Libertas Academica
FREEDOM TO RESEARCH

Open Access: Full open access to this and thousands of other papers at http://www.la-press.com.

Biomarker Insights

Integrative Genomic Analysis for the Discovery of Biomarkers in Prostate Cancer

Chindo Hicks[1–4], Tejaswi Koganti[1], Shankar Giri[1], Memory Tekere[5], Ritika Ramani[1], Jitsuda Sitthi-Amorn[1] and Srinivasan Vijayakumar[3]

[1]Cancer Institute, University of Mississippi Medical Center, Jackson, MS, USA. [2]Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. [3]Department of Radiation Oncology, University of Mississippi Medical Center, Jackson, MS, USA. [4]Department of Public Health Sciences, University of Lusaka, Lusaka, Zambia. [5]Department of Environmental Sciences, University of South Africa, UNISA Florida Campus, Florida, South Africa.

ABSTRACT: Genome-wide association studies (GWAS) have achieved great success in identifying single nucleotide polymorphisms (SNPs, herein called genetic variants) and genes associated with risk of developing prostate cancer. However, GWAS do not typically link the genetic variants to the disease state or inform the broader context in which the genetic variants operate. Here, we present a novel integrative genomics approach that combines GWAS information with gene expression data to infer the causal association between gene expression and the disease and to identify the network states and biological pathways enriched for genetic variants. We identified gene regulatory networks and biological pathways enriched for genetic variants, including the prostate cancer, IGF-1, JAK2, androgen, and prolactin signaling pathways. The integration of GWAS information with gene expression data provides insights about the broader context in which genetic variants associated with an increased risk of developing prostate cancer operate.

KEY WORDS: GWAS, genetic variants, gene expression, prostate cancer

CITATION: Hicks et al. Integrative Genomic Analysis for the Discovery of Biomarkers in Prostate Cancer. *Biomarker Insights* 2014:9 39–51 doi: 10.4137/BMI.S13729.

RECEIVED: November 27, 2013. **RESUBMITTED:** April 3, 2014. **ACCEPTED FOR PUBLICATION:** April 6, 2014.

ACADEMIC EDITOR: Karen Pulford, Associate Editor

TYPE: Original Research

FUNDING: The study was supported by the startup funds from the University of Mississippi Medical Center's Cancer Institute. The authors acknowledge this support.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

COPYRIGHT: © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

CORRESPONDENCE: chicks2@umc.edu

This paper was subject to independent, expert peer review by a minimum of two blind peer reviewers. All editorial decisions were made by the independent academic editor. All authors have provided signed confirmation of their compliance with ethical and legal obligations including (but not limited to) use of any copyrighted material, compliance with ICMJE authorship and competing interests disclosure guidelines and, where applicable, compliance with legal and ethical guidelines on human and animal research participants.

Introduction

Genome-wide association studies (GWAS) provide a comprehensive and an unbiased assessment of single nucleotide polymorphism (SNPs) (herein called genetic variants) associated with an increased risk of developing prostate cancer.[1,2] These findings are providing valuable clues about the emerging genetic susceptibility landscape of prostate cancer.[2] However, despite considerable progress, GWAS-defined loci, singly or in aggregate, typically explain only a small proportion of the heritable variation, and they do not typically inform the broader context in which the disease genes operate, thereby providing limited insights into the mechanisms driving prostate cancer.[2,3]

Prostate cancer is a polygenic disease originating from a more complex interplay between constellations of genetic alterations involving both common and rare variations and abroad range of nongenetic factors. These complex arrays of interacting factors affect entire network states and biological pathways that in turn increase or decrease the risk of developing prostate cancer or affect the severity of the disease.[4] Therefore, if we consider only the common genetic variants that are strongly associated with prostate cancer, the common genetic variants and rare variants that jointly have significant risk effects, but individually making a small contribution, will be missed.[5,6] Most notably, the genes and pathways that likely mediate the actions of SNP-containing genes and pathways may be missed.[5,6]

One way to account for the missing information from GWAS is by integrating GWAS information with gene

expression data and biological information. This unified approach has the potential to identify novel genes that are functionally related with genes containing genetic variants associated with an increased risk of developing prostate cancer and to provide insights about the broader context in which genetic variants operate. We have recently demonstrated this approach in breast cancer.[5] But to date, the vast amounts of information generated from GWAS have not been leveraged with gene expression data to link GWAS findings to the disease state and to identify molecular networks and biological pathways enriched for genetic variants in prostate cancer.

Over the past decade, considerable efforts and financial resources have been directed at identifying molecular signatures for prostate cancer using transcription profiling.[7–10] However, although these primary analyses have made great strides in deciphering the molecular basis of prostate cancer, they have been unsuccessful in determining which genes have causative roles as opposed to being consequences of the prostate cancer state.

The objectives of this study were threefold: (i) to determine whether genes containing genetic variants associated with an increased risk of developing prostate cancer are functionally related and associated with the disease state, (ii) to gain insights about the broader context in which the genetic variants operate by identifying gene regulatory networks and biological pathways enriched for genetic variants, and (iii) to identify novel genes (ie, genes not identified by GWAS). The rationale being that identification of novel genes could explain the missing variation. Our working hypothesis was that genes containing genetic variants associated with an increased risk of developing prostate cancer are associated with the disease state, and that these genes are functionally related and interact with one another and with novel genes in gene regulatory networks and biological pathways enriched for genetic variants. Throughout this report, we have defined SNPs as genetic variants associated with an increased risk of developing prostate cancer, and used these terms interchangeably.

Material and Methods

Sources of information on genetic variants and associated genes. The GWAS information on prostate cancer, specifically the genetic variants and associated genes used in this study were based on publicly available data obtained from the published reports on GWAS and the website hosting supplementary data on the respective reports. The details about methods of data collection including inclusion and exclusion criteria as well as quality control were based on the guidelines proposed by the Human Genome Epidemiology Network for systematic review of genetic associations[11–15] and have been reported elsewhere.[2] Here, we provide a brief but detailed description of the data used in this study.

We examined a total of 140 published reports on GWAS. The reports 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 40 studies included removal of studies with insufficient or incomplete information, reviews, studies reporting only intergenic regions, and studies with very small sample sizes. GWAS were eligible to be included if they met the following criteria: (i) must have been based on a case-control design using unrelated individuals, (ii) publications must have been of full length and published in peer-reviewed journals or online in English before September 2013, (iii) prostate cancer must have been diagnosed by histological examination, (iv) must be based on sample sizes of greater than 500 subjects in the cases controls, (v) the study must have provided sufficient information such that genotype frequencies for both prostate cancer cases and controls could be determined without ambiguity, and (vi) the study must have used recommended statistical methods for assessing evidence of association by taking into account covariates and accounting for population structure.[11]

We manually extracted the information from the studies meeting our eligibility criteria and the accompanying websites containing supplementary data. The extracted information included SNP identification number (SNP-rsID), the SNP P value indicating the magnitude and strength of association, the gene name to which the identified SNP map, the chromosome position of the gene, the sample sizes for cases and controls used to detect the association, and the study reporting the association. Evidence and credibility of association were assessed using the procedures previously reported.[11] This assessment included the amount of evidence as determined by the association of P value, extent of replication, protection from bias, and a composite of strong ($P \leq 10^{-8}$), moderate ($P \approx 10^{-5}$–10^{-7}), or weak association ($P \approx 10^{-2}$–10^{-4}). This search yielded 450 SNPs mapped to 172 genes from a population of over 350,000 cases and over 350,000 controls. In addition, the search identified 300 SNPs mapped to intergenic regions, but these were not included in the final analysis. The SNP (rs-IDs), gene names, and their locations on the chromosomes were verified using the dbSNP database[16] and the Human Genome Nomenclature Committee (HGNC) database.[17]

Sources and characteristics of gene expression data. We used publicly available gene expression data. The data were downloaded from the NCBI Gene Expression Omnibus under accession numbers GSE32448 and GSE17951. Methods regarding experimental design, sample preparations, and data processing have been described by the data originators.[18] Briefly, the data included a total of 234 samples (194 prostate cancer patients and 40 cancer-free controls) from the populations of European ancestry. All the data were processed using the Affymetrix platform based on the Human GeneChip U133Plus 2.0, which contains ~54,000 probes, using standard Affymetrix protocols. Expression data (average scaled difference values) were processed and normalized using the Affymetrix Microarray Analysis Software (MAS 5.0). The data were filtered out to remove spiked control genes. A total

of 154 SNP-containing genes were represented on the Chip and used in the analysis. The discrepancy between the number of genes (172) containing SNPs associated with an increased risk of developing prostate cancer and the number of SNP-containing genes represented on the Chip can be explained by two factors, namely, discrepancies in GWAS annotations and lack of representation of some SNP-containing genes on the U133Plus 2.0 Chip.

