Genetic and functional analyses implicate the *NUDT11*, *HNF1B*, and *SLC22A3* genes in prostate cancer pathogenesis

Chiara Grisanzio[a], Lillian Werner[b], David Takeda[a,c], Bisola C. Awoyemi[a], Mark M. Pomerantz[a,d], Hiroki Yamada[a,e], Prasanna Sooriakumaran[f], Brian D. Robinson[f], Robert Leung[f], Anna C. Schinzel[g], Ian Mills[h,i,j], Helen Ross-Adams[k], David E. Neal[k], Masahito Kido[e], Toshihiro Yamamoto[e], Gillian Petrozziello[a], Edward C. Stack[a], Rosina Lis[a,l], Philip W. Kantoff[a], Massimo Loda[a,d,g,l], Oliver Sartor[m], Shin Egawa[e], Ashutosh K. Tewari[f], William C. Hahn[a,c,d,g], and Matthew L. Freedman[a,g,n,1]

[a]Department of Medical Oncology, [b]Department of Biostatistics and Computational Biology, and [g]Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, MA 02115; [c]Cancer Program and [n]Medical and Population Genetics Program, Broad Institute of Harvard and MIT, Cambridge, MA 02142; [d]Department of Medicine and [l]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115; [e]Department of Urology, Jikei University School of Medicine, Minato-ku, Tokyo 105-8461, Japan; [f]Department of Urology, Lefrak Center of Robotic Surgery and Prostate Cancer Institute, Weill Cornell Medical College, New York, NY 10065; [h]Centre for Molecular Medicine Norway (NCMM), University of Oslo, N-0349 Oslo, Norway, Departments of [i]Urology and [j]Cancer Prevention, Oslo University Hospitals, N-0370 Oslo, Norway [k]Uro-Oncology Research Group, Cancer Research United Kingdom, Cambridge Research Institute, Cambridge CB2 0RE, United Kingdom; and [m]Department of Medicine and Urology, Tulane Cancer Center, Tulane University School of Medicine, New Orleans, LA 70112

Edited* by Francis S. Collins, National Institutes of Health, Bethesda, MD, and approved May 25, 2012 (received for review January 18, 2012)

One of the central goals of human genetics is to discover the genes and pathways driving human traits. To date, most of the common risk alleles discovered through genome-wide association studies (GWAS) map to nonprotein-coding regions. Because of our relatively poorer understanding of this part of the genome, the functional consequences of trait-associated variants pose a considerable challenge. To identify the genes through which risk loci act, we hypothesized that the risk variants are regulatory elements. For each of 12 known risk polymorphisms, we evaluated the correlation between risk allele status and transcript abundance for all annotated protein-coding transcripts within a 1-Mb interval. A total of 103 transcripts were evaluated in 662 prostate tissue samples [normal (n = 407) and tumor (n = 255)] from 483 individuals [European Americans (n = 233), Japanese (n = 127), and African Americans (n = 123)]. In a pooled analysis, 4 of the 12 risk variants were strongly associated with five transcripts (*NUDT11*, *MSMB*, *NCOA4*, *SLC22A3*, and *HNF1B*) in histologically normal tissue ($P \leq 0.001$). Although associations were also observed in tumor tissue, they tended to be more attenuated. Previously, we showed that *MSMB* and *NCOA4* participate in prostate cancer pathogenesis. Suppressing the expression of *NUDT11*, *SLC22A3*, and *HNF1B* influences cellular phenotypes associated with tumor-related properties in prostate cancer cells. Taken together, the data suggest that these transcripts contribute to prostate cancer pathogenesis.

expression quantitative trait loci | prostate cancer risk SNPs | multi-ethnic

Genome-wide association studies (GWAS) have identified over 40 risk loci strongly associated with prostate cancer (PCa) (1–10). Similar to other complex traits, the majority of the loci are located in introns and intergenic regions. This observation is in stark contrast to Mendelian traits, which typically harbor mutations in protein-coding sequences. Nucleotide changes in protein-coding regions and their effects on amino acids are readily deciphered using the genetic code. Therefore, fine mapping and gene identification for Mendelian traits typically consists of resequencing exons. On the other hand, predicting the functional consequences of DNA changes in the nonprotein-coding portion of the genome presents a major challenge because of our comparatively poorer understanding of this region. Therefore, identifying the genes underlying complex traits is not straightforward.

