An Integrative Genomics Approach for Associating Genome-Wide Association Studies Information With Localized and Metastatic Prostate Cancer Phenotypes

Biomarker Insights
Volume 12: 1–14

DOI: 10.1177/1177271917695810

Chindo Hicks[1], Ritika Ramani[2], Oliver Sartor[3], Ritu Bhalla[4], Lucio Miele[1], Zodwa Dlamini[5] and Njabulo Gumede[5]

[1]Department of Genetics, Louisiana State University Health Sciences Center New Orleans, New Orleans, LA, USA. [2]Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. [3]Department of Medicine, Tulane University, New Orleans, LA, USA. [4]Department of Pathology, Louisiana State University Health Sciences Center New Orleans, New Orleans, LA, USA. [5]Department of Biology, Mangosuthu University of Technology, Durban, South Africa.

ABSTRACT: High-throughput genotyping has enabled discovery of genetic variants associated with an increased risk of developing prostate cancer using genome-wide association studies (GWAS). The goal of this study was to associate GWAS information of patients with primary organ–confined and metastatic prostate cancer using gene expression data and to identify molecular networks and biological pathways enriched for genetic susceptibility variants involved in the 2 disease states. The analysis revealed gene signatures for the 2 disease states and a gene signature distinguishing the 2 patient groups. In addition, the analysis revealed molecular networks and biological pathways enriched for genetic susceptibility variants. The discovered pathways include the androgen, apoptosis, and insulinlike growth factor signaling pathways. This analysis established putative functional bridges between GWAS discoveries and the biological pathways involved in primary organ–confined and metastatic prostate cancer.

KEYWORDS: Integrative genomics analysis, GWAS, primary metastatic prostate cancer

RECEIVED: November 3, 2014. **ACCEPTED:** February 5, 2017.

PEER REVIEW: Six peer reviewers contributed to the peer review report. Reviewers' reports totaled 1429 words, excluding any confidential comments to the academic editor.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Chindo Hicks, Department of Genetics, Louisiana State University Health Sciences Center New Orleans, 533 Bolivar Street, New Orleans, LA 70012, USA. Email: chick3@lsuhsc.edu

Introduction

Prostate cancer is the most common noncutaneous malignancy and the second leading cause of cancer-related death in men older than 40 years in the United States.[1] In 2015, an estimated 220 000 men were diagnosed with prostate cancer, and 27 540 of these died of the disease.[1] Although there has been significant progress in the treatment of prostate cancer, several challenges persist such as the means to match patients with targeted therapies and understanding the molecular mechanisms underlying genetic predisposition to tumor aggressiveness. Nearly all the mortalities from prostate cancer are due to metastatic disease, typically through tumors that evolve to be hormone refractory or castrate resistant.[2] Advances in microarray technology over the past decade have enabled molecular classification of subtypes of prostate cancer[3,4] and discovery of potential clinically actionable biomarkers.[5] However, this approach has been unsuccessful in determining which genes are potential drivers of prostate cancer aggressiveness as opposed to being consequences of the disease state.

Over the past decade, considerable efforts have been directed at discovering the genetic variants associated with an increased risk of developing prostate cancer using genome-wide association studies (GWAS).[6,7] These studies have revealed the genetic variants, primarily single-nucleotide polymorphisms (SNPs) and genes associated with an increased risk of developing prostate cancer.[6,7] However, GWAS information has not been leveraged and integrated with gene expression data from primary organ–confined and metastatic prostate cancers to identify molecular signatures predictive of disease aggressiveness. A limited number of GWAS have reported genetic susceptibility variants associated with tumor aggressiveness,[8–16] prostate cancer progression and mortality,[17,18] and survival.[19] However, most of the GWAS-identified loci are not phenotype specific, and their association with clinical phenotypes remains poorly understood. This limited progress must be viewed against the recognition that most GWAS on prostate cancer were not designed to identify genetic variants associated with tumor-specific clinical phenotypes as they were largely focused on men diagnosed with prostate cancer irrespective of the disease state. This knowledge gap has hampered translation of GWAS discoveries into clinically actionable biomarkers to improve human health.

Integrative genomics that combines GWAS information with gene expression data has the promise of associating genetic susceptibility variants and genes with the clinical phenotypes.[7] Indeed, several groups including ours have reported integration of GWAS information with gene expression data in prostate cancer.[20–25] However, the association of GWAS discoveries with primary organ–confined and metastatic disease has not been reported. Similarly, discovery of molecular networks and biological pathways enriched for genetic susceptibility variants involved

in the 2 disease states has not been reported. The objectives of this study were manifold: (1) to investigate whether genes containing genetic susceptibility variants associated with and increased risk of developing prostate cancer are associated with primary organ–confined and metastatic tumors, (2) to discover a gene signature enriched for genetic susceptibility variants which distinguishes the 2 patient groups, (3) to determine whether the genes containing genetic susceptibility variants are functionally related and involved in similar biological processes, and (4) to identify molecular networks and biological pathways enriched for genetic variants which drive the 2 clinical phenotypes. Throughout this investigation, we have defined SNPs identified from GWAS as genetic susceptibility variants and used the genes containing these genetic variants as the units of association.

Materials and Methods

Genetic susceptibility variants and associated genes

Genetic variants and genes used in this study were derived from publicly available data obtained from published reports on GWAS and the websites hosting supplementary data for the respective reports.[7,22] Data collection was based on the guidelines proposed by the Human Genome Epidemiology Network for systematic review of genetic associations.[26–30] The details pertaining collection of data used in this study have been reported in our previous published reports.[7,22] Here, we provide a brief but detailed description of the data used in this study.

We reviewed a total of 140 published reports on GWAS. The reports were screened by title, abstract, and full-text review to identify the studies meeting our eligibility criteria. After screening, 100 studies that met our eligibility criteria were selected and subjected to further detailed review. 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 (ie, studies containing <500 subjects in cases and controls). For the remaining 100 studies used in this study, they were considered eligible if they met the following criteria: (1) must have been based on a case-control study design using unrelated individuals, (2) publications must have been of full length and published in peer-reviewed journals or online in English language before July 2015, (3) prostate cancer must have been diagnosed by histological examination, (4) the sample sizes must be more than 500 for the cases and more than 500 for the controls, (5) the study must have provided sufficient information such that genotype frequencies for both prostate cancer and controls can be discerned without ambiguity, and (6) the studies must have used the appropriate and recommended statistical methods to infer the associations by taking into account the covariates and accounting for population structure.[26]

We manually extracted the information from the 100 studies meeting our eligibility criteria and the accompanying Web sites containing supplementary data. The extracted information included SNP identification number (rs-ID); evidence of association as determined by the *P* value; a composite of strong ($P \leqslant 10^{-6}$), moderate ($P = 10^{-4}$–10^{-5}), and weak association ($P = 10^{-2}$–10^{-3}); gene name; and associated chromosome position to which the SNPs map as determined by the dbSNP database[31] and the Human Genome Nomenclature database.[32] This search yielded more than 400 SNPs mapped to 172 genes from a population of more than 350 000 cases and more than 350 000 controls. Table SA (Supplementary Material) provides information about the genetic variants and references or published reports from which they were derived.

A concern with GWAS studies is publication bias (a bias that occurs when only the most significant SNPs are reported, also known as "the winner curse"). To address publication bias, we included genes containing genetic variants with strong, moderate, and weak associations. The premise is that the presence of genetic variants with strong, moderate, and weak associations in genes of similar biological functions is likely to give a degree of confidence that the associations are likely genuine and could potentially have functional impact on clinical phenotypes.

Gene expression data

We used publicly available gene expression data generated from tumor and control samples derived from the white women. The data were downloaded from the National Center for Biotechnology Information's Gene Expression Omnibus under accession numbers GSE1431 and GSE6604.[33–35] The methods regarding experimental design, sample preparations, and data processing have been described by the data originators.[33–35] Briefly, the data included a total of 194 samples distributed as follows: primary organ–confined tumors, N = 88; metastatic tumors, N = 25; and cancer-free controls, N = 81. These tissues were collected between 2002 and 2004. The protocols used for sample collection have been fully described by the data originators.[33–35] All the data were processed on the Affymetrix platform using the Human GeneChip U95A and 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. The final data matrix consisted of expression profiles of ~12 000 probes. The probes were mapped to gene names using the batch query in the NETAFX an Affymetrix database which contains gene symbols.

The U95A Chip represents about 9000 unique genes. Of the 172 genes containing genetic variants associated with an increased risk of developing prostate cancer, 89 genes were represented on the Chip. As a result of limited Chip capacity, some genes containing SNPs associated with an increased risk of developing prostate cancer were not represented on the Chip. Genes not represented on the Chip were not evaluated for gene expression, but instead were evaluated using gene ontology (GO) information,[36] network, and pathway analysis to assess their functional relationships with genes represented on the Chip as described in the "Data analysis" section.