Data analysis. We performed supervised analysis comparing expression levels of SNP-containing genes between prostate cancer samples and matched cancer-free control samples using a *t* test. The goal of this analysis was to determine whether genes containing genetic variants associated with an increased risk of developing prostate cancer are associated with the disease state and to identify a molecular signature of SNP-containing genes and novel genes distinguishing the two groups. We performed additional supervised analysis on the whole data sets to identify novel genes (ie, genes not identified by GWAS) which were highly significantly differentially expressed between cases and controls. Permutation test was applied to reliably estimate the *P* values. We used a false discovery rate (FDR)[19] to correct for multiple hypothesis testing. Because of small sample sizes, we did not partition the data into test and validation sets, but instead we used an out of sample (leave-one-out) validation procedure to identify genes with predictive power.[20] Supervised analyses were performed using GenePattern[21] and Pomelo II software packages.[22]

To determine whether genes containing SNPs associated with an increased risk of developing prostate cancer are functionally related and have similar patterns of expression profiles with each other and with novel genes, we performed unsupervised analysis using hierarchical clustering. We computed the Pearson correlation coefficients between all possible pairs of significantly differentially expressed genes. Using the Pearson correlation coefficient as the distance measure and the complete linkage method, the genes were subjected to hierarchical clustering using GenePattern.[21] Before clustering, gene expression data were normalized using the median normalization and were standardized and centered.[23] In addition, we performed gene ontology (GO)[24] analysis to gain insights about the molecular functions, biological processes, and cellular components in which the SNP-containing and novel genes are involved. To gain insights about the broader context in which the genes containing genetic variants associated with an increased risk of developing prostate cancer and novel genes operate, we performed network and pathway analyses and visualization using the Ingenuity pathway analysis (IPA) program (http://www.ingenuity.com).[25]

Results

Associating GWAS information with disease state. One of the primary goals of this investigation was to determine whether genes containing SNPs associated with an increased risk of developing prostate cancer are associated with the disease state. We addressed this question by comparing gene expression levels of SNP-containing genes between prostate cancer samples and matched cancer-free control samples. Our working hypothesis was that genes containing SNPs associated with an increased risk of developing prostate cancer are significantly differentially expressed between cancer patients and cancer-free controls. Out of a total of 154 genes containing SNPs associated with an increased risk of developing prostate cancer evaluated, 131 genes were found to be significantly ($P < 0.05$) differentially expressed between prostate cancer and control samples. Table 1 shows a list of 47 genes containing SNPs with strong associations and SNPs replicated in multiple independent GWAS. A complete list of all the 131 genes containing SNPs associated with an increased risk of developing prostate cancer that were found to be significantly associated with the disease state are presented in Table S1, provided as supplementary data to this report.

Among the significantly differentially expressed SNP-containing genes identified, 29 genes including *FGF10*, *MLPH*, *ZBTB38*, *ZNF652*, *AR*, *AR15*, *CCHCR1*, *FAM84B*, *FSHR*, *GGCX*, *IL16*, *IRX4*, *MYEOV*, *SKIL*, *BIK*, *C2ORF43*, *CTBP2*, *EEFSEC*, *EHBP1*, *FOXP4*, *GPRC6A*, *HNF1B*, *ITGA6*, *JAZF1*, *KLK3*, *MSMB*, *NUDT11*, *PDLIM5*, and *TET2* contained SNPs with strong ($P < 10^{-8}$) associations (Table 1). Another 33 genes including *BIK*, *BMP5*, *C2ORF43*, *CASP3*, *CNGB3*, *CTBP2*, *EEFSEC*, *EHBP1*, *FOXP4*, *FREM1*, *GPRC6A*, *HERC2*, *HNF1B*, *ITGA6*, *JAZF1*, *KLK15*, *KLK2*, *KLK3*, *LMTK2*, *LOC729852*, *MSMB*, *MSR1*, *NCOA4*, *NKX3-1*, *NSMCE2*, *NUDT11*, *PDLIM5*, *RFX6*, *SLC22A3*, *SLC25A37*, *TET2*, *TNSF10*, and *TNRC6B* contained SNPs replicated in multiple independent studies (Table 1).

Interestingly, the genes containing SNPs with moderate ($P \sim 10^{-5}$–$P \sim 10^{-7}$) to weak ($P \sim 10^{-2}$–$P \sim 10^{-4}$) associations were found to be significantly associated with the disease state (Table S1). This is a significant finding given that relatively few SNPs have *P* values sufficiently small and/or replicated in multiple independent studies to give conclusive evidence of association.[26] The identification of SNP-containing genes that are significantly differentially expressed between tumor and control samples confirmed our hypothesis that genes containing genetic variants associated with an increased risk of developing prostate cancer are associated with the disease state. A small number (23 genes) of the 154 SNP-containing genes evaluated did not exhibit significant differences in expression levels between tumor and control samples. This could partially be explained by the genetic heterogeneity inherent in GWAS.[27]

The findings from published GWAS reports explain only a small proportion of the heritable variation,[28] raising the question "where is the missing variation?". To partially address this question, we performed additional supervised analysis comparing gene expression levels between prostate cancer patients and cancer-free controls on the whole data set to identify

Table 1. List of significantly differentially expressed genes containing SNPs with strong associations and SNPs replicated in multiple independent studies.

GENE SYMBOL	CHROMOSOME POSITION	SNP GWAS *P* VALUE	GENE EXPRESSION *P* VALUE
BIK	22q13.31	$1.30 \times 10^{-12} - 3.01 \times 10^{-3}$	5.0×10^{-6}
BMP5	6p12.1	$3.0 \times 10^{-2} - 4.0 \times 10^{-2}$	5.0×10^{-6}
C2ORF43	2p24.1	$7.5 \times 10^{-8} - 1.0 \times 10^{-2}$	5.0×10^{-6}
CASP3	4q34	$4.0 \times 10^{-2} - 4.0 \times 10^{-2}$	5.0×10^{-6}
CNGB3	8q21.3	$2.79 \times 10^{-2} - 1.0 \times 10^{-2}$	5.0×10^{-6}
CTBP2	10q26.13	$2.7 \times 10^{-8} - 3.0 \times 10^{-2}$	5.0×10^{-6}
EEFSEC	3q21.3	$2.3 \times 10^{-8} - 3.30 \times 10^{-3}$	5.0×10^{-6}
EHBP1	2p15	$7.7 \times 10^{-9} - 4.0 \times 10^{-2}$	5.0×10^{-6}
FGF10	5p13-p12	4.0×10^{-8}	5.0×10^{-6}
FOXP4	6p21.1	$7.6 \times 10^{-8} - 2.0 \times 10^{-2}$	5.0×10^{-6}
FREM1	9p22.3	$2.0 \times 10^{-3} - 2.0 \times 10^{-3}$	3.9×10^{-2}
GPRC6A	6q22.31	$1.6 \times 10^{-12} - 2.0 \times 10^{-3}$	2.0×10^{-2}
HERC2	15q13	$5.20 \times 10^{-5} - 4.0 \times 10^{-3}$	5.0×10^{-6}
HNF1B	17q12	$1.13 \times 10^{-25} - 1.0 \times 10^{-2}$	5.0×10^{-6}
ITGA6	2q31.1	$9 \times 10^{-23} - 2.0 \times 10^{-2}$	5.0×10^{-6}
JAZF1	7p15.2-p15.1	$7.05 \times 10^{-14} - 4 \times 10^{-2}$	5.0×10^{-6}
KLK15	19q13.4	$2.7 \times 10^{-4} - 1.0 \times 10^{-2}$	1.8×10^{-4}
KLK2	19q13.33	$9.0 \times 10^{-3} - 4.0 \times 10^{-2}$	5.0×10^{-6}
KLK3	19q13.41	$1.6 \times 10^{-24} - 3 \times 10^{-10}$	5.0×10^{-6}
LMTK2	7q22.1	$1.1 \times 10^{-9} - 2 \times 10^{-2}$	5.0×10^{-6}
LOC729852	7p21.3	$8.0 \times 10^{-3} - 1.0 \times 10^{-2}$	5.0×10^{-6}
MLPH	2q37.2	4.0×10^{-8}	5.0×10^{-6}
MSMB	10q11.2	$8.7 \times 10^{-29} - 1.0 \times 10^{-2}$	3.0×10^{-5}
MSR1	8p22	$9.0 \times 10^{-3} - 2.0 \times 10^{-2}$	5.0×10^{-6}
NCOA4	10q11.2	$5.6 \times 10^{-3} - 7.0 \times 10^{-3}$	5.0×10^{-6}
NKX3–1	8p21.2	$5.52 \times 10^{-7} - 7.0 \times 10^{-3}$	5.0×10^{-6}
NSMCE2	8q24.13	$5.0 \times 10^{-4} - 4.0 \times 10^{-2}$	5.0×10^{-6}
NUDT11	Xp11.22-p11.1	$1.00 \times 10^{-47} - 4.0 \times 10^{-2}$	5.0×10^{-6}
PDLIM5	4q22	$4.2 \times 10^{-15} - 7.30 \times 10^{-2}$	5.0×10^{-6}
RFX6	6q22.31	$3.1 \times 10^{-6} - 4.43 \times 10^{-5}$	2.0×10^{-4}
SLC22A3	6q25.3	$9.3 \times 10^{-7} - 2.0 \times 10^{-3}$	5.0×10^{-6}
SLC25A37	8p21.2	$3.0 \times 10^{-2} - 2.64 \times 10^{-1}$	5.0×10^{-6}
TET2	4q24	$6.74 \times 10^{-10} - 1.2 \times 10^{-2}$	5.0×10^{-6}
TNFSF10	3q26	$7.34 \times 10^{-5} - 2.0 \times 10^{-3}$	9.9×10^{-4}
TNRC6B	22q13	$5 \times 10^{-7} - 1.22 \times 10^{-3}$	5.0×10^{-6}
ZBTB38	3q23	2.0×10^{-8}	5.0×10^{-6}
ZNF652	17q21.32	3.4×10^{-13}	5.0×10^{-6}
AR	Xq12	1.0×10^{-8}	5.0×10^{-6}
ARL15	5p15.2	5.4×10^{-19}	5.0×10^{-6}
CCHCR1	6p21.3	3.2×10^{-8}	5.0×10^{-6}
FAM84B	8q24.21	4.0×10^{-10}	5.0×10^{-6}
FSHR	2p21-p16	5.0×10^{-8}	8.0×10^{-4}
GGCX	2p12	3.0×10^{-15}	8.0×10^{-4}
IL16	15q26.3	9.8×10^{-8}	6.4×10^{-2}
IRX4	5p15.33	3.9×10^{-18}	6.4×10^{-2}
MYEOV	11q13.2	8.30×10^{-10}	5.0×10^{-6}
SKIL	3q26	7.0×10^{-22}	5.0×10^{-6}