RNA transcript abundance is a heritable trait and is thus amenable to genetic mapping. Genetic polymorphisms, such as SNPs, associated with transcript levels are typically referred to as expression quantitative trait loci (eQTLs) (ref. 11 and references therein). In fact, recent data demonstrates that alleles discovered through GWAS are enriched for eQTLs (12). One can therefore seek transcript levels that are correlated with risk allele status. Prior studies reveal that this is a promising strategy to link nonprotein-coding risk alleles with their target genes (13–17).

A number of factors deserve careful consideration when undertaking an eQTL analysis. Selection of which tissue to evaluate is important. Previous studies demonstrate that many eQTLs are tissue-specific (14, 15, 18–20). To uncover genes related to PCa pathogenesis, we reasoned that it was appropriate to evaluate transcripts from prostatectomy specimens. Tissue state is another factor: should normal or tumor tissue be studied? Because there is no clear precedent, we studied both types of tissue. Finally, should all transcripts or a subset be studied? This point directly relates to the power of the study to detect a particular magnitude of effect. Prior studies have shown that transcripts in the vicinity of the eQTL are easier to detect because eQTLs tend to have a larger effect size on nearby transcripts (11, 21). Therefore, all annotated transcripts within 1 Mb of each risk locus were evaluated in a total of 662 samples (407 normal and 255 tumor).

The majority of GWAS have been performed in individuals of European ancestry. Many (but not all) of the common prostate risk loci discovered in men of European ancestry tend to replicate in other ethnic groups. To evaluate if eQTL/target gene associations were similar across ethnic groups, we performed this study in tissues derived from European American (EA), African American (AA), and Japanese men.

Functional analyses are important for further elucidating the biology of the genes implicated through gene-expression analysis. A possible scenario is that gene-expression levels are associated with a risk locus, yet are not related to the trait under study (i.e., the gene is in fact regulated by the risk region, but does not play a role in the trait). Therefore, genes that were associated with risk loci were further evaluated in functional assays.

Given the rapidity with which GWAS are discovering regions associated with complex traits, gene identification has become

Author contributions: C.G., W.C.H., and M.L.F. designed research; C.G., D.T., B.C.A., M.M.P., H.Y., and A.C.S. performed research; P.S., B.D.R., R. Leung, I.M., H.R.-A., D.E.N., M.K., T.Y., G.P., E.C.S., R. Lis, P.W.K., M.L., O.S., S.E., A.K.T., and M.L.F. contributed new reagents/analytic tools; C.G., L.W., D.T., M.M.P., A.C.S., W.C.H., and M.L.F. analyzed data; and C.G. and M.L.F. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: freedman@broadinstitute.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1200853109/-/DCSupplemental.

a bottleneck. The goal of this study was to outline a strategy for identifying genes underlying human PCa pathogenesis.

Results

Association Between PCa Risk Loci and Expression Levels of Candidate Genes in Normal and Tumor Prostate Tissue. We genotyped 12 PCa risk loci (1–7) and evaluated expression levels of 103 genes within a 1-Mb interval of each locus (Table S1). Gene-expression analyses were performed on radical prostatectomy specimens from 483 subjects of diverse ancestral backgrounds (*Materials and Methods*). A total of 662 samples (407 histologically normal and 255 tumor prostate tissues) were evaluated (Table S2).

The data are presented based on the European American samples because the GWAS have been predominantly performed in these populations. Normal and tumor tissues were analyzed independently. Subsequently, a pooled analysis of the normal samples of all individuals (adjusting for cohort) was performed for the nominally significant associations (Table 1 and Table S3). Nonsignificant eQTL/target genes are presented in Table S4.