Indeed, many data sources exist in the public domain, including data derived from serum samples and tissue microarrays. However, because gene expression can be tissue, time, and platform specific, and because of the lack of clinical information for many existing publicly available data sets, our analysis focused on using gene expression data generated from primary organ–confined and metastatic tumor samples generated using the Affymetrix platform and U95A Human Chip only, a limitation that we readily acknowledge.

Data analysis

We compared gene expression levels between patients diagnosed with primary organ–confined tumors and matched controls, between patients diagnosed with metastatic prostate cancer and matched control samples, and between patients with organ-confined tumors and patients with metastatic tumors. Significant differences in gene expression levels between cases and controls and between the 2 disease states were assessed using a permutation *t* test as implemented in Pomelo[37] and GenePattern[38] software packages. Due to small sample sizes for metastatic prostate cancer, we did not partition the data into test and validation sets as such an approach would lead to bias resulting from sampling errors. Instead, we used the leave-one-out cross-validation procedure as our prediction and validation model to identify genes with predictive power.[39] This approach has been used successfully in gene expression data analysis with limited sample size to eliminate bias.[39] We used the false discovery rate (FDR) procedure[40] to correct for multiple hypothesis testing. Genes were ranked based on the *P* values and the FDR, and highly significantly differentially expressed genes were selected for each comparison.

We performed hierarchical clustering using GenePattern[38] to identify genes with similar patterns of expression profiles for primary organ–confined and metastatic disease and a gene signature distinguishing the 2 disease states. Prior to clustering, gene expression data were normalized using the median normalization, standardized and centered.[41] We performed GO analysis to gain insights about the molecular functions and biological processes in which the genes containing genetic variants and other genes not identified by GWAS are involved. We performed additional analysis using the Ingenuity Pathway Analysis (IPA)[42] to identify molecular networks and biological pathways enriched for genetic variants.

Results

We investigated whether genes containing genetic susceptibility variants are associated with primary organ–confined and metastatic tumors and are differentially expressed between the 2 disease states. We further conducted investigations to identify molecular networks and biological pathways enriched for genetic susceptibility variants in each disease state. This section describes our findings.

Associating genes containing genetic variants with primary tumors

To investigate whether genes containing genetic susceptibility variants associated with prostate cancer are associated with primary organ–confined tumors, we compared gene expression levels between patients with primary organ–confined tumors and matched control samples. After correcting for multiple hypothesis testing, the analysis produced a signature of 53 significantly ($P < .05$) differentially expressed genes associated with primary organ–confined tumors. Thirty-two genes were highly significantly ($P < 10^{-5}$) differentially expressed and predictive of primary organ–confined tumors (Table 1). The list included the genes *TERT, BIK, POLR2E, KLK3, LILRA, CTBBP2*, *ZNF652, FSHR, PRPH, MSMB, LMTK2*, and *CCHCR1* containing SNPs strongly ($P < 10^{-8}$) associated with an increased risk of developing prostate cancer (Table 1). Interestingly, the list included the genes *PARK7, TNFSF10, RUVBL1, GLI2*, and *KRT8* containing genetic variants with moderate association and the genes *MYC, IGF2, STAT3, TTLL1, HOXB13, KLK13, KLK2, KCNQ1, RAB14, ERG, VAMP8, PRKCI, INS*, and *TH* containing SNPs with weak associations (Table 1). Most notably, the analysis produced 4 genes, *TERT, KLK3, TNFS10*, and *MSMB*, containing genetic susceptibility variants associated with tumor aggressiveness. A complete list of genes containing SNPs associated with an increased risk of developing prostate cancer is presented in Table S1 provided as supplementary data to this report. Additional analysis on the whole data set revealed a signature of 100 highly significantly (10^{-6}) differentially expressed genes not identified by GWAS.

Associating genes containing genetic variants with metastatic tumors

To investigate the association between GWAS information and metastatic disease, we compared gene expression levels between patients with metastatic tumors and matched control samples. The analysis produced a signature of 49 genes significantly ($P < .05$) associated with metastatic disease. Twenty-nine genes were highly significantly ($P < 10^{-5}$) associated with metastatic prostate cancer (Table 2). The list included the genes *AR, MSMB, FERMT2, PDLIM5, POU5F1, WWOX, CTBP2, C9ORF3, ITGA6, EHBP1, KLK3, FSHR, IL16*, and *ZNF652* containing SNPs strongly associated with prostate cancer (Table 2). The list also included the genes *PARK7, SLC19A2* containing SNPs with moderate association, and the genes *VDR, PRKCI, CDKN2A, BMPR1A, CPNE3, TTLL1, PIK3R1, INS, AKR1C3, SKIL, TCF7L2*, and *IRS2* with weak associations (Table 2).

Interestingly, the list included 5 genes, *MSMB, AR, POU5F1, WWOX*, and *KLK3*, containing genetic variants associated with aggressive prostate cancer. A complete list of genes containing SNPs associated with an increased risk of developing prostate cancer is presented in Table S2 provided

Table 1. List of the top 32 most highly significantly differentially expressed genes containing SNPs associated with an increased risk of developing prostate cancer that were found to be associated with primary organ–confined prostate cancer.

GENE SYMBOL	CHROMOSOME POSITION	SNP-ID	GWAS *P* VALUE	GE *P* VALUE	FDR
MYC	8q24	rs10090154	4.00×10^{-2}	5.00×10^{-6}	1.85×10^{-5}
IGF2	11p15.5	rs7127900	3.06×10^{-2}	5.00×10^{-6}	1.85×10^{-5}
KRT8	12q13.13	rs4919743	3.64×10^{-4}	5.00×10^{-6}	1.85×10^{-5}
TERT	5p15.33	rs2242652	2.70×10^{-24}	5.00×10^{-6}	1.85×10^{-5}
STAT3	17q21	rs744166	3.00×10^{-2}	5.00×10^{-6}	1.85×10^{-5}
BIK	22q13.31	rs5759167	1.30×10^{-12}	5.00×10^{-6}	1.85×10^{-5}
POLR2E	19p13.3	rs3787016	7.22×10^{-7}	5.00×10^{-6}	1.85×10^{-5}
TTLL1	22q13.1	rs5759167	3.01×10^{-3}	5.00×10^{-6}	1.85×10^{-5}
KLK3	19q13.41	rs2735839	6.45×10^{-37}	5.00×10^{-6}	1.85×10^{-5}
LILRA3	19q13.4	rs103294	5.34×10^{-16}	5.00×10^{-6}	1.85×10^{-5}
HOXB13	17q21.32	rs8556	6.00×10^{-3}	5.00×10^{-6}	1.85×10^{-5}
KLK13	19q13.33	rs2736433	5.00×10^{-2}	5.00×10^{-6}	1.85×10^{-5}
KLK2	19q13.33	rs2735839	9.00×10^{-3}	5.00×10^{-6}	1.85×10^{-5}
KCNQ1	11p15.5	rs231362	1.00×10^{-2}	5.00×10^{-6}	1.85×10^{-5}
MTNR1B	11q21-q22	rs10830963	3.00×10^{-2}	5.00×10^{-6}	1.85×10^{-5}
GLI2	2q14	rs11122834	5.00×10^{-6}	5.00×10^{-6}	1.85×10^{-5}
CTBP2	10q26.13	rs4962416	2.70×10^{-8}	5.00×10^{-6}	1.85×10^{-5}
RAB14	9q32-q34.11	rs942152	2.00×10^{-3}	5.00×10^{-6}	1.85×10^{-5}
ZNF652	17q21.32	rs7210100	3.00×10^{-13}	5.00×10^{-6}	1.85×10^{-5}
FSHR	2p21-p16	rs2268363	5.00×10^{-8}	5.00×10^{-6}	1.85×10^{-5}
ERG	21q22.3	rs2836370	1.00×10^{-2}	5.00×10^{-6}	1.85×10^{-5}
RUVBL1	3q21	rs7641133	1.0×10^{-4}	5.00×10^{-6}	1.85×10^{-5}
VAMP8	2p12-p11.2	RS10187424	5.00×10^{-2}	5.00×10^{-6}	1.85×10^{-5}
PRPH	12q12-q13	rs10875943	6.9×10^{-12}	5.00×10^{-6}	1.85×10^{-5}
PRKCI	3q26.3	rs4955720	7.00×10^{-3}	1.00×10^{-5}	3.56×10^{-5}
INS	11p15.5	rs7127900	3.06×10^{-3}	1.50×10^{-5}	5.13×10^{-5}
TNFSF10	3q26	rs3774315	7.34×10^{-5}	2.00×10^{-5}	6.59×10^{-5}
MSMB	10q11.2	rs10993994	8.70×10^{-29}	3.00×10^{-5}	9.21×10^{-5}
LMTK2	7q22.1	rs6465657	1.10×10^{-9}	3.00×10^{-5}	9.21×10^{-5}
CCHCR1	6p21.3	rs130067	3.20×10^{-8}	4.50×10^{-5}	1.33×10^{-4}
TH	11p15.5	rs7127900	3.06×10^{-3}	8.50×10^{-5}	2.44×10^{-4}
PARK7	1p36.23	rs6703670	9.09×10^{-4}	9.00×10^{-5}	2.50×10^{-4}

Abbreviations: FDR, false discovery rate; GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism; GE, gene expression.

as supplementary data to this report. To determine whether any of the associations discovered using gene expression overlap between the 2 disease states investigated, we examined their *P* values. Seven genes (*PRKCI*, *MSMB*, *CTBP2*, *INS*, *PARK7*, *KLK3*, and *ZNF652*) were strongly associated with both primary organ–confined and metastatic disease. Additional

Table 2. List of the top 29 most highly significantly differentially expressed genes containing SNPs associated with an increased risk of developing prostate cancer that were found to be associated with metastatic prostate cancer.

