer as demonstrated in this study, individual genes must be understood in the context of molecular networks and biological pathways. Based on this reasoning, in this study, we did not use meta-analysis because our main goal was to gain insights about the broader biological context in which the genetic variants and associated genes operate and to identify the molecular mechanisms underlying GWAS findings. This is an endeavor that could not be achieved by meta-analysis. However, our method is complementary to GWAS analysis and relies on GWAS findings.

The analysis presented in this study demonstrates the power of network and pathway analyses combined with biological information to gain insights about the broader biological context in which genetic variants and associated genes operate and to identify novel genes. To our knowledge, this is the first study in prostate cancer to provide a comprehensive catalogue of genetic variants, to assess their credibility, and to integrate GWAS information with biological knowledge through functional, network, and pathway analyses to identify the molecular mechanisms underlying GWAS findings in prostate cancer. However, limitations must be acknowledged. It is conceivable that some of the genetic variants, particularly those with very weak associations and those not replicated in multiple independent studies, are false discoveries. But many are likely to be genetic variants with genuine associations with small effects. We could not independently validate the genetic variants, and, therefore, the results in this study should be interpreted conservatively. We have used publicly available information, and our GWAS information relies on the methods used in the reports from which data were extracted. Use of publicly available data has many limitations including use of heterogeneous methods and sample sizes, genotyping errors, methods not accounting for population structures, and admixing of the populations, all which could influence the results in this study. Accounting for those factors was beyond the scope of this study. Although we were very careful in selecting the reports from which the data were extracted as described in the Methods section, some errors could still exist. Our study did not evaluate the genes containing SNPs associated with increased risk of developing prostate cancer in the disease state using gene expression information. However, gene expression can be population-specific, and these studies were conducted in many different populations. It is also worth noting that our study did not consider the environmental and socioeconomic factors. Conducting a gene expression study that encompasses all the populations represented in this study along with environmental factors and socioeconomic conditions was beyond the scope of this report, and that is a weakness that we readily acknowledge. Work on integrating GWAS information with gene expression data in prostate cancer

targeting specific populations is ongoing and will be reported elsewhere.

In summary, we have developed a comprehensive catalogue and assessed the credibility of genetic variants associated with increased risk of developing prostate cancer. We demonstrated that integrative analysis combining GWAS information with biological information through function, network, and pathway analysis provides insights about the broader biological context in which genetic variants and associated genes operate. Most notably, we showed that integration of GWAS information with network and pathway analysis enables identification of novel genes. The approach provides putative functional bridges between GWAS information and gene regulatory networks and biological pathways, thereby serving as a powerful approach to elucidating the molecular mechanisms underlying GWAS findings. In conclusion, we show that the emerging genetic susceptibility landscape of prostate cancer is complex and involves interactions between constellations of genes containing genetic variants (with both small and large effects) interacting with one another and with a broad range of novel genes that are potential mediators. These complex arrays of interacting genes map to gene regulatory networks and biological pathways enriched for genetic variants associated with an increased risk of developing prostate cancer and/or are key drivers of the disease. More research is needed to test the ability of the identified genetic variants, genes, and biological pathways to function as clinically actionable biomarkers or potential targets for drug development.

Author Contributions

Conceived and designed the experiments: CH, SV, LM. Analyzed the data: CH, TK. Wrote the first draft of the manuscript: CH, SV, TK, LM. Contributed to the writing of the manuscript: CH, SV, LM, TK. Agree with manuscript results and conclusions: CH, SV, LM, TK. Jointly developed the structure and arguments for the paper: CH, SV, LM, TK. Made critical revisions and approved final version: CH, SV, LM, TK. All authors reviewed and approved of the final manuscript.

Funding

This work was supported by funding from the Cancer Institute and the ARIC Cancer Group.

Competing Interests

Author(s) disclose no potential conflicts of interest.

Disclosures and Ethics

As a requirement of publication the authors have provided signed confirmation of their compliance with ethical and legal obligations including but not limited to compliance with ICMJE authorship and competing interests guidelines, that the article is neither under consideration for publication nor published elsewhere, of their compliance with legal and ethical guidelines concerning human and animal research participants (if applicable), and that permission has been obtained for reproduction of any copyrighted material. This article was subject to blind, independent, expert peer review. The reviewers reported no competing interests.