significantly differentially expressed novel genes (ie, genes not reported in GWAS) between tumor and control samples. Because of the large number of genes analyzed, we used a very stringent threshold to select the differentially expressed genes. After correcting for multiple hypothesis testing, we identified 200 highly significantly ($P < 10^{-6}$, FDR = 0) differentially expressed novel genes. The 200 genes and their estimates of *P* values are presented in Table S2 provided as supplementary data to this report.

Co-expression and functional analysis. To determine whether the SNP-containing and novel genes are involved in the same molecular functions, biological processes, and cellular components, we performed GO analysis. We hypothesized that the genes containing genetic variants associated with an increased risk of developing prostate cancer are functionally related with one another and with novel genes. 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 are potentially genuine even if none of the genetic variants is individually highly significant. The results of GO analysis for genes containing SNPs associated with an increased risk of developing prostate cancer and novel genes are provided in Table S3 provided as supplementary data to this report. GO analysis revealed that the genes containing SNPs associated with an increased risk of developing prostate cancer are functionally related with one another and with novel genes (Table S3). Interestingly, the genes containing SNPs with strong associations and SNPs replicated in multiple independent GWAS were found to be functionally related with genes containing SNPs with weak to moderate associations.

To further gain insights about the functional relationships of the identified genes, we performed unsupervised analysis on significantly differentially expressed SNP-containing and novel genes to identify co-expressed genes with similar patterns of expression profiles. We hypothesized that SNP-containing genes exhibit similar patterns of expression profiles or behave similarly. We further hypothesized that SNP-containing genes and novel genes have similar patterns of expression profiles. The rationale is that co-expression is correlated with functional relationships, such as physical interaction between the encoded proteins, although co-expression does not necessarily imply a causal relationship among transcript levels.[23]

The results of co-expression analysis for the 131 SNP-containing genes only are presented in Figure 1. The results of co-expression analysis for the novel genes only are presented in Figure 2. In both the cases, hierarchical clustering revealed similarities in patterns of gene expression profiles among the genes under study. Genes containing SNPs associated with an increased risk of developing prostate cancer revealed similar patterns of expression profiles with one another (Fig. 1). The patterns of gene expression profiles for the SNP-containing genes only (Fig. 1) were more spurious than the patterns of expression profiles for the novel genes only (Fig. 2). This can be partially explained by the heterogeneity inherent in GWAS data. Interestingly, genes containing SNPs with strong associations and SNPs replicated in multiple independent studies were found to interact with genes containing SNPs with weak to moderate associations (Fig. 1). This is a significant finding given that relatively few SNPs have *P* values sufficiently small to give conclusive evidence of association. Although it is conceivable that some of the SNPs with weak associations could be false positives, the presence of associated SNPs in functionally related genes with similar patterns of expression profiles gives convincing evidence that some of the SNPs and associated genes are genuine even if none of the SNPs individually is strongly associated with prostate cancer.

The similarity in patterns of expression profiles of genes containing SNPs associated with an increased risk of developing prostate cancer was a significant finding. However, it has become evident in prostate cancer research that much of the genetic risk remains unexplained. Therefore, to address this question, we performed unsupervised analysis combining SNP-containing genes with novel genes. The goal was to determine whether genes containing genetic variants associated with an increased risk of developing prostate cancer are functionally related and have similar patterns of expression profiles with novel genes. The results showing patterns of expression profiles for the combined set of genes are presented in Figure 3. Pattern recognition analysis revealed that correlated expression patterns occur between genes containing SNPs associated with an increased risk of developing prostate cancer and novel genes (Fig. 3). Interestingly, genes containing SNPs with weak to moderates associations were found to be co-expressed and functionally related with novel genes (Fig. 3).

Gene regulatory networks and biological pathways. Although pattern recognition analysis using hierarchical clustering provides a high-level overview, it is difficult to gain the broader context in which genes containing genetic variants associated with an increased risk of developing prostate cancer and novel genes operate. Genetic variants and associated genes carry out their functions through intricate molecular networks and biological pathways. To address this question, we performed network and pathway analysis to determine whether SNP-containing and novel genes interact with one another and to identify the gene regulatory networks and biological pathways enriched for genetic variants. We hypothesized that genetic variants and associated genes affect entire network states and biological pathways which in turn increase or decrease the risk of developing the disease or affect the severity of the disease.

First, we performed network analysis using SNP-containing genes only. Our working hypothesis was that genes containing genetic variants associated with an increased risk of developing prostate cancer interact with one another in gene regulatory networks. Thus, the goal of our analysis was to identify molecular networks enriched for genetic variants associated with an increased risk of developing prostate cancer.

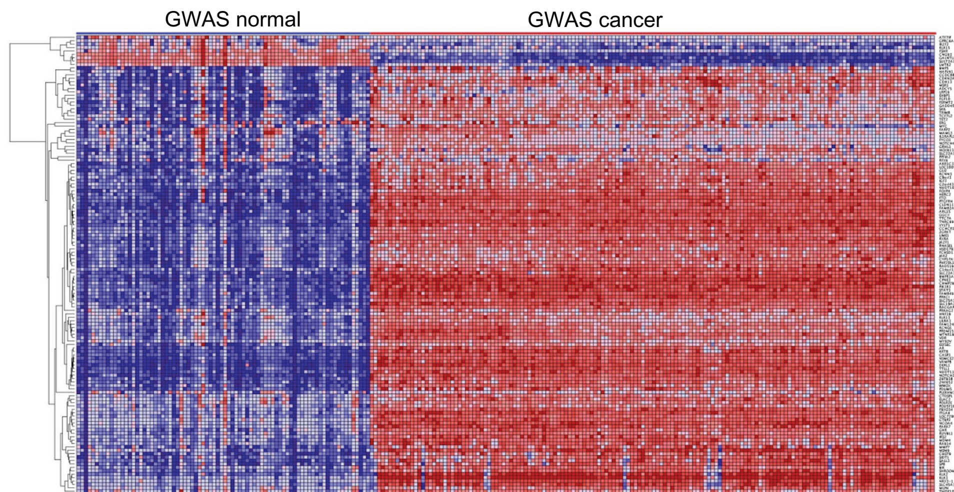

Figure 1. Patterns of gene expression profiles for the 131 significantly differentially expressed genes containing SNPs associated with an increased risk of developing prostate cancer. The heat map is based on 194 cancer patients and 40 cancer-free controls. The columns indicate samples and the rows indicate the genes. Red color indicates upregulation and blue color indicates downregulation.