GENE	CHROMOSOME	SNP-ID	GWAS *P* VALUE	UNADJ. *P*	FDR_INDEP
VDR	12q12-q14	rs7975128	2.00×10^{-2}	1.00×10^{-7}	1.00×10^{-7}
AR	Xq12	rs5919432	1.20×10^{-8}	1.00×10^{-7}	1.00×10^{-7}
PRKCI	3q26.3	rs4955720	7.00×10^{-3}	1.00×10^{-7}	1.00×10^{-7}
CDKN2A	9p21	rs10811661	4.00×10^{-2}	1.00×10^{-7}	1.00×10^{-7}
MSMB	10q11.2	rs10993994	8.70×10^{-29}	1.00×10^{-7}	1.00×10^{-7}
FERMT2	14q22.1	rs8008270	1.78×10^{-14}	1.00×10^{-7}	1.00×10^{-7}
PDLIM5	4q22	rs17021918	4.00×10^{-14}	1.00×10^{-7}	1.00×10^{-7}
TNRC6B	22q13	rs7291691	5.61×10^{-5}	1.00×10^{-7}	1.00×10^{-7}
BMPR1A	8q24	rs11597689	3.00×10^{-2}	1.00×10^{-7}	1.00×10^{-7}
POU5F1	6p21.33	rs6983267	7.50×10^{-6}	1.00×10^{-7}	1.00×10^{-7}
CPNE3	8q21	rs4961199	2.79×10^{-2}	1.00×10^{-7}	1.00×10^{-7}
WWOX	16q23.3-q24.1	rs11150069	9.43×10^{-6}	1.00×10^{-7}	1.00×10^{-7}
CTBP2	10q26.13	rs4962416	2.70×10^{-8}	1.00×10^{-7}	1.00×10^{-7}
TTLL1	22q13.1	rs5759167	3.01×10^{-3}	1.00×10^{-7}	1.00×10^{-7}
C9orf3	9q22	rs3802458	4.00×10^{-6}	1.00×10^{-7}	1.00×10^{-7}
PIK3R1	5q13.1	rs13156223	4.00×10^{-2}	2.00×10^{-7}	9.00×10^{-7}
INS	11p15.5	rs7127900	3.06×10^{-3}	4.00×10^{-7}	2.10×10^{-6}
AKR1C3	10p15-p14	rs4881400	3.00×10^{-2}	6.00×10^{-7}	3.10×10^{-6}
ITGA6	2q31.1	rs12621278	9.00×10^{-23}	8.00×10^{-7}	4.00×10^{-6}
SKIL	3q26	RS10936632	5.00×10^{-2}	3.10×10^{-6}	1.36×10^{-5}
EHBP1	2p15	rs721048	8.00×10^{-9}	3.10×10^{-6}	1.36×10^{-5}
PARK7	1p36.23	rs6703670	9.09×10^{-4}	3.60×10^{-6}	1.51×10^{-5}
KLK3	19q13.41	rs2735839	6.45×10^{-37}	7.00×10^{-6}	2.76×10^{-5}
FSHR	2p21-p16	rs2268363	5.00×10^{-8}	1.07×10^{-5}	4.05×10^{-5}
IL16	15q26.3	rs7175701	9.80×10^{-8}	1.54×10^{-5}	5.60×10^{-5}
ZNF652	17q21.32	rs7210100	3.00×10^{-13}	2.13×10^{-5}	7.46×10^{-5}
TCF7L2	10q25.3	rs7903146	9.00×10^{-3}	2.27×10^{-5}	7.66×10^{-5}
SLC19A2	1q23.3	rs3765227	1.26×10^{-4}	4.16×10^{-5}	1.35×10^{-4}
IRS2	13q34	rs7986346	6.00×10^{-3}	7.17×10^{-5}	2.25×10^{-4}

Abbreviations: FDR, false discovery rate; GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism.

analysis on the whole data set produced a signature of 100 highly significantly (10^{-6}) differentially expressed genes not identified by GWAS.

Differences in gene expression levels between primary and metastatic tumors

Some primary organ–confined tumors progress to metastatic disease. The molecular mechanisms underlying disease progression are poorly understood. To identify a signature of genes containing genetic susceptibility variants which distinguish patients with primary organ–confined tumors from patients with metastatic disease, we compared gene expression levels between the 2 patient groups. After correcting for multiple hypothesis testing, this analysis produced a signature of 56 significantly ($P < 0.05$) differentially expressed genes. Thirty genes were highly significantly differentially expressed (Table 3). The list included the genes *MSMB*, *AR*, *C9ORF3*, *WWOX*,

Table 3. List of the top 30 most highly significantly differentially expressed genes between primary organ–confined prostate cancer and metastatic prostate cancer.

GENE NAME	CHROMOSOME	SNP-ID	GWAS *P* VALUE	UNADJ. *P*	FDR_INDEP
BMPR1A	8q24	rs11597689	3.00×10^{-2}	5.00×10^{-6}	1.78×10^{-5}
MSMB	10q11.2	rs10993994	8.70×10^{-29}	5.00×10^{-6}	1.78×10^{-5}
TTLL1	22q13.1	rs5759167	3.01×10^{-3}	5.00×10^{-6}	1.78×10^{-5}
AR	Xq12	rs5919432	1.20×10^{-8}	5.00×10^{-6}	1.78×10^{-5}
C9orf3	9q22	rs3802458	4.00×10^{-6}	5.00×10^{-6}	1.78×10^{-5}
PRKCI	3q26.3	rs4955720	7.00×10^{-3}	5.00×10^{-6}	1.78×10^{-5}
MYC	8q24	rs10090154	4.00×10^{-2}	5.00×10^{-6}	1.78×10^{-5}
WWOX	16q23.3-q24.1	rs11150069	9.43×10^{-6}	5.00×10^{-6}	1.78×10^{-5}
PIK3R1	5q13.1	rs13156223	4.00×10^{-2}	5.00×10^{-6}	1.78×10^{-5}
KRT8	12q13.13	rs4919743	3.64×10^{-4}	5.00×10^{-6}	1.78×10^{-5}
FERMT2	14q22.1	rs8008270	1.78×10^{-14}	5.00×10^{-6}	1.78×10^{-5}
BIK	22q13.31	rs5759167	1.30×10^{-12}	5.00×10^{-6}	1.78×10^{-5}
ZNF652	17q21.32	rs7210100	3.00×10^{-13}	5.00×10^{-6}	1.78×10^{-5}
CTBP2	10q26.13	rs4962416	2.70×10^{-8}	5.00×10^{-6}	1.78×10^{-5}
PDLIM5	4q22	rs17021918	4.00×10^{-14}	5.00×10^{-6}	1.78×10^{-5}
EHBP1	2p15	rs721048	8.00×10^{-9}	5.00×10^{-6}	1.78×10^{-5}
PARK7	1p36.23	rs6703670	9.09×10^{-4}	5.00×10^{-6}	1.78×10^{-5}
MMP7	11q21-q22	rs11568818	1.56×10^{-11}	5.00×10^{-6}	1.78×10^{-5}
RAB14	9q32-q34.11	rs942152	2.00×10^{-3}	5.00×10^{-6}	1.78×10^{-5}
HOXB13	17q21.32	rs8556	6.00×10^{-3}	5.00×10^{-6}	1.78×10^{-5}
GADD45A	1p31.2	rs520820	2.00×10^{-2}	5.00×10^{-6}	1.78×10^{-5}
NCOA4	10q11.2	rs10761581	2.00×10^{-3}	5.00×10^{-6}	1.78×10^{-5}
BMP5	6p12.1	rs3734444	3.00×10^{-2}	5.00×10^{-6}	1.78×10^{-5}
GHR	5p14-p12	rs2940919	3.77×10^{-4}	5.00×10^{-6}	1.78×10^{-5}
VAMP8	2p12-p11.2	RS10187424	5.00×10^{-2}	5.00×10^{-6}	1.78×10^{-5}
IGF2	11p15.5	rs7127900	3.06×10^{-3}	1.00×10^{-5}	3.30×10^{-5}
POU5F1	6p21.33	rs6983267	7.50×10^{-6}	1.00×10^{-5}	3.30×10^{-5}
LILRA3	19q13.4	rs103294	5.34×10^{-16}	4.00×10^{-5}	1.30×10^{-5}
VDR	12q12-q14	rs7975128	2.00×10^{-2}	6.00×10^{-5}	1.80×10^{-5}
GLI2	2q14	rs11122834	5.00×10^{-6}	8.00×10^{-5}	2.40×10^{-5}

Abbreviations: FDR, false discovery rate; GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism.