References

1. American Cancer Society. Cancer facts and figures 2012. Atlanta: American Cancer Society; 2012. http://www.cancer.org/cancer/cancerbasics/index.
2. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;106(23):9362–7.
3. Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature*. 2009;461:218–23.
4. Luo L, Peng G, Zhu Y, Dong H, Amos CI, Xiong M. Genome-wide gene pathway analysis. *Europ J Hum Genet*. 2010;18:1045–53.
5. Hicks C, Asfour R, Pannuti A, Miele L. An integrative genomics approach to biomarker discovery in breast cancer. *Cancer Inform*. 2011;10:185–204.
6. Ioannidis JP, Boffetta P, Little J, et al. Assessment of cumulative evidence on genetic associations: Interim guidelines. *Intl J Epidemiol*. 2008; 37:120–32.
7. Khoury MJ, Bertram I, Boffetta P, et al. Genome-wide association studies, field synopses, and the development of the knowledge base on genetic variation in human diseases. *Am J Epidemiol*. 2009;170:269–79.
8. Sagoo GS, Little J, Higgins JP. Systematic reviews of genetic association studies. Human Genome Epidemiology Network. *PLoS Med*. 2009;6:e28.
9. Moher D, Liberati A, Tetzlaff J. Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151:264–9.
10. Liberati A, Altman DG, Tetzlaff J, Mulrow C, et al. The PRISMA statement for reporting systematic reviews and meta-analyses studies that evaluate health care interventions: Explanation and elaboration. *PLoS Med*. 2009;6(7):e1000100.
11. Hicks C, Kumar R, Pannuti A, et al. An integrative genomics approach for associating GWAS information with triple-negative breast cancer. *Cancer Inform*. 2013;12:1–20.
12. Zhang B, Beeghly-Fadiel A, Long J, Wei Z. Genetic variants associated with breast-cancer risk: Comprehensive research synopsis, meta-analysis, and epidemiology evidence. *Lancet*. 2011;12:477–88.
13. Grasso CS, Wu YM, Robinson DR, et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature*. 2012;487(7406):239–43.
14. Berger MF, Lawrence MS, Demichelis F, et al. The genomic complexity of primary human prostate cancer. *Nature*. 2011;470(7333):214–20.
15. Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med*. 2010;363(2):166–76.
16. Neal BM, Sham PC. The future of association studies. Gene-based analysis and replication. *Am J Hum Genet*. 2004;75:353–62.

17. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontolgy Consortium. *Nat Genet*. 2000;25(1):25–9.
18. Ingenuity Pathways Analysis (IPA) system. Redwood, CA: Ingenuity Systems, Inc.; http://www.ingenuity.com/.
19. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst*. 2004;96:434–42.
20. Chanock SJ, Manolio T, Boehnke M, et al; NCI-NHGRI Working Group on Replication in Association Studies. Replicating genotype-phenotype associations. *Nature*. 2007;447:655–60.
21. Ioannidis JP. Non-replication and inconsistency in the genome-wide association setting. *Hum Hered*. 2007;64:203–13.
22. Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet*. 2006;7:781–91.
23. Pompanon F, Bonin A, Bellemain E, Taberlet P. Genotyping errors: causes, consequences and solutions. *Nat Rev Genet*. 2005;6:847–59.
24. Clayton DG, Walker NM, Smyth DJ, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet*. 2005;37:1243–6.
25. Page GP, George V, G RC, Page PZ, Allison DB. Are there yet? Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. *Am J Hum Genet*. 2003;73:711–9.
26. Kwon EM, Holt SK, Fu R, et al. Androgen metabolism and JAK/STAT pathway genes and prostate cancer risk. *Cancer Epidemiol*. 2012;36:347–53.
27. Feng J, Sun J, Kim S-T, Lu Y, et al. A genome-wide survey over the ChIP-On-Chip identified androgen receptor-binding genomic regions identifies a novel prostate cancer susceptibility locus at 12q13.13. *Cancer Epidemiol Biomarkers Prev*. 2011;20(11):2396–403.
28. Kaarbo M, Klokk TI, Saatcioglu F. Androgen signaling and its intervention with other signaling pathways in prostate cancer. *Bioessays*. 2007;29:1227–38.
29. Rittenhouse HG, Finlay JA, Mikolajczyk SD, Partin AW. Human kallikrein 2 (hk2) and prostate-specific antigen (PSA): two closely related, but distinct, kallikreins in the prostate. *Crit Rev Clin Lab Sci*. 1998;35:275–368.
30. Chen Y, Sawyers CL, Scher HI. Targeting the androgen receptor pathway in prostate cancer. *Curr Opin Pharmacol*. 2008;8:440–8.
31. Heinlein CA, Chang C. Androgen receptor in prostate cancer. *Endoc Rev*. 2004;25(2):276–308.
32. Scher HI, Sawyers CL. Biology of progressive, castration-resistant prostate cancer: Directed therapies targeting the androgen-receptor signaling axis. *J Clin Oncol*. 2005;23(32):8253–61.
33. Knusen KE, Penning TM. Partners in crime: deregulation of AR activity and androgen synthesis in prostate cancer. *Trends Endocrinol Metab*. 2010;21(5):315–23.
34. Papatsoris A, Karamouzis MV, Papavassiliou AG. Novel insights into the implication of the IGF-1 network in prostate cancer. *Trends Mol Med*. 2005;11(2):52–5.
35. Dagvadorj A, Collins S, Jomain J-B, et al. Autocrine prolactin promotes cancer cell growth via janus kinase-2-signal transducer and activator of transcription-5a/b signaling pathway. *Endocrinology*. 2007;148:3089–101.
36. Amin Al, Olama A, Kote-Jarai Z, Schumacher FR, et al. A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease. *Hum Mol Genet*. 2013;22(2):408–15.

Supplementary Tables

Table S1. A comprehensive list of single nucleotide polymorphisms (herein called genetic variants) and associated genes associated with increased risk of developing prostate cancer and published GWAS reports denoted by the PubMed ID and actual reference from which the data were extracted.

Table S2. A comprehensive list of molecular functions, biological process, and cellular components in which genes containing single nucleotide polymorphisms associated with increased risk of developing prostate cancer are involved as determined by GO analysis. Also included in the Table are the GO IDs and GO terms associated with the genes as reported in the GO Database.