In the histologically normal tissue of the EA population, 8 of the 12 PCa risk variants (rs9364554, rs6465657, rs10993994, rs4962416, rs10896449, rs4430796, rs2737839, and rs5945619) are nominally ($P \leq 0.05$) associated with at least one candidate gene ($P = 0.04–6.9 \times 10^{-46}$) (Fig. 1 and Table 1). In tumor tissue, four of these eight variants (rs10993994, rs5945619, rs6465657, and rs9364554) are associated with the identical genes that are associated with these variants in normal tissue. Three of these four alleles possess a more attenuated level of significance in tumors ($P = 0.03–6.3 \times 10^{-10}$) (Table 1). We had previously described the association between one risk variant, rs10993994, and *MSMB* and *NCOA4* in both normal and tumor tissue (15). Two additional risk loci (rs2660753 and rs6465657) show modest associations ($P = 0.03–0.04$) with transcripts in tumor, but are not present in normal tissue.

In the Japanese normal tissue, three variants that were also significant in the EA normal tissue (rs10993994, rs4430796, and rs5945619) were associated with one or more genes ($P = 0.02–1.4 \times 10^{-12}$) (Fig. 1 and Table 1). In the tumor tissue of Japanese individuals, these same three variants were also significantly associated with transcript levels of the same genes with the exception of *HNF1B* ($P = 0.04–8.7 \times 10^{-7}$) (Table 1).

In the AA normal tissue samples, four variants (rs10993994, rs4430796, rs5945619, and rs6465657) showed an association with a subset of genes significant also in the EA normal tissue ($P = 0.046–1.4 \times 10^{-19}$) (Fig. 1 and Table 1). In the tumor prostate tissue of AA individuals, these four variants were also significantly associated with one or more genes that were also significant in EA tissue ($P = 0.04–2.98 \times 10^{-5}$) (Table 1).

In a pooled analysis across all of the normal tissue samples ($n = 407$), we observed strong associations ($P \leq 0.001$) between four risk SNPs (rs10993994, rs5945619, rs4430796, and rs9364554) and the transcript levels of five genes: *MSMB*, *NCOA4*, *NUDT11*, *HNF1B*, and *SLC22A3* (Table 1).

Functional Studies: Alteration of Gene-Expression Levels in Prostate Cancer Cell Lines. We next evaluated the functional relevance of the genes that were significantly associated with risk allele status. We previously showed that *MSMB* and *NCOA4* participate in PCa tumorigenesis (15). The genetic data demonstrate that risk variants rs5945619 and rs4430796 are associated with increased expression of *NUDT11* and *HNF1B*, respectively, whereas