FERMT2, *BIK*, *ZNF652*, *CTBP2*, *PDLIM5*, *EHBP1*, *MMP7*, *POU5F1*, *LILRA3*, and *GLI2* containing SNPs strongly associated with the risk of developing prostate cancer (Table 3). Intriguingly, the list included the genes *GHR*, *KRT8*, and *PARK7* containing SNPs with moderate association and genes *BMPR1A*, *TTLL1*, *PRKCI*, *MYC*, *PIK3R1*, *RAB14*, *HOXB13*, *GADD45A*, *NCOA4*, *BMP5*, *VAMP8*, *IGF*, and *VDR* containing SNPs with weak association (Table 3). Among the identified genes included *MSMB*, *AR*, *WWOX*, and *KLK3* containing SNPs associated with aggressive tumors. A complete list of genes containing SNPs associated with an increased risk of developing prostate cancer is presented in Table S3 provided as

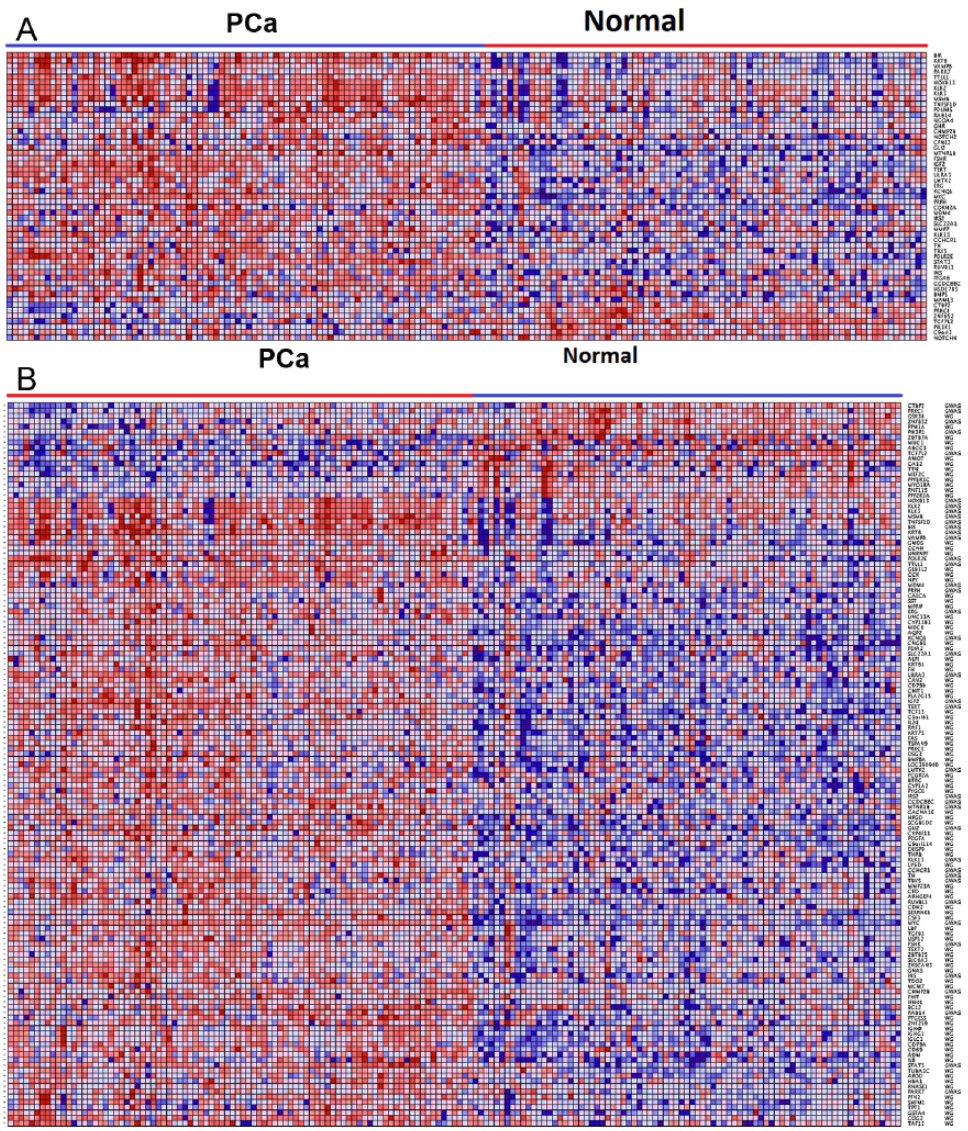

Figure 1. (A) Patterns of expression profiles for the 53 genes containing genetic variants strongly associated with an increased risk of developing prostate cancer evaluated in primary organ–confined prostate cancer (PCa) and control or normal samples. Patients are represented in columns and genes in rows. Red and blue colors indicate upregulation and downregulation, respectively. (B) Patterns of expression profiles for the 53 genes containing genetic variants strongly associated with an increased risk of developing prostate cancer and the 100 non–genome-wide association studies genes evaluated in PCa prostate cancer and control or normal samples. Patients are represented in columns and genes in rows. Red and blue colors indicate upregulation and downregulation, respectively.

supplementary data to this report. Further analysis on the whole data set produced a signature of 100 highly significantly (10^{-6}) differentially expressed genes not identified by GWAS.

Patterns of gene expression profiles in primary and metastatic tumors

Having established the association between genes containing genetic susceptibility variants with clinical phenotypes, we performed hierarchical clustering and GO analysis. We sought to investigate whether the discovered GWAS genes have similar patterns of expression profiles and are functionally related to one another and to non-GWAS genes. This analysis was performed for each disease state and between the 2 disease states.

Figure 1A shows the patterns of gene expression profiles for the top 53 genes containing genetic susceptibility variants associated with primary organ–confined tumors. The analysis revealed similarities in patterns of gene expression profiles and functional relationships among genes containing genetic susceptibility variants (Figure 1A). We discovered a cluster of 46 genes upregulated in tumors; among them, genes *KLK3, MSMB, TNFSF10, TERT*, and *KLK2* containing genetic variants are associated with tumor aggressiveness. Combined analysis of GWAS identified and non-GWAS genes revealed similarities in patterns of gene expression profiles between the 2 sets of genes (Figure 1B).

The patterns of gene expression profiles for the top 29 genes containing genetic susceptibility variants strongly