The results of network analysis based on genes containing genetic variants associated with an increased risk of developing prostate cancer only are presented in Figure 4. In the networks, the edges represent interactions and the nodes represent genes. Network analysis based on SNP-containing genes only revealed five top networks with scores ranging from 22 to 34. These networks were merged and consolidated into one network (Fig. 4). Network analysis revealed that genes containing SNPs associated with an increased risk of developing prostate cancer are functionally related and interact with one another in complex gene regulatory networks confirming our hypothesis. In addition, we identified genes that were not identified by GWAS. Associated network functions included genes involved in cellular growth and proliferation, cellular development, organismal development, cell-mediated immune response, cellular movement, cellular assembly and

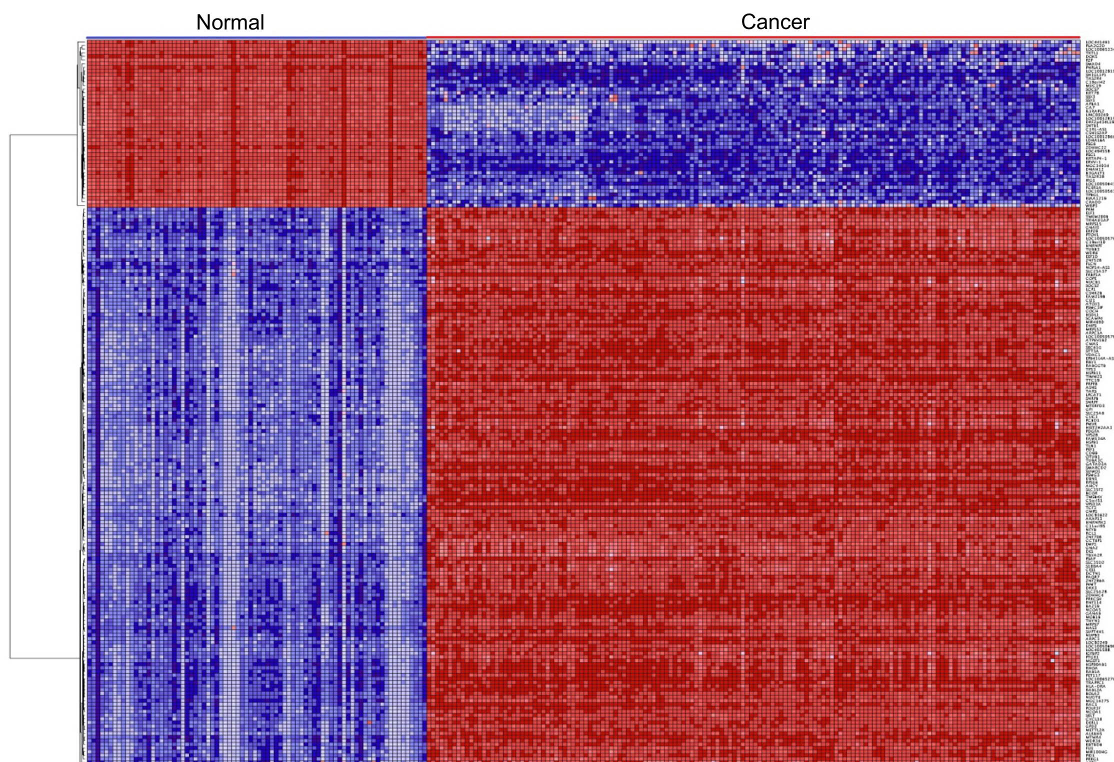

Figure 2. Patterns of gene expression profiles for the 200 significantly differentially expressed novel genes only distinguishing cancer patients from cancer-free controls. The heat map is based on 194 cancer patients and 40 cancer-free controls. The columns indicate samples and the rows indicate the genes. Red color indicates upregulation and blue color indicates downregulation.

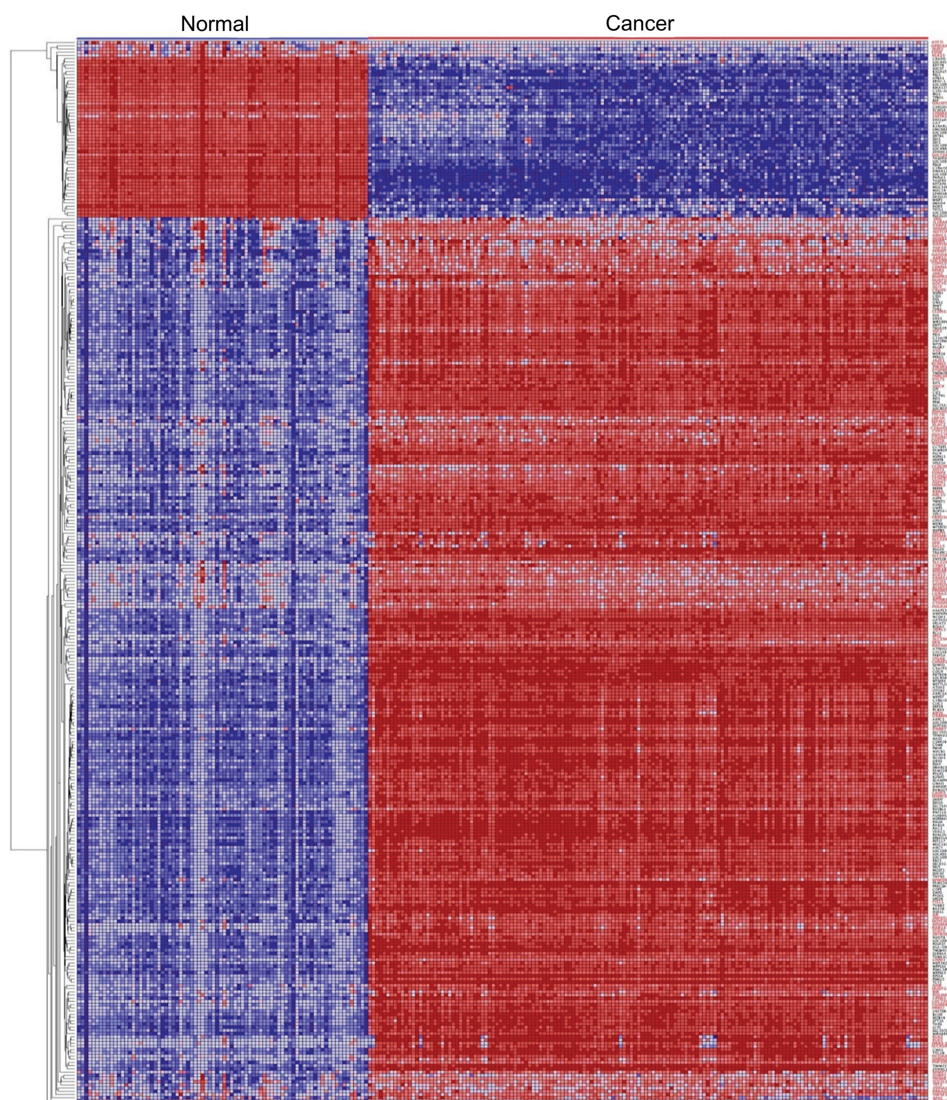

Figure 3. Patterns of gene expression profiles for the 331 significantly differentially expressed genes containing SNPs associated with an increased risk of developing prostate cancer and novel genes combined. The heat map is based on 194 cancer patients and 40 cancer-free controls. The columns indicate samples and the roles indicate the genes. Red color indicates upregulation and blue color indicates downregulation. The red and black fonts for gene symbols denote SNP-containing and novel genes, respectively.

organization, DNA replication, DNA recombination and repair, cancer, gene expression, RNA damage and repair, RNA post-transcription modification, and reproductive system disease. In addition, we identified the upstream regulators that included the androgen receptor (AR), androgen, dihydrotestosterone, *LEF1*, and testosterone.