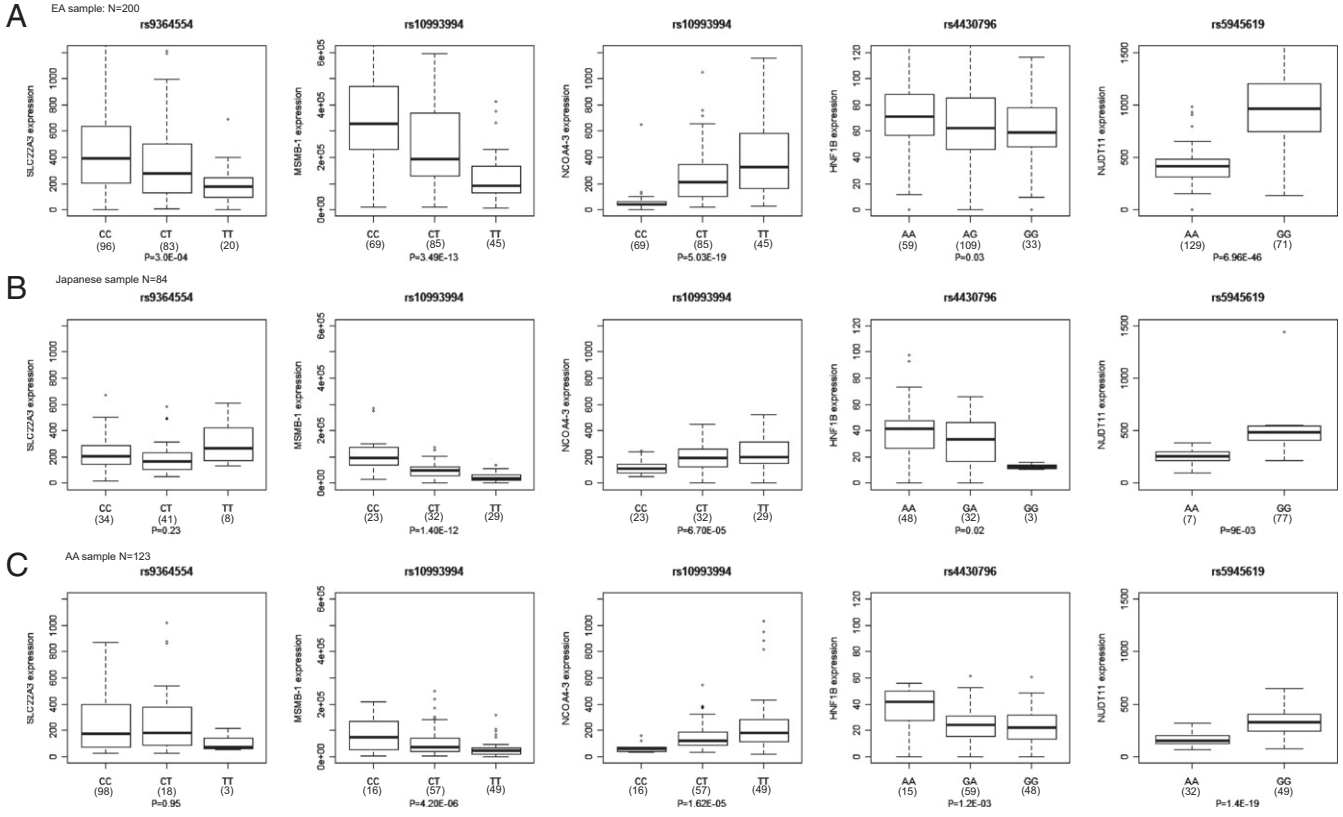

Fig. 1. RNA expression of significantly associated genes in normal prostate tissue of EA, Japanese, and AA individuals. Each distribution is summarized as a boxplot. The horizontal line within the box represents the median of the distribution, and the hinges of the box represent the 25th and 75th percentiles. The *P* value for each graph denotes the significance of association between expression and genotype. (*A*) Expression in histologically normal EA tissue ($n = 200$). (*B*) Expression in histologically normal Japanese tissue ($n = 84$). (*C*) Expression in histologically normal AA tissue ($n = 123$).

Table 1. Prostate cancer risk variants associated with gene expression levels in individuals of European, Japanese, and African ancestries