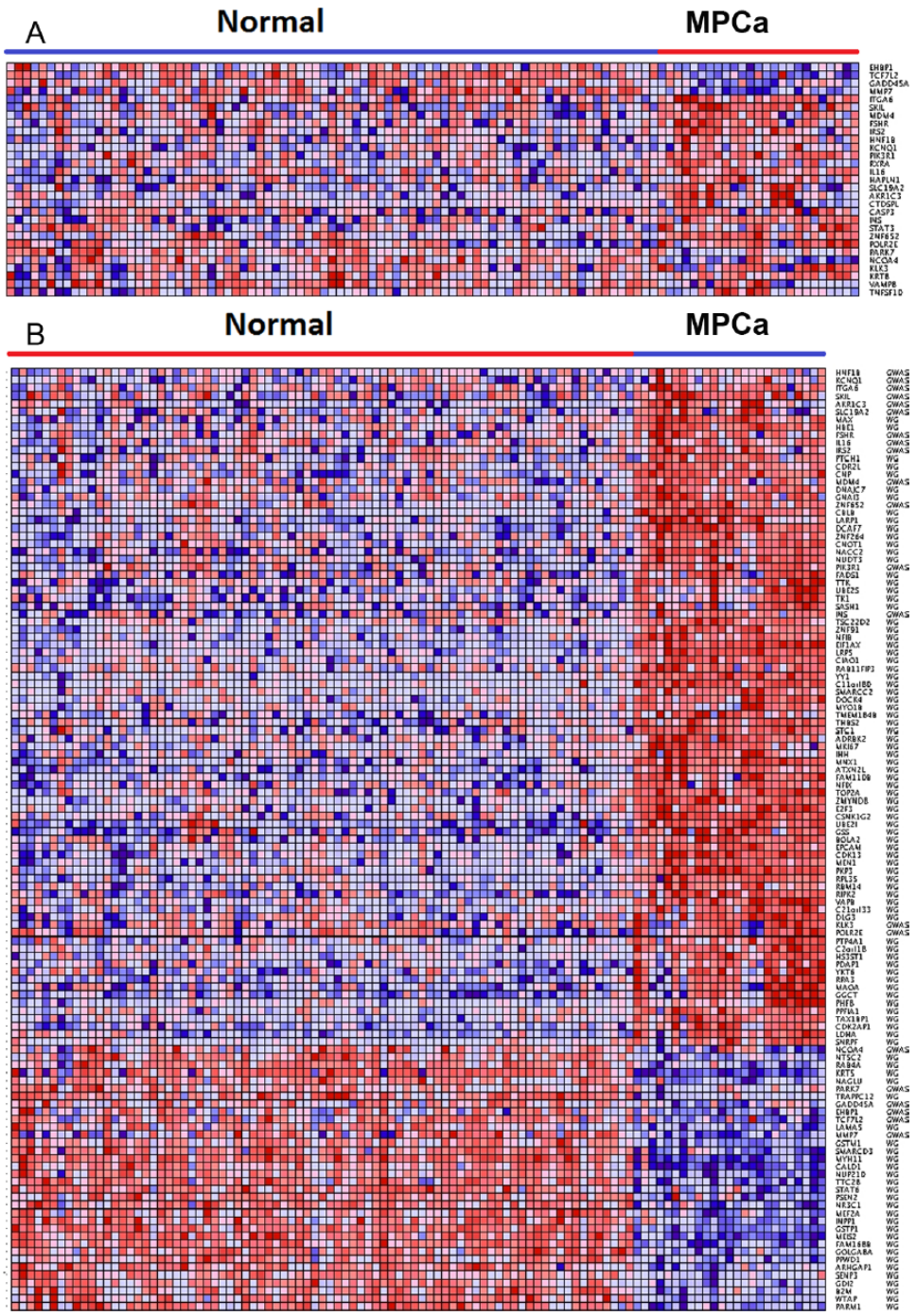

Figure 2. (A) Patterns of expression profiles for the 29 genes containing genetic variants strongly associated with an increased risk of developing prostate cancer evaluated in metastatic prostate cancer (MPCa) and control or normal samples. Patients are represented in columns and genes in rows. Red and blue colors indicate upregulation and downregulation, respectively. (B) Patterns of expression profiles for the 29 genes containing genetic variants strongly associated with an increased risk of developing prostate cancer and 100 non–genome-wide association studies genes evaluated in MPCa and control or normal samples. Patients are represented in columns and genes in rows. Red and blue colors indicate upregulation and downregulation, respectively.

associated with metastatic prostate cancer are presented in Figure 2A. The analysis revealed a set of 24 genes upregulated in metastatic disease; among them, the *KLK3* gene is associated with tumor aggressiveness (Figure 2A). The remaining 5 genes (*EHBP1, TCF7L2, GADD45A, MMP7*, and *NCOA4*) were downregulated (Figure 2A). Overall, the genes containing genetic susceptibility variants had similar patterns of expression profiles and were functionally related (Figure 2A). Combined analysis of GWAS and non-GWAS genes revealed similarities in patterns of expression profiles between the 2 sets of genes (Figure 2B).

The patterns of expression profiles for the 34 genes containing genetic susceptibility variants distinguishing patients with metastatic disease from those with primary organ–confined

tumors are presented in Figure 3A. The signature revealed 2 distinct clusters of genes. A cluster containing 9 genes (*CTBP2, AR, PRKC1, ZNF652, PDLIMS, VDR, PIK3R1, POU5F1*, and *TNRC68*) consistently upregulated in metastatic prostate tumors and downregulated in primary organ–confined tumors (Figure 3A). This cluster included the *AR* and *POU5F1* genes which contain genetic variants associated with aggressive tumors. The other cluster included a set of 25 genes consistently downregulated in metastatic prostate tumors (Figure 3A). Among the genes downregulated in metastatic tumors were *TERT, KLK2, MSMB*, and *WWOX* containing genetic variants associated with tumor aggressiveness (Figure 3A). Further analysis combining GWAS and non-GWAS genes revealed similarities and functional relationships between the 2 sets of genes (Figure 3B).

Overall, the analysis revealed similarities in patterns of expression profiles and functional relationships among GWAS and non-GWAS genes in all the 3 cases evaluated. However, as expected, there was considerable variation in patterns of gene expression profiles. This can be explained in part by the diversity of populations and phenotypes from which GWAS discoveries were derived.

Association with molecular networks and biological pathways

To gain insights about the broader biological context in which the genetic susceptibility variants operate, we performed network and pathway analysis. We sought to identify molecular networks and biological pathways enriched for genetic susceptibility variants in each disease state. The results showing the molecular networks enriched for genetic susceptibility variants in primary organ–confined prostate cancer are presented in Figure 4. Network analysis revealed that the genes containing genetic susceptibility variants with strong associations—*MSMB, KLK3, FGF10, AR, NKX3-1, TERT*, P*OU5F1, RFX6, FOXP4, ITGA6, CTBP2, FSMR, NCOA4*, and *SQRDL* (in red font)—are functionally related and interact with one another (Figure 4). The analysis also revealed that genes containing genetic variants with moderate-to-weak associations—*PRPH, MYC, GLI2, RUVBL1, HOXB13, IGF2, ERG, KLK2, PRKCI*, P*OLR2E*, and *KRT8* (in blue font)—are functionally related and interact with genes containing genetic variants with strong association (Figure 4). In addition, genes containing genetic susceptibility variants were found to interact with genes *RELA, RIPK3, UBC, EWSR1*, and *ANKRD7* not reported in GWAS (Figure 4).

Functional analysis of the genes in highly significant networks using GO information revealed sets of genes predicted to be involved in cellular development, cellular growth and proliferation, cancer, hereditary disorder, organism injury and abnormalities, tissue development, cellular development, cell death and survival, cell-mediated immune response, and cellular compromise. Pathway analysis revealed sets of genes involved in the androgen, androgen biosynthesis, growth hormone, and apoptosis signaling pathways. A complete list of genes containing genetic susceptibility variants and genes not identified by GWAS is presented in Table S7 provided as supplementary data to this report. Also presented in Table S4 is the information on molecular functions and biological processes in which the genes associated with primary organ–confined prostate cancer are involved.

The results showing the molecular networks enriched for genetic susceptibility variants for the metastatic disease are presented in Figure 5. The analysis revealed molecular networks enriched for genetic susceptibility variants (Figure 5). The networks included the genes *NKX3-1, KLK3, FGF10, AR, MSMB, ITGA6, NCOA4, CTBP2, TERT, CASP3*, and *PIK6R1* containing genetic susceptibility variants strongly associated with an increased risk of developing prostate cancer. The network also included genes *PARK7, TCF7L2, CDKN2A, IRS2*, and *VDR* containing genetic variants with moderate-to-weak associations and non-GWAS genes (Figure 5). Functional analysis using GO information revealed sets of genes predicted to be involved in posttranslation modification, nucleic acid metabolism, small-molecule biochemistry, cancer, endocrine system development and function, cancer, and gene expression. A complete list of genes containing genetic susceptibility variants and genes not identified by GWAS along with the biological processes and molecular functions in which they are presented in Table S5 provided as supplementary data to this report. The genes *AR, KLK3, NKX3.1, ITGA6, MSMB, CTBP2*, and *NCOA4* overlapped between the 2 disease states (Figures 4 and 5). Pathway analysis revealed biological pathways enriched for genetic susceptibility variants, including prostate cancer, AKT, P53, apoptosis, VDR, and AR signaling pathways.

As noted earlier in this report, some primary organ–confined prostate cancers progress to metastatic disease. To address this question, we performed additional networks and pathway analysis using a set of genes distinguishing the 2 disease states. We sought to identify molecular networks and biological pathways enriched for genetic susceptibility variants that are potential drivers of disease progression, and we performed network and pathway analyses using the set of genes distinguishing the 2 disease states. The analysis revealed molecular networks enriched for genetic susceptibility variants (Figure 6). The network included the genes *CTBP2, NKX3.1, TERT, KLK3, ITGA6, FGF10, THADA*, and *SKIL* containing genetic variants strongly associated with prostate cancer. Also identified were the genes *IGF2, AR, HOXB13, PARK7, NCOA4, VDR, KRT8*, and *GADD45A* containing SNPs with moderate-to-weak associations and the genes *AKT, RELA, EGFR, ER, HDAC1, UBC*, and *ELAVL1* not identified in GWAS.