Interestingly, genes containing SNPs with strong associations and SNPs replicated in multiple independent studies were found to be functionally related and interact with genes containing genetic variants with weak to moderate associations (Fig. 4). Among the genes containing SNPs strongly associated with prostate cancer mapping to the networks included the genes *AR*, *MSMB*, *FERMT2*, *HNF1B*, *BIK*, *KLK3*, *ITGA6*, *NKX3-1*, *JAZF1*, *NOTCH4*, *ARL15*, and *RAD51B* (Fig. 4). Genes containing genetic variants with weak to moderate associations mapped to molecular networks included the genes *KLK2*, *ERG*, *PRKCI*, *RXRA*, *NOTCH2*, *VDR*, *SULT2A1*, *PHF20L1*, *CTDSPL*, *GADD45A*, and *CPNE3* (Fig. 4).

Next, we performed network analysis combining SNP-containing and novel genes. Our working hypothesis was that genes containing genetic variants associated with an increased risk of developing prostate cancer interact with novel genes. Thus, the goal of network analysis in this case was to identify molecular networks in which the two sets of genes show functional relationships and interact with each other. The results of network analysis based on the combined set of genes are presented in Figure 5. Network analysis revealed that genes containing genetic variants associated with an increased risk of developing prostate cancer interact with novel genes in gene regulatory networks (Fig. 5). Among the genes identified in network analysis included genes involved in cancer,

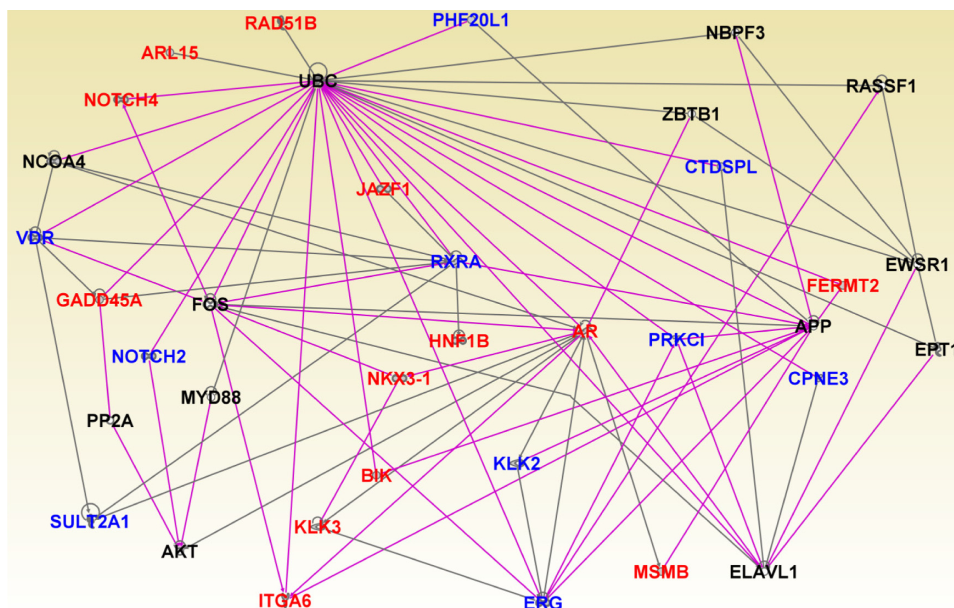

Figure 4. Gene regulatory networks obtained using only genes containing SNPs associated with an increased risk of developing prostate cancer. The nodes depict the genes, edges indicate direct interactions, genes in red color fonts contain genetic variants with strong associations and/or replicated in multiple independent studies, genes in blue color fonts contain genetic variants with moderate to weak associations. The novel genes (genes not identified by GWAS) are represented in black color fonts.

immunological disease, DNA replication, recombination and repair, cell-to-cell signaling interaction, tissue development, cell death and survival, organ morphology, drug metabolism, endocrine system development and function, lipid metabolism, and carbohydrate metabolism. In addition, we identified the upstream regulators including the *AR*, cadmium chloride, androgen, *CD437*, and *CTNNB1*; all of which have been implicated in prostate cancer.

Among the genes containing SNPs significantly associated with prostate cancer, those found to be interacting with

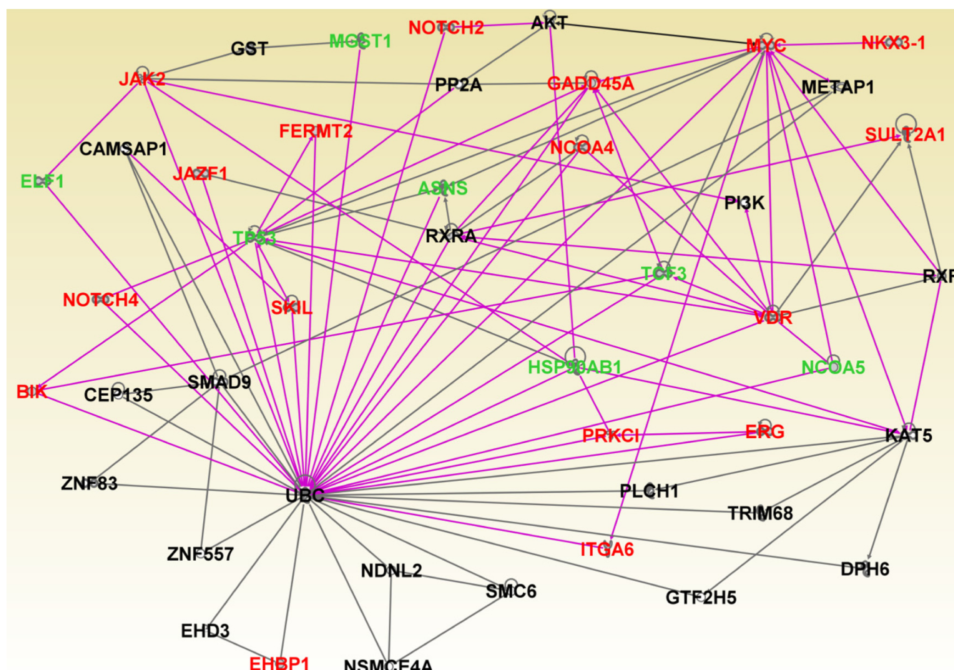

Figure 5. Gene regulatory networks obtained using genes containing SNPs associated with an increased risk of developing prostate cancer and novel genes. The nodes depict the genes, edges indicate direct interactions, genes in red color fonts contain genetic variants with strong associations and/or replicated in multiple independent studies. Novel (genes not identified by GWAS) differentially expressed genes are represented in green. Other functionally related novel genes captured through network analysis are represented in black color fonts.

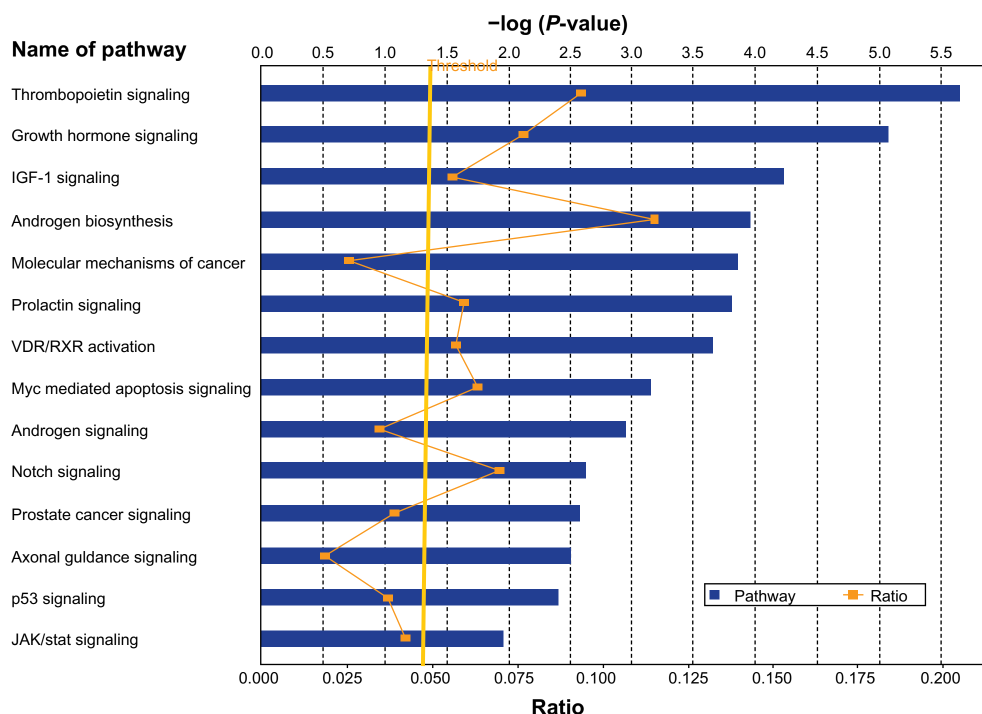

Figure 6. Biological pathways significantly enriched for genetic variants that are strongly associated with an increased risk of developing prostate cancer. Pathway analysis was based only on SNP-containing genes. The vertical thick yellow line indicates the threshold level of the *P* value on a log scale for declaring that the pathway is significantly enriched for genetic variants after correcting for multiple testing. The zigzagging orange line is the ratio of the SNP-containing genes used as the input to the total number of molecules in the specific pathway.

novel genes included *JAK2*, *NOTCH2*, *JAZF1*, *FERMT2*, *SKIL*, *NOTCH4*, *BIK*, *EHBP1*, *ITGA6*, *ERG*, *PRKCI*, *VDR*, *NOOA4*, *GADD45A*, *MYC*, *NKX3-1*, and *SULT2A1* (Fig. 5). Intriguingly, genes containing genetic variants with weak to moderate associations were found to be functionally related and interacting with novel genes in gene regulatory networks. The results of functional relationship and interactions between SNP-containing and novel genes revealed in both Figures 4 and 5 suggest that some of the actions of genes containing SNPs associated with an increased risk of developing prostate cancer may be mediated by novel genes.