			EA individuals				Japanese individuals				AA individuals				N tissue, all populations	
Chromosome	SNP(s)	Genes	*P* values N (*n* = 200)	Correlation with reference allele	*P* values T (*n* = 96)	Correlation with reference allele	*P* values N (*n* = 84)	Correlation with reference allele	*P* values T (*n* = 114)	Correlation with reference allele	*P* values N (*n* = 123)	Correlation with reference allele	*P* values T (*n* = 45)	Correlation with reference allele	*P* values (*n* = 407)	Correlation with reference allele
2p15	rs721048	*TMEM17*	0.71		0.38		0.91		0.54		8.00E-03	Increases	0.22		0.94	
3p12.1	rs2660753	*VGLL3*	0.09		0.04	Decreases	0.41		0.29		0.22		0.83		0.11	
		CHMP2B	0.32		0.22		0.23		0.02	Increases	0.04	Increases	0.85		0.01	
		POU1F1-1	0.54		0.95		0.97		0.72		0.03	Increases	0.17		0.03	
6q25.3	rs9364554	*MAS1*	0.26		0.23		0.11		0.47		0.72		0.03	Increases	0.10	
		SLC22A3	**3.00E-04**	**Decreases**	**0.03**	**Decreases**	**0.23**		**0.05**	**Increases**	**0.95**		**0.29**		**5.00E-04**	**Decreases**
7p15.2	rs10486567	*HIBADH*	0.96		0.87		0.99		0.42		0.01	Decreases	0.33		0.28	
7q21.3	rs6465657	*ASNS*	4.40E-02	Increases	2.00E-03	Increases	0.76		0.03	Decreases	0.06		6.00E-03	Increases	0.62	
		BAIAP2L1	0.20		0.12		0.66		0.02	Increases	0.98		0.77		0.30	
		BHLHA15	0.03	Increases	4.00E-03	Increases	0.11		6.00E-03	Increases	4.60E-02	Increases	0.31		0.003	
		BRI3	0.70		0.96		0.52		0.65		4.60E-03	Increases	0.06		0.62	
		LMTK2	0.14		0.03	Increases	0.64		0.03	Increases	0.45		0.59		0.06	
8q24	rs6983267	*POU5F1B*	0.76		0.40		0.51		0.27		0.02	Decreases	0.59		0.48	
10q11.23	rs10993994	***MSMB1***	**3.49E-13**	**Decreases**	**8.70E-05**	**Decreases**	**1.40E-12**	**Decreases**	**8.70E-07**	**Decreases**	**4.20E-06**	**Decreases**	**3.30E-05**	**Decreases**	**4.00E-19**	**Decreases**
		MSMB2	**2.10E-12**	**Decreases**	**3.00E-04**	**Decreases**	**2.70E-08**	**Decreases**	**2.80E-05**	**Decreases**	**2.00E-03**	**Decreases**	**2.00E-04**	**Decreases**	**3.00E-13**	**Decreases**
		NCOA4-1	**1.77E-12**	**Increases**	**0.14**		**0.01**	**Increases**	**0.43**		**6.20E-07**	**Increases**	**0.09**		**2.00E-18**	**Increases**
		NCOA4-3	**5.03E-19**	**Increases**	**8.30E-05**	**Increases**	**6.70E-05**	**Increases**	**0.02**	**Increases**	**1.62E-05**	**Increases**	**0.04**	**Increases**	**8.00E-25**	**Increases**
10q26.13	rs4962416	*CTBP2-2*	0.01	Increases	0.08		0.08		0.75		0.46		0.56		0.10	
		ZRANB1	7.40E-03	Increases	0.26		0.88		0.29		0.52		0.45		0.33	
11q13.2	rs10896449	*MTL5-2*	0.03	Decreases	0.51		0.66		0.89		0.91		0.72		0.54	
		ORAOV1	0.05	Decreases	0.65		0.59		0.95		0.38		0.84		0.81	
17q12	rs4430796	*C17orf78*	0.88		0.84		0.61		0.11		0.04	Decreases	0.18		0.66	
		DDX52	0.28		0.21		0.13		0.09		6.00E-03	Increases	0.02	Increases	0.49	
		HNF1B	**0.03**	**Increases**	**0.20**		**0.02**	**Increases**	**0.15**		**1.20E-03**	**Increases**	**0.13**		**2.00E-04**	**Increases**
		SYNRG	0.81		0.25		0.83		0.62		6.70E-03	Increases	0.06		0.59	
		TADA2A-2	0.38		0.55		0.74		0.04	Increases	0.10		0.21		0.10	
19q	rs2735839	*CD33*	0.71		0.29		0.37		0.86		0.01	Decreases	0.90		0.35	
		KLK1	0.03	Increases	0.63		0.38		0.22		0.20		0.51		0.17	
		KLK15	0.87		0.58		0.36		0.51		0.04	Increases	0.89		0.40	
		KLK3	0.05	Decreases	0.63		0.61		0.65		0.61		0.49		0.37	
		KLK5	0.62		0.22		0.35		0.90		0.04	Decreases	0.38		0.36	
		KLK7	0.75		0.60		0.39		0.93		0.04	Decreases	0.99		0.41	
		KLK8	1.00E-03	Decreases	0.95		0.55		0.85		0.79		0.94		0.38	
		KLK9	0.38		0.17		0.35		0.59		0.04	Decreases	0.09		0.36	
		SPIB	0.26		0.42		0.38		0.25		0.04	Decreases	0.82		0.31	
Xp11.22	rs5945619	*CXorf67*	0.44		0.17		7.00E-04	Increases	0.82		0.55		0.21		0.05	
		MAGED1	0.61		0.69		0.78		0.66		0.47		0.40		0.49	
		NUDT11	**6.96E-46**	**Increases**	**6.30E-10**	**Increases**	**9.00E-03**	**Increases**	**8.30E-05**	**Increases**	**1.40E-19**	**Increases**	**2.98E-05**	**Increases**	**1.00E-47**	**Increases**

Highlighted in boldface are the genes that maintained significance in a pooled analysis across populations. N, normal; T, tumor.

rs9364554 is associated with decreased *SLC22A3* transcript abundance. Functional studies were performed in four cell lines: two PCa cell lines, LNCaP and PC3 (*Materials and Methods*), and two immortalized prostate epithelial cell lines, LHSAR and RWPE1.