Functional analysis of the genes in the networks revealed genes predicted to be involved in cellular growth and

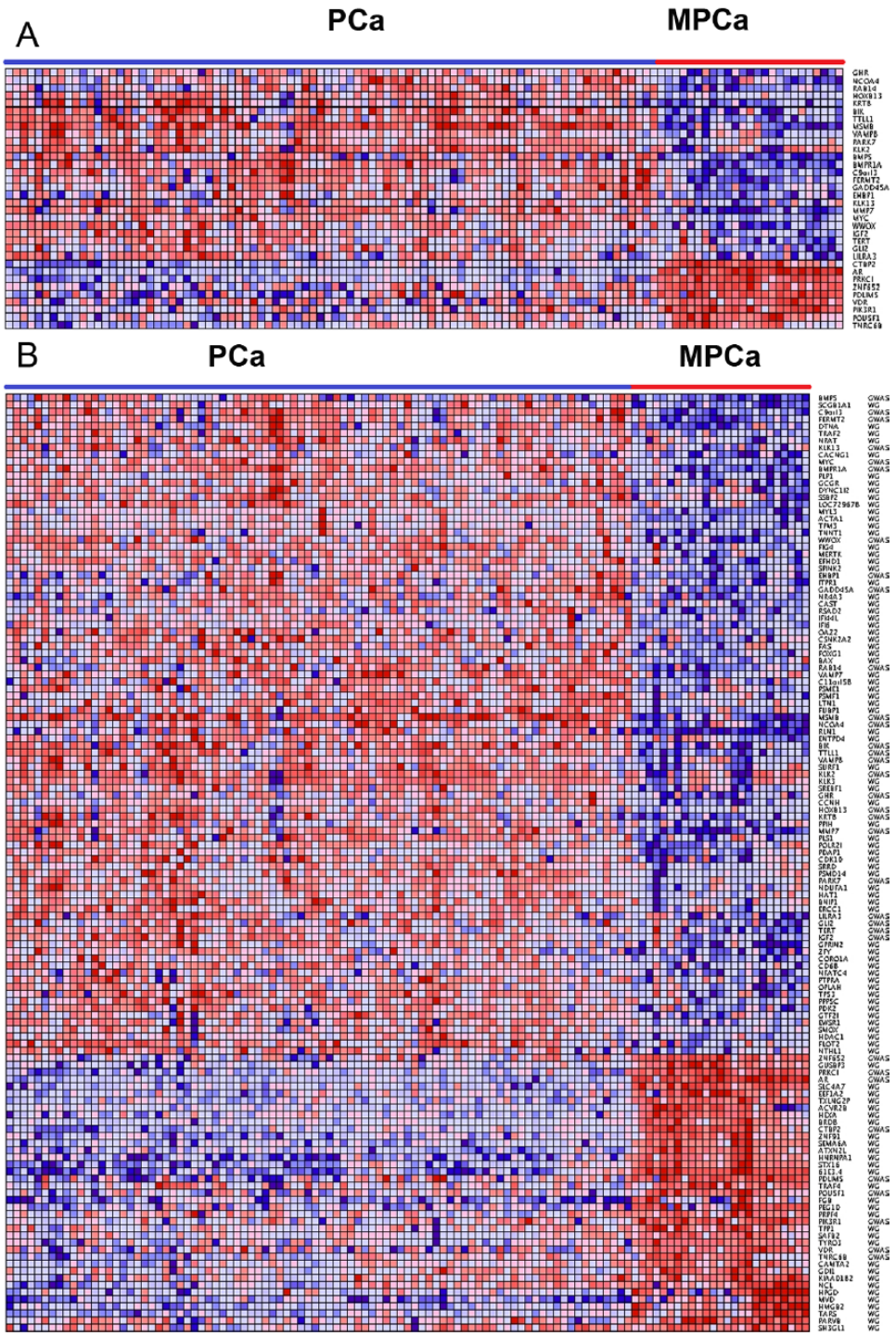

Figure 3. (A) Patterns of expression profiles for the 34 genes containing genetic variants strongly associated with an increased risk of developing prostate cancer evaluated in primary organ–confined prostate cancer (PCa) and metastatic prostate cancer (MPCa). Patients are represented in columns and genes in rows. Red and blue colors indicate upregulation and downregulation, respectively. (B) Patterns of expression profiles for the 34 genes containing genetic variants strongly associated with an increased risk of developing prostate cancer and 100 non–genome-wide association studies genes evaluated in PCa and MPCa. Patients are represented in columns and genes in rows. Red and blue colors indicate upregulation and downregulation, respectively.

proliferation, cellular development, cellular movement, hair and skin development and function, organ morphology, cancer, organ development, molecular transport, developmental disorders, cellular growth and proliferation, cell death and survival, reproductive system development and function, cellular assembly and organization, and lipid metabolism. A complete list of novel and SNP-containing genes showing the biological processes and molecular functions in which they are involved in primary organ–confined and metastatic prostate cancer is presented in Table S6 provided as supplementary data to this

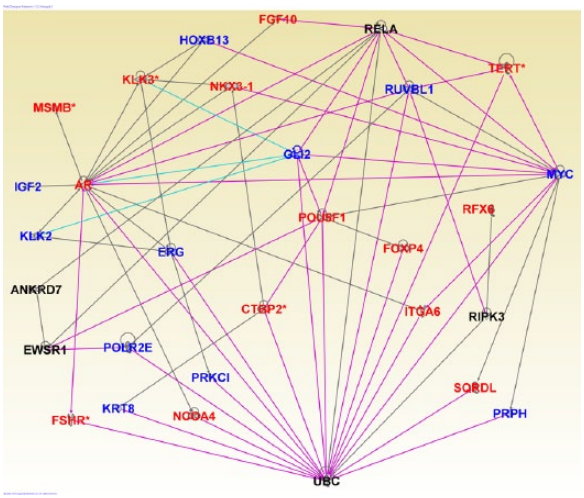

Figure 4. Gene interaction networks based on the set of genes containing genetic variants associated with an increased risk of developing prostate cancer and non–genome-wide association studies (GWAS) genes found to be highly significantly differentially expressed between patients with primary organ–confined prostate cancer and controls or normal samples. Genes containing single-nucleotide polymorphisms (SNPs) with strong associations are marked in red fonts. Blue fonts indicate genes containing SNPs with weak-to-moderate association. Genes in black font are non-GWAS genes. Nodes indicate the genes and the edges indicate interactions based on functional relationships. Information on functional relationships of the genes and biological processes in which the genes are involved is provided in Supplementary Table S4.

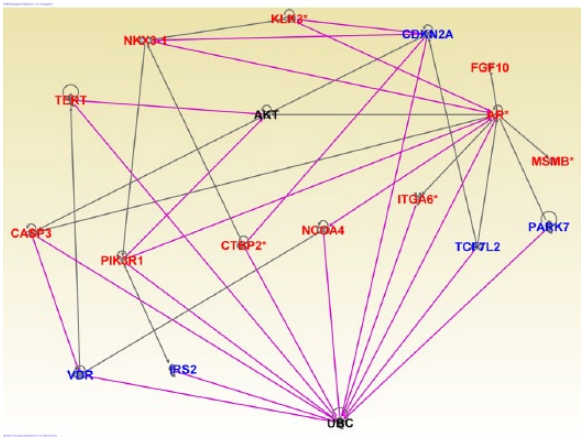

Figure 5. Gene interaction networks based of the set of genes containing genetic variants associated with an increased risk of developing prostate cancer and non–genome-wide association studies (GWAS) genes found to be highly significantly differentially expressed between patients with metastatic prostate cancer and controls or normal samples. Genes containing single-nucleotide polymorphisms (SNPs) with strong associations are marked in red fonts. Blue fonts indicate genes containing SNPs with weak-to-moderate association. Genes in black font are non-GWAS genes. Nodes indicate the genes and the edges indicate interactions based on functional relationships. Information on functional relationships of the genes and biological processes in which the genes are involved is provided in Supplementary Table S5.

report. Pathway analysis revealed the growth hormone and prostate cancer signaling pathways as the potential drivers of metastatic disease.

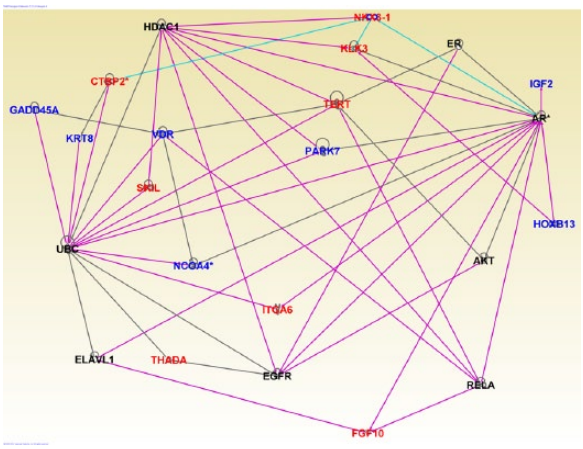

Figure 6. Gene interaction networks based of the set of genes containing genetic variants associated with an increased risk of developing prostate cancer and non–genome-wide association studies (GWAS) genes found to be highly significantly differentially expressed between patients with primary organ–confined prostate cancer and metastatic prostate cancer. Genes containing single-nucleotide polymorphisms (SNPs) with strong associations are marked in red fonts. Blue fonts indicate genes containing SNPs with weak-to-moderate association. Genes in black font are non-GWAS genes. Nodes indicate the genes, and the edges indicate interactions based on functional relationships. Information on functional relationships of the genes and biological processes in which the genes are involved is provided in Supplementary Table S6.