To further gain insights about the broader biological context in which the genetic variants operate, we performed pathway analysis. We hypothesized that the SNP-containing and novel genes carryout their joint actions in biological pathways. Thus, the goal of pathway analysis was to identify the pathways enriched for genetic variants and to establish putative functional bridges between GWAS findings and the biological pathways. First, we performed pathway analysis using only genes containing genetic variants associated with an increased risk of developing prostate cancer. Subsequently, we repeated same analysis using a combined set of SNP-containing and novel genes.

The results of pathway analysis based on SNP-containing genes only are presented in Figure 6. Pathway analysis revealed many biological pathways significantly enriched for genetic variants. Among the identified pathways included the thrombopoeitin, growth hormone, IGF-1, androgen biosynthesis, molecular mechanisms of cancer, prolactin signaling, VDR/RXR activation, MYC-mediated apoptosis signaling, androgen signaling, NOTCH signaling, prostate cancer signaling, axonal guidance signaling, P53 signaling, and the JAK/STAT signaling pathways (Fig. 6). Interestingly, the genes containing genetic variants with strong association and the genes containing SNPs with weak to moderate association were mapped to the same biological pathways.

The results of pathway analysis based on SNP-containing and novel genes are presented in Figure 7. Path analysis revealed that genes containing genetic variants associated with an increased risk of developing prostate cancer and novel genes map to the same pathways confirming our hypothesis. Among the identified pathways included the VDR/RxR activation, prolactin, prostate cancer, thrombopoietin, IGF-1, growth hormone, molecular mechanisms of cancer, Myc-mediated apoptosis, death receptor, JAK/STAT, androgen, PI3/AKT, and apoptosis signaling pathways (Fig. 7). A close examination of the results based on GWAS information alone and the results based on the combined set of SNP-containing and novel genes revealed differences in the ranking of the identified pathways. For example, the NOTCH and P53 signaling pathways found to be among the top pathways when only GWAS information used (Fig. 6) were not among the top pathways identified in combined analysis (Fig. 7). Conversely, a novel pathway, the death receptor signaling pathway,

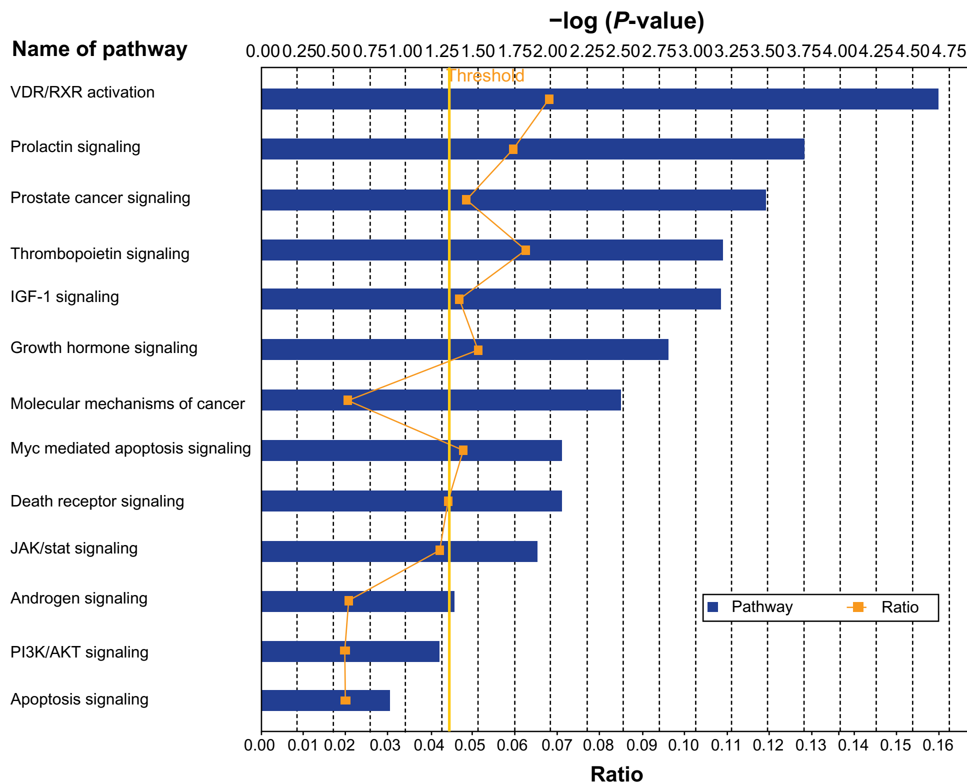

Figure 7. Biological pathways enriched for genetic variants significantly associated with an increased risk of developing prostate cancer. Pathway analysis included SNP-containing and novel genes. The vertical thick yellow line indicates the threshold level of the *P* value on a log scale for declaring that the pathway is significantly enriched for genetic variants after correcting for multiple testing. The zigzagging orange line is the ratio of the SNP-containing and novel genes used as the input to the total number of molecules in the specific pathway.

was identified in combined analysis (Fig. 7) suggesting that important biological pathways could be missed by focusing only on GWAS information. The interaction between SNP-containing and novel genes in biological pathways suggests that some of the biological activities of genes containing SNPs associated with an increased risk of developing prostate cancer may be mediated by novel genes.

The rationale is that such genes and the biological pathways they map to could serve as potential clinically actionable biomarkers. To investigate the relevance of the SNP-containing and novel genes associated with the disease state in the context of prostate cancer pathogenesis, we conducted literature mining. The goal was to determine whether any of the SNP-containing genes associated with the disease state have been implicated in prostate cancer. The rationale is that such genes and the biological pathways they map could serve as potential clinically actionable biomarkers. A summary of the results of this analysis is given below.

Interestingly, we found that many of the identified genes are involved in prostate cancer. Here, we provide a summary of our findings. One of the most prominent genes identified through network analysis was the novel gene *UBC*, which was found to be interacting with many SNP-containing as well as novel genes. The *UBC* regulates the *AR* events that influence disease progression.[29] The ubiquitin E3 ligase a ring finger protein (*RNF*)-6 has been shown to promote *AR* activity through selective modulation of co-factor recruitment such as the androgen receptor-associated (*ARA*) protein.[30] This function is enhanced in castration-resistant tumors.[31] The MDM2-mediated ubiquitylation of AR results in receptor destabilization and loss of activity.[32]

The human AR has been shown to play a critical role in the growth and differentiation of normal prostate gland as well as in the development of prostate cancer.[33] The *AR*, which was found to be overexpressed in prostate cancer in this study, is required for prostate cancer growth in all stages, including the relapsed, "androgen-independent" tumors in the presence of very low levels of androgens.[33] The cellular *AR* levels have been shown to be correlated with primary and metastatic lesions. *PTEN* loss and *AR* amplification have been associated with disease progression to lethal, metastatic castration-resistant prostate cancer.[33] *TP53* regulates critical prostate cancer pathways and is a crucial therapeutic target in prostate cancer.[33]

The human kallikreins *KLK2* and *KLK3* are used as diagnostic and prognostic markers.[34,35] The most studied of the kallikreins is *KLK3* popularly known as prostate-specific antigen (PSA), a widely used clinical tumor biomarker for detection and monitoring of prostate cancer progression.[34,35] Expression of PSA is prostate-specific, regulated by andro-

gens, and increased PSA levels may be an indication of prostate abnormalities.[34,35] The *KLK2* has very close structural homology to *KLK3*, and both genes are adjacent to each other on chromosome 19. Like the *KLK3*, the *KLK2* is androgen regulated and may have utility as a prostate cancer biomarker in conjunction with the *KLK3*, which is expressed at lower levels compared with the *KLK2* in poorly differentiated tumors.[34,35] The *NKX3-1* has been shown to be a regulator of prostate epithelial growth and differentiation.[34,35] These results indicate that AR-targeted SNP-containing genes represent potential clinically actionable biomarkers in prostate cancer.