LNCaP is an androgen-sensitive cell line from a lymph node metastasis. In this cell line, suppression of *NUDT11* and *HNF1B* inhibited the proliferation/viability by 63% and 73.2%, respectively (Fig. 2 and Fig. S1) and decreased anchorage-independent colony formation by 86.3% in *NUDT11* and 89.5% in *HNF1B* (Fig. 3 and Fig. S2). *SLC22A3* knockdown in LNCaP cells resulted in decreased proliferation/viability by 74.3% (Fig. 2 and Fig. S1), and had no effect on anchorage-independent colony formation. PC3 is an androgen-independent cell line derived from bone metastasis. Because *SLC22A3* is not expressed in the PC3 cells, results are only reported for *NUDT11* and *HNF1B*. In this cell line, suppression of *NUDT11* and *HNF1B* did not affect cell proliferation/viability (Fig. 2 and Fig. S1). Anchorage-independent colony formation in PC3 cells was inhibited by 56.9% by knockdown of *NUDT11* levels, but no effect was observed by suppression of *HNF1B* (Fig. 3 and Fig. S2). LHSAR is an immortalized prostate epithelial cell line. In LHSAR, suppression of *NUDT11*, *HNF1B*, and *SLC22A3* inhibited the proliferation/viability by 59%, 66.1%, and 71%, respectively (Fig. 2 and Fig. S1). RWPE1 is also an immortalized prostate epithelial cell line. In this cell line, suppression of *NUDT11*, *HNF1B*, and *SLC22A3* resulted in decreased proliferation/viability by 29.2%, 60.9%, and 89.5%, respectively (Fig. 2 and Fig. S1). These observations suggest that expression changes in *NUDT11*, *HNF1B*, and *SLC22A3* influence cellular phenotypes associated with tumor related characteristics.

Discussion

In this study we systematically characterized 12 PCa risk loci both genetically and functionally to gain insight into the genes driving PCa pathogenesis. We implicate *NUDT11*, *HNF1B*, and *SLC22A3* as playing a role in PCa initiation. To our knowledge, this study is one of the largest and most diverse of its kind to bridge the gap between nonprotein-coding risk alleles and their target genes in PCa.

Most GWAS have been performed in populations of European ancestry (1–6, 10). As of this writing, one PCa GWAS has been performed for each of African and Asian ancestries (8, 9). Replication studies for variants discovered in European populations have also been performed in these populations (22–28). The 10q, 17q, and Xp variants (rs10993994, rs4430796, and rs5945619) are associated with PCa risk in men of European, Japanese, and African ancestries (8, 23–25), and these polymorphisms are also associated with *MSMB*, *NCOA4*, *HNF1B*, and *NUDT11* transcripts across all three ethnic groups. In contrast, associations between PCa risk and the 6q locus (rs9364554) have been more variable in Japanese and African men (8, 22–25, 27, 28). Consistent with this observation, a strong association between rs9364554 and *SLC22A3* transcript levels is only seen in the EA population. If this region does confer risk across all populations, then differential linkage disequilibrium structures between the not-yet-discovered causal allele and rs9364554 may exist between the populations. Alternatively, the region may only confer risk in populations of European ancestry because of gene–gene or gene–environment interactions. Fine-mapping studies will help to resolve this issue.

eQTL associations in tumor tissue generally demonstrated attenuated *P* values compared with normal tissue. Several possible reasons exist for this observation. First, a true difference in the eQTL/gene association between normal and tumor tissue may