Discussion and Clinical Significance

We integrated GWAS information with gene expression data to identify gene signatures enriched for genetic susceptibility variants associated with primary organ–confined and metastatic tumors and distinguishing the 2 disease states. The results revealed gene signatures associated with and distinguishing the 2 disease states. Several studies have combined GWAS information with gene expression data.[20,22,23,43] However, this is the first study to report gene signatures enriched for genetic variants involved in primary organ–confined and metastatic prostate cancer and to distinguish the 2 disease states. The clinical significance of the results from this study can be summarized as follows:

1. *Patient stratification.* The discovery of gene signatures enriched for genetic susceptibility variants involved in primary organ–confined and metastatic tumors is significant. Because genetic variants regulate gene expression,[44,45] they could potentially be used to stratify patients to guide and or prioritize treatment options. For example, genetic variants (mapped to genes) rs2267437 (*XRCC6*), rs3815824 (*MVP*),[12,46] rs871135(*POU5F1P1*), 47 rs3774315 (*TNFSF10*), rs6497287 (HERC2),[16] rs1571801 (*DAP2IP*),[10] and rs2735839 (*KLK3*)[8,47] are associated with more aggressive prostate cancer. Thus, if confirmed or validated, these genetic variants and genes could be used to stratify and prioritize patients for treatment on the basis of potential disease aggressiveness.

2. *Patient screening using PSA.* Common genetic susceptibility variants could be used to improve the interpretation of serum PSA levels.[48,49] Literature reports estimate that 40% to 45% of the inter-individual variability in measured serum PSA concentrations can be explained by genetic factors,[50] Studies have shown that SNPs in or near the *KLK3* gene which encodes PSA can influence serum PSA concentrations and subsequently affect the frequency of prostate cancer screening and detection,[48–51] the interpretation of serum PSA values,[48,49] and the performance of PSA as a screening too.[52,53] Although our study did not focus on correlating GWAS information with PSA levels, a GWAS study has shown that several genetic variants associated with an increased risk of developing prostate cancer are strongly associated with serum PSA.[48]
3. *Reduction of unnecessary biopsies.* In a clinical setting, abnormal PSA values and digital rectal examinations determine the need for prostate biopsy.[8] However, literature evidence suggests that only 30% to 40% of patients with abnormal PSA values or physical examination are routinely diagnosed with prostate cancer on transrectal ultrasound-guided biopsy.[54] Recent studies have shown that genetic susceptibility variants are predictive of prostate cancer diagnosis on biopsy in men of European ancestry[55,56] and different racial/ethnic populations.[57–59] Thus, if confirmed, the genetic variants discovered in this study could be used to reduce the number of potentially unnecessary biopsies.
4. *Risk prediction.* One of the challenges in prostate cancer is risk prediction. Family history and genetic variants are measures of genetic susceptibility to prostate cancer.[8] Several studies have used genetic variants to calculate a genetic risk score (GRS) in different populations.[8] Published reports have suggested that GRS is significantly better in discriminative ability than knowledge of family history.[60,61]

The genetic variants and respective genes associated with tumor aggressiveness discovered in this study if confirmed could be used for calculating the GRS and improving risk prediction algorithms currently in use.

Limitations

This study provides insights about the global biological context in which genetic variants associated with an increased risk of developing prostate cancer operate in primary organ–confined and metastatic tumors. However, limitations of the study must be acknowledged. The GWAS information used in this study is biased toward men of European ancestry, and gene expression data used in this study were derived from the white population. Genetic susceptibility and level of prostate cancer aggressiveness vary among race and ethnic populations.[59,62–64] In addition, gene expression can vary and differ among populations,[65] and the data set used in this study is small. Larger studies are needed to evaluate perturbations and potential clinical utility of genes and variants associated with an increased risk of developing prostate cancer in different race and ethnic populations. In this study, we did not study allele-specific expression and the functional impact of genetic susceptibility variants on genes and pathways. However, several published reports have documented allele-specific expression[66–69] in human populations and more recently in prostate cancer.[70–72] Recently, Larson et al[70] reported a comprehensive evaluation of *cis*-regulatory variation in the human prostate transcriptome using gene-level allele-specific expression. More recently, Whitington et al[71] elucidated the gene regulatory mechanisms underpinning prostate cancer susceptibility. Moreover, evidence from the literature indicates that genetic variants regulate and account for differences in gene expression among populations.[73,74]

Conclusions

The results in this report demonstrate that integrative analysis combining GWAS information with gene expression data is a powerful approach for linking genetic predisposition to clinical phenotypes in prostate cancer. This exploratory study establishes putative functional bridges between GWAS discoveries and biological pathways in primary organ–confined and metastatic tumors. Validation of these genetic variants using large test and validation data will permit additional tools that can be incorporated into algorithms that can improve clinical practice and risk prediction in prostate cancer.

Acknowledgements

The authors thank Louisiana University Health Sciences Center—School of Medicine, the Center for Clinical and Translational Science at the University of Alabama, Mangosuthu University of Technology, and the South African Medical Research Council for supporting this research project. They confirm that the funders had no influence over the study design, content of the article, or selection of this journal.

Author Contributions

CH, RR, ZD, NG, LM, OS, and NG conceived and designed the experiments. CH and RR analyzed the data. CH, RR, OS, RB, LM, ZD and NG wrote the first draft of the manuscript. CH, RR, ZD, NG, and LM contributed to the writing of the manuscript. CH, RR, OS, and RB agree with manuscript results and conclusions. CH jointly developed the structure and arguments for the paper. CH, RR, LM, RB, OS, ZD, and NG made critical revisions and approved final version. All authors reviewed and approved the final manuscript.

REFERENCES

1. Cancer Facts & Figures 2015. American Cancer Society. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2015.html.
2. 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;7:600–611.