Several biological pathways enriched for genetic variants identified in this study have been implicated in prostate cancer. The AR signaling pathway is one of the critical biological pathways associated with prostate cancer.[30,34–36] In the initial stages, prostate cancer is dependent on androgens for growth, which is the basis for androgen ablation therapy, although in most cases, prostate cancer progresses to a hormone refractory phenotype for which there is no effective therapy available at present.[33,34] Studies have also shown that AR signaling plays a major role in advanced prostate cancer, which is hormone refractory or androgen independent.[33–36] Therefore, directed therapies targeting the androgen–receptor axis may be effective.[37]

The *IGF-1* signaling pathway increases the transactivation potential of *AR*.[38] In addition, various reports have implicated the *IGF-1* signaling pathway in modulating *AR* signaling.[34,38] For example, *IGF-1* and *IGFBP-3* coordinate in prostate cancer pathogenesis.[38] Large-scale epidemiological studies have demonstrated that there is a correlation between high serum levels of *IGF-1* and low levels of *IGFBP-3*, a serum protein that regulates the binding of free *IGF*-1 to *IGF* receptor (*IGFR*) and increased the risk of developing prostate cancer.[39] The potential molecular mechanisms include *AR* phosphorylation[40] and *AR* translocation[40] or stimulation of expression of *AR* co-factors.[41] As an example, *IGF-1* has been shown to promote the formation of a complex involving *AKT*, *AR*, and *MDM2*, which results in phosphorylation-dependent ubiquitylation and degradation of AR by a proteasome-dependent mechanism.[42] The *P*53 signaling pathway modulates the genes involved in DNA damage and cell-cycle regulation, and therefore is a crucial target in the development of therapeutic strategies and early interventions in prostate cancer.[43]

However, it is worth noting that adaptation of prostate cancer cells to androgen deprivation may involve both amplification and mutations of the AR.[44] In addition to alterations of AR function, protein kinase pathways activated by peptide hormones and local growth factors are known to promote proliferation and survival of prostate cancer cells either directly or through stimulation of *AR* actions.[45] One such growth factor-initiated protein kinase signaling pathway in prostate cancer is the prolactin-Janus kinase (*JAK2*) signaling pathway[46] identified in this study. In addition to the pathways discussed above, we identified the *JAK/STAT* pathway that has been associated with the risk of developing prostate cancer.[47] The discovery of the *JAK/STAT* pathway found in this study is consistent with the literature reports.[48] Taken together, these results demonstrate that genetic variants associated with an increased risk of developing prostate cancer are likely to dysregulate entire molecular networks and biological pathways, which in turn may increase or decrease the risk of developing prostate cancer or affect the severity of the disease. The discovery of multiple pathways enriched for genetic variants suggests that pathway crosstalk may be involved.

Discussion

This research was conducted to infer the causal association between gene expression and the disease, to understand the broader context which the genetic variants operate and to establish functional bridges between GWAS findings and molecular networks and biological pathways. The unique feature of our analysis is the integration of GWAS information with an intermediate phenotype (ie, gene expression data). This study demonstrates that integrative genomics combining GWAS information with gene expression data provides a unified approach for linking GWAS information to the disease state and for the discovery of gene regulatory networks and biological pathways enriched for genetic variants.

There are several novel features in using this approach. First, the results demonstrate that the genetic variants are likely to affect entire network states and biological pathways. Specifically, the results revealed that, in the context of prostate cancer as a polygenic disease, the disease state is an emergent property of molecular networks and biological pathways that are likely to be affected by many genetic variants (both common and rare), rather than one or a few genetic variants with strong associations. Second, the approach allows identification of novel genes not identified by GWAS. Therefore, this is a powerful approach for explaining the missing variation. These are significant findings in light of the small fraction of the heritable variation explained by the variants identified thus far.

Third, the results revealed that genes containing genetic variants strongly associated with an increased risk of developing prostate cancer interact with genes containing genetic variants with weak to moderate associations and novel genes. This suggests that some of the biological activities of genes containing genetic variants with strong associations may be mediated by novel genes and genes containing genetic variants with weak to moderate associations. The results demonstrate that integrating genotype with gene expression data holds the promise of not only identifying the affected networks and biological pathways, but also providing important insights about the broader context in which the genetic variants operate. This is critical to identifying potential targets for the development of novel therapeutic strategies, early interventions, and translating GWAS discoveries into clinically actionable biomarkers to improve human health.

In the published literature, many reports using a pathway-based approach in GWAS analyses have been reported.[6,49,50] Our intention in this report was not to refute the methods of GWAS analysis, but to offer a complementary approach that provides new insights that were not provided by single-SNP GWAS analysis in prostate cancer. An important finding in this study is that genetic variants and associated genes carryout their functions through intricate gene regulatory networks and biological pathways. Identifying the genetic variants that map to biological pathways associated with prostate cancer is an excellent first step to uncovering the drivers of prostate cancer. In the published literature on GWAS, the functions of many SNPs associated with an increased risk of developing prostate cancer have not been well characterized.[2] However, as demonstrated in this study, the functions of particular pathways such as the *IGF*-1 and the AR signaling pathways have been much better characterized.[34–38]

Many dysregulated pathways identified in this study including IGF-1, AR, prostate cancer, Notch, and *P53* signaling pathways were enriched for genetic variants, suggesting that the genetic variants associated with an increased risk of developing prostate cancer act through biological pathways. This indicates that the networks and biological pathways can be used as an organizational framework for dissecting the molecular mechanisms underpinning prostate cancer development and progression. These pathways could also be used as candidates for the development of novel therapeutic and early intervention strategies. Beyond the identification of network states and biological pathways enriched for genetic variants, there is a critical and urgent need to delineate the molecular mechanisms by which the genetic variants dysregulate the network states and biological pathways. The integrative approach reported here has another novel feature, that is, it can be used to identify genes and biological pathways to prioritize for targeted massively paralleled sequencing.[2]

However, it is worth noting that, although the integrative genomics approach presented here has the promise to address many longstanding questions as demonstrated in this study, limitations of the study must be acknowledged. First, it is worth noting that our approach relies on using publicly available GWAS information and gene expression data. Thus, our analysis is subjected to all the limitations inherent in such data, including but not limited to genotyping and experimental errors, publication bias, lack of uniformity in data analysis methods, population heterogeneity and stratification, sampling errors, etc. Second, although this is a holistic approach and accounts for all the genetic variants in the genes, it provides no information about allele-specific expression. Therefore, it is difficult to discern the effects of individual SNPs on gene expression and the pathways. It is also conceivable that some of the SNPs and associated genes from GWAS may be false associations. Nevertheless, it is worth noting that our goal here was to establish the association between GWAS information and the disease state, and to gain insights about the broader context in which the genetic variants operate.

Importantly, genetic variants regulate gene expression, and previous studies have reported allele-specific expression in human populations[51–53] and correlations between risk variants and clinical parameters for genes and genetic variants in breast cancer.[54] Finally, an important consideration is that the overwhelming majority of GWAS particularly in earlier studies on prostate cancer, most of which were used in this study, were limited to populations of European ancestry, and gene expression used in this study was based on populations of European ancestry. Given that genetic susceptibility and gene expression in prostate cancer can be population specific, the variability in gene expression between and among populations, along with the fact that common genetic variants can account for differences in expression among ethnic groups.[55–57] The results found in this study cannot be over generalized. We recommend that future studies should include other racial/ethnic populations.

In conclusion, this investigation demonstrates that integrative genomics analysis is a powerful approach for associating GWAS information with the disease state, identifying novel genes, and gaining insights about the broader context in which the genetic variants associated with an increased risk of developing prostate cancer operate. The identification of functionally related genes interacting in gene regulatory networks and biological pathways enriched for genetic variants provides a framework and a critical step toward understanding the molecular mechanisms underpinning prostate cancer development and the discovery of biomarkers for the development of novel therapeutic strategies. More research is needed to understand the molecular mechanisms by which the genetic variants regulate gene expression in different populations and different disease states.