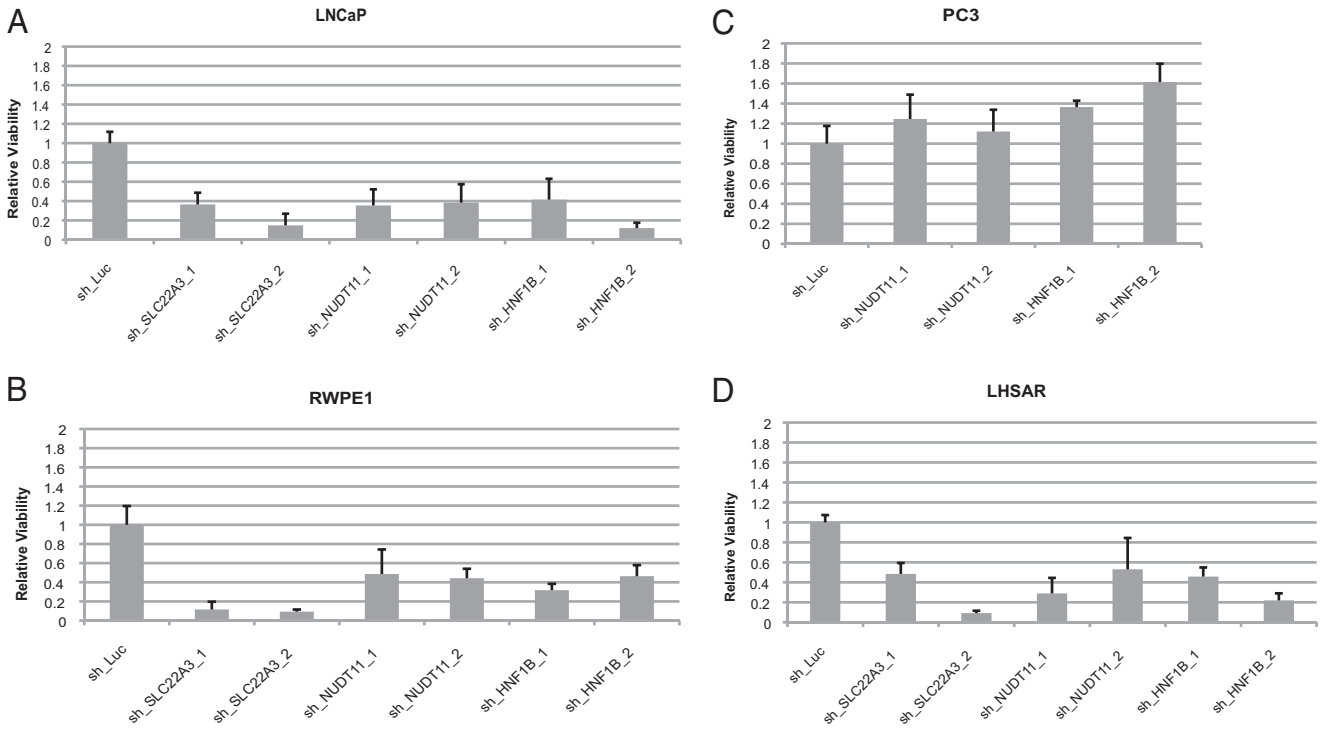

Fig. 2. Suppression of risk loci genes has differential effects on cell viability of prostate cell lines. The indicated genes were suppressed by lentivirally mediated delivery of shRNA and viability measured 6 d after selection using Cell Titer Glo (Promega) assay. Specifically, *SLC22A3*, *NUDT11*, and *HNF1B* were each suppressed with two different shRNAs targeting different sites of the genes (*Materials and Methods*). shRNA targeting luciferase (sh_Luc) was used as a negative control. Data represent the average and SD of at least two independent experiments. Cell viability of LNCaP (*A*), RWPE (*B*), and LHSAR (*D*) cell lines was significantly inhibited by suppression of all targeted candidate genes. Cell viability of PC3 cells (*C*) was not significantly affected by suppression of targeted candidate genes, compared with Luciferase control.