3. Lapointe J, Li C, Higgins JP, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A*. 2004;101: 811–816.
4. 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;62:4499–4506.
5. Penney KL, Sinnott JA, Fall K, et al. mRNA expression signature of Gleason grade predicts lethal prostate cancer. *J Clin Oncol*. 2011;29:2391–2396.
6. 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;106:9362–9367.
7. 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;12:175–191.
8. Helfand BT, Roehl KA, Cooper PR, et al. Associations of prostate cancer risk variants with disease aggressiveness: results of the NCI-SPORE Genetics Working Group analysis of 18,343 cases. *Hum Genet*. 2015;134:439–450.
9. Berndt SI, Wang Z, Yeager M, et al. Two susceptibility loci identified for prostate cancer aggressiveness. *Nat Commun*. 2015;6:6889.
10. Duggan D, Zheng SL, Knowlton M, et al. Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst*. 2007;99:1836–1844.
11. 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;22:408–415.
12. Pal P, Xi H, Guha S, et al. Common variants in 8q24 are associated with risk for prostate cancer and tumor aggressiveness in men of European ancestry. *Prostate*. 2009;69:1548–1556.
13. Weng PH, Huang YL, Page JH, et al. Polymorphisms of an innate immune gene, toll-like receptor 4, and aggressive prostate cancer risk: a systematic review and meta-analysis. *PLoS ONE*. 2014;9:e110569.
14. Nam RK, Zhang W, Siminovitch K, et al. New variants at 10q26 and 15q21 are associated with aggressive prostate cancer in a genome-wide association study from a prostate biopsy screening cohort. *Cancer Biol Ther*. 2011;12:997–1004.
15. Machiela MJ, Chen C, Liang L, et al. One thousand genomes imputation in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium aggressive prostate cancer genome-wide association study. *Prostate*. 2013;73: 677–689.
16. FitzGerald LM, Kwon EM, Conomos MP, et al. Genome-wide association study identifies a genetic variant associated with risk for more aggressive prostate cancer. *Cancer Epidemiol Biomarkers Prev*. 2011;20:1196–1203.
17. Gallagher DJ, Vijai J, Cronin AM, et al. Susceptibility loci associated with prostate cancer progression and mortality. *Clin Cancer Res*. 2010;16:2819–2832.
18. Penney KL, Pyne S, Schumacher FR, et al. Genome-wide association study of prostate cancer mortality. *Cancer Epidemiol Biomarkers Prev*. 2010;19: 2869–2876.
19. Szulkin R, Karlsson R, Whitington T, et al. Genome-wide association study of prostate cancer-specific survival. *Cancer Epidemiol Biomarkers Prev*. 2015;24: 1796–1800.
20. Penney KL, Sinnott JA, Tyekucheva S, et al. Association of prostate cancer risk variants with gene expression in normal and tumor tissue. *Cancer Epidemiol Biomarkers Prev*. 2015;24:255–260.
21. 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:S13.
22. Hicks C, Koganti T, Giri S, et al. Integrative genomic analysis for the discovery of biomarkers in prostate cancer. *Biomark Insights*. 2014;9:39–51.
23. Sircar K, Huang H, Hu L, et al. Integrative molecular profiling reveals asparagine synthetase is a target in castration-resistant prostate cancer. *Am J Pathol*. 2012;180:895–903.
24. Xu X, Hussain WM, Vijai J, et al. Variants at IRX4 as prostate cancer expression quantitative trait loci. *Eur J Hum Genet*. 2014;22:558–563.
25. Jiang J, Cui W, Vongsangnak W, Hu G, Shen B. Post genome-wide association studies functional characterization of prostate cancer risk loci. *BMC Genomics*. 2013;14:S9.
26. Ioannidis JP, Boffetta P, Little J, et al. Assessment of cumulative evidence on genetic associations: interim guidelines. *Int J Epidemiol*. 2008;37:120–132.
27. Khoury MJ, Bertram I, Boffetta P, et al. Genome-wide association studies, field synopses, and the development of the knowledge base on genetic variation and human diseases. *Am J Epidemiol*. 2009;170:269–279.
28. Sagoo GS, Little J, Higgins JP. Systematic reviews of genetic association studies. Human Genome Epidemiology Network. *PLoS Med*. 2009;6:e28.
29. 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;151:264–269.
30. 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:e1000100.
31. dbSNP. http://www.ncbi.nlm.nih.gov/SNP/.
32. Human Genome Nomenclature Committee (HGNC). http://www.genenames.org/.
33. Stuart RO, Wachsman W, Berry CC, et al. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. *Proc Natl Acad Sci U S A*. 2004;101:615–620.
34. Chandran UR, Dhir R, Ma C, Michalopoulos G, Becich M, Gilbertson J. Differences in gene expression in prostate cancer, normal appearing prostate tissue adjacent to cancer and prostate tissue from cancer free organ donors. *BMC Cancer*. 2005;5:45.
35. Chandran UR, Ma C, Dhir R, et al. Gene expression profiles of prostate cancer reveal involvement of multiple molecular pathways in the metastatic process. *BMC Cancer*. 2007;7:64.
36. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25:25–29.
37. Morrissey ER, Diaz-Uriarte R. Pomelo II: finding differentially expressed genes. *Nucleic Acids Res*. 2009;37:W581–W586.
38. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P. GenePattern 2.0. *Nat Genet*. 2006;38:500–501.
39. Ramacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. *J Comput Biol*. 2002;9:505–511.
40. Benjamini Y, Yosef H. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Society*. 1995;57:289–300.
41. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*. 1998;95: 14863–14868.
42. *Ingenuity Pathways Analysis (IPA) System*. Redwood, CA: Ingenuity Systems. http://www.ingenuity.com/.
43. Gorlov IP, Gallick GE, Gorlova OY, et al. GWAS meets microarray: are results from genome-wide association studies and gene-expression profiling consistent? prostate cancer as an example. *PLoS ONE*. 2009;4:e6511.
44. Cheung VG, Spelman RS. Genetics of human gene expression: mapping DNA variants that influence gene expression. *Nat Rev Genet*. 2009;10:595–604.
45. Dimas AS, Deutsch S, Stranger BE, et al. Common regulatory variation impacts gene expression in cell type-dependent manner. *Science*. 2009;325: 1246–1250.
46. Henríquez-Hernández LA, Valenciano A, Foro-Arnalot P, Álvarez-Cubero MJ. Association between single-nucleotide polymorphisms in DNA double-strand break repair genes and prostate cancer aggressiveness in the Spanish population. *Prostate Cancer Prostatic Dis*. 2016;19:28–34.
47. Kader AK, Sun J, Isaacs SD, et al. Individual and cumulative effect of prostate cancer risk-associated variants on clinicopathologic variables in 5,895 prostate cancer patients. *Prostate*. 2009;69:1195–1205.
48. Gudmundsson J, Besenbacher S, Sulem P, Gudbjartsson DF. Genetic correction of PSA values using sequence variants associated with PSA levels. *Sci Transl Med*. 2010;2:62ra92.
49. Helfand BT, Loeb S, Hu Q, et al. Personalized prostate specific antigen testing using genetic variants may reduce unnecessary prostate biopsies. *J Urol*. 2013;189:1697–1701.
50. Bansal A, Murray DK, Wu JT, Stephenson RA, Middleton RG, Meikle AW. Heritability of prostate-specific antigen and relationship with zonal prostate volumes in aging twins. *J Clin Endocrinol Metab*. 2000;85:1272–1276.
51. Loeb S. Germline sequence variants and prostate-specific antigen interpretation. *Clin Chem*. 2011;57:662–663.
52. Vickers AJ, Cronin AM, Roobol MJ, et al. A four-kallikrein panel predicts prostate cancer in men with recent screening: data from the European Randomized Study of Screening for Prostate Cancer, Rotterdam. *Clin Cancer Res*. 2010;16:3232–3239.
53. Vickers A, Cronin A, Roobol M, et al. Reducing unnecessary biopsy during prostate cancer screening using a four-kallikrein panel: an independent replication. *J Clin Oncol*. 2010;28:2493–2498.
54. Cormio L, Scattoni V, Lorusso F, et al. Prostate cancer detection rates in different biopsy schemes. Which cores for which patients? *World J Urol*. 2014;32:341–346.
55. Ren S, Xu J, Zhou T, et al. Plateau effect of prostate cancer risk-associated SNPs in discriminating prostate biopsy outcomes. *Prostate*. 2013;73:1824–1835.
56. Aly M, Wiklund F, Xu J, et al. Polygenic risk score improves prostate cancer risk prediction: results from the Stockholm-1 cohort study. *Eur Urol*. 2011;60: 21–28.
57. Akamatsu S, Takahashi A, Takata R, et al. Reproducibility, performance, and clinical utility of a genetic risk prediction model for prostate cancer in Japanese. *PLoS ONE*. 2012;7:e46454.
58. Jiang H, Liu F, Wang Z, et al. Prediction of prostate cancer from prostate biopsy in Chinese men using a genetic score derived from 24 prostate cancer risk-associated SNPs. *Prostate*. 2013;73:1651–1659.
59. Xu Z, Bensen JT, Smith GJ, Mohler JL, Taylor JA. GWAS SNP replication among African American and European American men in the North Carolina-Louisiana Prostate Cancer Project (PCaP). *Prostate*. 2011;71:881–891.

60. Kader AK, Sun J, Reck BH, et al. Potential impact of adding genetic markers to clinical parameters in predicting prostate biopsy outcomes in men following an initial negative biopsy: findings from the REDUCE trial. *Eur Urol*. 2012;62:953–961.
61. Sun J, Na R, Hsu FC, et al. Genetic score is an objective and better measurement of inherited risk of prostate cancer than family history. *Eur Urol*. 2013;63:585–587.
62. Han Y, Signorello LB, Strom SS, et al. Generalizability of established prostate cancer risk variants in men of African ancestry. *Int J Cancer*. 2015;136:1210–1217.
63. Hoffmann TJ, Van Den Eeden SK, Sakoda LC, Jorgenson E. A large multiethnic genome-wide association study of prostate cancer identifies novel risk variants and substantial ethnic differences. *Cancer Discov*. 2015;5:878–891.
64. Gusev A, Shi H, Kichaev G, Pomerantz M, Li F. Atlas of prostate cancer heritability in European and African-American men pinpoints tissue-specific regulation. *Nat Commun*. 2016;7:10979.
65. Powell IJ, Dyson G, Land S, et al. Genes associated with prostate cancer are differentially expressed in African American and European American men. *Cancer Epidemiol Biomarkers Prev*. 2013;22:891–897.
66. Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. Allelic variation in human gene expression. *Science*. 2002;297:1143.
67. Buckland PR. Allele-specific gene expression differences in humans. *Hum Mol Genet*. 2004;13:R255–R260.
68. Paracios R, Gazave E, Goni J, et al. Allelle-specific gene expression is widespread across the genome and biological processes. *PLoS ONE*. 2009;4:e4150.
69. Meyer KB, Maja A-T, O'Reilly M, et al. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol*. 2008;6:e108.
70. Larson NB, McDonnell S, French AJ, et al. Comprehensively evaluating cis-regulatory variation in the human prostate transcriptome by using gene-level allele-specific expression. *Am J Hum Genet*. 2015;96:869–882.
71. Whitington T, Gao P, Song W, et al. Gene regulatory mechanisms underpinning prostate cancer susceptibility. *Nat Genet*. 2016;48:387–397.
72. Jin HJ, Jung S, DebRoy AR, Davuluri RV. Identification and validation of regulatory SNPs that modulate transcription factor chromatin binding and gene expression in prostate cancer. *Oncotarget*. 2016;7:54616–54626.
73. Spielman RS, Bastone LA, Burdick JT, et al. Common genetic variants account for differences in gene expression among ethnic populations. *Nat Genet*. 2007;39:226–231.
74. 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;82:631–640.