Author Contributions

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

Supplementary Data

Table S1. List of genes and SNPs associated with an increased risk of developing prostate cancer and estimates of *P* values for genes derived from gene expression data; and GWAS association *P* values for the SNPs.

Table S2. List of novel genes (genes not identified by GWAS) that were highly significantly ($P < 10^{-6}$) differentially expressed between prostate cancer and cancer-free controls and their estimates of *P* values.

Table S3. Combined list of GWAS identified and novel genes (genes not identified by GWAS) and their functional

relationships depicted by molecular function, biological component process, and cellular component in which they are involved.

REFERENCES

1. 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 USA*. 2009;106(23):9362–7.
2. Hicks C, Miele L, Koganti T, Vijayakumar S. Comprehensive assessment and network analysis of the emerging genetic susceptibility landscape of prostate cancer. *Cancer Inform*. 2013;12:175–91.
3. Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet*. 2010;86:581–91.
4. Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature*. 2009;461:218–23.
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. Peng G, Luo L, Zhu Y, Dong H, Amos CI, Xiong M. Genome-wide gene pathway analysis. *Eur J Hum Genet*. 2010;18:1045–53.
7. Calvo A, Xiao N, Kang J, et al. Alterations in gene expression profiles during prostate cancer progression: functional correlations to tumorigenicity and down-regulation of selenoprotein-P in mouse and human tumors. *Cancer Res*. 2002;62(18):5325–35.
8. Paulo P, Ribeiro FR, Santos J, et al. Molecular subtyping of primary prostate cancer reveals specific and shared target genes of different ETS rearrangements. *Neoplasia*. 2012;14(7):600–11.
9. LaTulippe E, Satagopan J, Smith A, et al. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer Res*. 2002;62(15):4499–506.
10. Chetcuti A, Margan S, Mann S, et al. Identification of differentially expressed genes in organ-confined prostate cancer by gene expression array. *Prostate*. 2001;47(2):132–40.
11. Ioannidis JP, Boffetta P, Little J, et al. Assessment of cumulative evidence on genetic associations: Interim guidelines. *Int J Epidemiol*. 2008;37:120–32.
12. Khoury MJ, Bertram L, 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.
13. Sagoo GS, Little J, Higgins JP. Systematic reviews of genetic association studies. Human Genome Epidemiology Network. *PLoS Med*. 2009;6:e28.
14. 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.
15. Liberati A, Altman DG, Tetzlaff J, 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.
16. dbSNP. Available at http://www.ncbi.nlm.nih.gov/SNP/
17. Human Genome Nomenclature Committee (HGNC) Database. Available at http://www.genenames.org/
18. Jia Z, Wang Y, Sawyers A, et al. Diagnosis of prostate cancer using differentially expressed genes in stroma. *Cancer Res*. 2011;71(7):2476–87.
19. Benjamini Y, Hochberg Y. Controlling the false discovery rat: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*. 1995;57(1):289–300.
20. Ramacher MD, Mcshane LM, Simon R. A paradigm for class prediction using gene expression profiles. *J Comput Biol*. 2002;9(3):505–11.
21. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P. GenePattern 2.0. *Nat Genet*. 2006;38(5):500–1.
22. Morrissey ER, Diaz-Uriarte R. Pomelo II: finding differentially expressed genes. *Nucleic Acids Res*. 2009;37:W581–6.
23. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Nat Acad Sci U S A*. 1998;95: 14863–8.
24. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25–9.
25. Ingenuity Pathways Analysis (IPA). *Ingenuity Pathways Analysis (IPA) system*. Redwood, CA: Ingenuity Systems, Inc. Available at http://www.ingenuity.com/
26. Broer L, Lill CM, Schuur M, et al. Distinguishing true from false positives in genomic studies: p values. *Eur J Epidemiol*. 2013;28(2):131–8.
27. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet*. 2001;29(3):306–9.
28. Galvan A, Ioannidis JP, Dragani TA. Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. *Trends Genet*. 2010;26(3):132–41.
29. Knudsen KE, Penning TM. Partners in crime: deregulation of AR activity and androgen synthesis in prostate cancer. *Trends Endocrinol Metab*. 2010;21(5):315–24.
30. Chen S, Kesler CT, Paschal BM, Balk SP. Androgen receptor phosphorylation and activity are regulated by an association with protein phosphatase 1. *J Biol Chem*. 2009;284:25576–84.
31. Xu K, Shimelis H, Linn DE, et al. Regulation of androgen receptor transcriptional activity and specificity by RNF6-induced ubiquitination. *Cancer Cell*. 2009;15:270–82.
32. Gaughan L, Logan IR, Neal DE, Robson CN. Regulation of androgen receptor and histone deacetylase 1 by Mdm2-mediated ubiquitylation. *Nucleic Acids Res*. 2005;33:13–26.
33. Kaarbø M, Klokk TI, Saatcioglu F. Androgen signaling and its interactions with other signaling pathways in prostate cancer. *Bioessays*. 2007;29(12):1227–38.
34. Zambon CF, Prayer-Galetti T, Basso D, et al. Effectiveness of the combined evaluation of KLK3 genetics and free-to-total prostate specific antigen ratio for prostate cancer diagnosis. *J Urol*. 2012;188(4):1124–30.
35. 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.
36. Heinlein CA, Chang C. Androgen receptor in prostate cancer. *Endocr Rev*. 2004;25(2):276–308.
37. 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.
38. Papatsoris AG, Karamouzis MV, Papavassiliou AG. Novel Insights into the implications of the IGF-1 network in prostate cancer. *Trends Mol Med*. 2005;11(2):52–5.
39. Chan JM, Stampfer MJ, Giovannucci E, et al. Plasma insulin-like growth factor-I and prostate cancer risk: a prospective study. *Science*. 1998;279:563–6.
40. Wu JD, Haugk K, Woodke L, Nelson P, Coleman I, Plymate SR. Interaction of IGF signaling and the androgen receptor in prostate cancer progression. *J Cell Biochem*. 2006;99:392–401.
41. Gregory CW, He B, Johnson RT, et al. A mechanism for androgen receptor-mediated prostate cancer recurrence after androgen deprivation therapy. *Cancer Res*. 2001;61:4315–9.
42. Lin HK, Wang L, Hu YC, Altuwaijri S, Chang C. Phosphorylation dependent ubiquitylation and degradation of androgen receptor by Akt require Mdm2 E3 ligase. *EMBO J*. 2002;21:4037–48.
43. De Luca P, Moiola CP, Zalazar F, Gardner K, Vazquez ES, De Siervi A. BRCA1 and P53 regulate critical prostate cancer pathways. *Prostate Cancer Prostatic Dis*. 2013;16:233–8.
44. Linja MJ, Visakorpi T. Alterations of androgen receptor in prostate cancer. *J Steroid Biochem Mol Biol*. 2004;92:255–65.
45. Feldman BJ, Feldman D. The development of androgen-independent prostate cancer. *Nat Rev Cancer*. 2001;1:34–45.
46. Dagvadorj A, Collins S, Jomain JB, et al. Autocrine prolactin promotes prostate cancer cell growth via janus kinase-2-signal transducer and activator transcription-5a/b signaling pathway. *Endocrinology*. 2007;148(7):3089–101.
47. 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.
48. Jia P, Liu Y, Zhao Z. Integrative pathway analysis of genome-wide association studies and gene expression data in prostate cancer. *BMC Syst Biol*. 2012; 6(suppl 3):513.
49. Chen X, Wang L, Hu B, Guo M, Barnard J, Zhu X. Pathway-based analysis of genome-wide association studies using supervised principal components. *Genet Epidemiol*. 2010;34:716–24.
50. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genome-wide association studies. *Am J Hum Genet*. 2007;81:1278–83.
51. Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. Allelic variation in human gene expression. *Science*. 2002;297:1143.
52. Buckland PR. Allele-specific gene expression differences in humans. *Hum Mol Genet*. 2004;13(2):R255–60.
53. Palacios R, Gazave E, Goñi J, et al. Allele-specific gene expression is wide-spread across the genome and biological processes. *PLoS One*. 2009;4(1):e4150.
54. Meyer KB, Maia AT, O'Reilly M, et al. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol*. 2008;6(5):e108.
55. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG. Common genetic variants account for differences in gene expression among ethnic populations. *Nat Genet*. 2007;39(2):226–31.
56. Zhang W, Duan S, Kistner EO, et al. Evaluation of genetic variation contributing to differences in gene expression between populations. *Am J Hum Genet*. 2008;82:631–40.
57. Stranger BE, Nica AC, Forrest MS, et al. Population genomics of human gene expression. *Nat Genet*. 2007;39(10):1217–24.