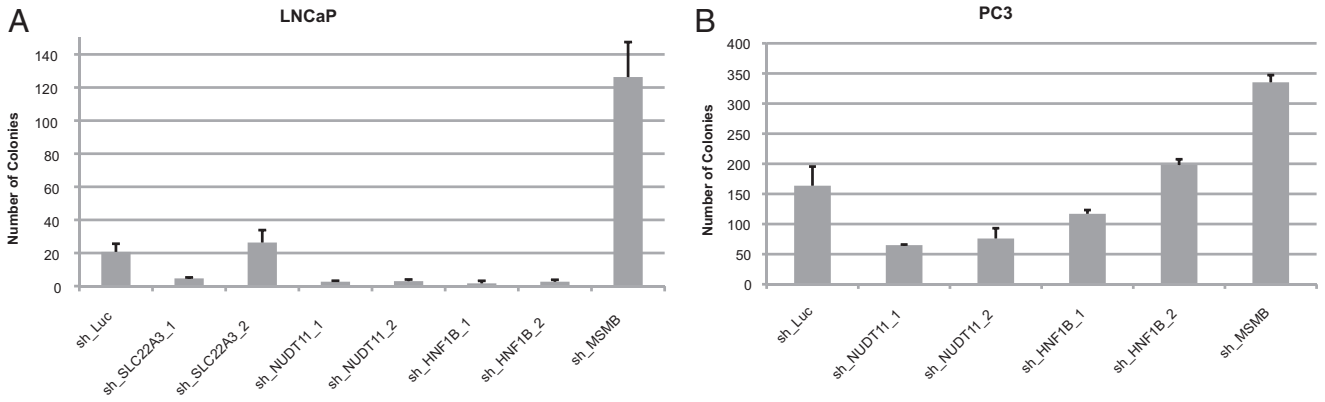

Fig. 3. Suppression of risk loci genes is associated with decreased anchorage-independent growth of prostate cancer cell lines. Soft agar colony formation assays were performed using LNCaP (*A*) (*P* values = 0.03, sh_SLC22A3_1; 0.49, sh_SLC22A3_2; 0.01, sh_NUDT11_1; 0.02, sh_NUDT11_2; 0.01, sh_HNF1B_1; 0.03 sh_HNF1B_2; 0.007, sh_MSMB) and PC3 (*B*) (*P* values = 0.03, sh_NUDT11_1; 0.01, sh_NUDT11_2; 0.08, sh_HNF1B_1; 0.25 sh_HNF1B_2; 0.02, sh_MSMB) cells that were infected with lentivirus expressing the indicated shRNA. Specifically, *SLC22A3*, *NUDT11*, and *HNF1B* were suppressed with two different shRNAs targeting different sites of the genes (*Materials and Methods*). shRNA targeting luciferase (sh_Luc) was used as a negative control. shRNA targeting MSMB (sh_MSMB) was used as a positive control, as previously shown to increase anchorage-independent colony formation. *P* values, paired two-tailed *t* test. Data represent the average and SD of at least two independent experiments.

exist, suggesting that these associations are more important for tumor initiation than maintenance and progression. Second, tumor-acquired changes (e.g., copy-number, methylation, and chromatin alterations) can influence gene expression. To evaluate the impact of somatic copy number alterations, we searched a publicly available copy number database (http://www.broadinstitute.org/tumorscape/pages/portalHome.jsf). Copy-number alterations at the significantly associated transcripts were not prevalent in a set of 92 PCa (Table S5). Finally, the sample size, and therefore statistical power, differs between the number of normal and tumor tissues.

Several risk loci did not demonstrate associations (or were marginally associated) with gene-expression levels (Table S4). Absence of association may imply various possibilities. Gene expression can vary over space and time. Therefore, an association may only be revealed at a particular developmental timepoint or in a distinct cell type (e.g., stem cell). Furthermore, effects on transcript abundance may occur in tissues other than the target tissue (e.g., stroma). The risk alleles may be acting through other types of transcripts, such as short or long noncoding RNAs. It is also conceivable that associations exist with transcripts beyond the 1-Mb interval interrogated here. In addition, most gene-expression studies, including this one, measure steady-state transcript levels; variants may be acting at a point not detected at steady state (e.g., rate of transcription). Finally, effects on gene-expression levels may be subtle, and require even larger sample sizes.

Associations between risk alleles and transcripts are restricted to prostate tissue. In an eQTL database consisting of six other tissue types, we did not find evidence of eQTL/gene associations that we observed in prostate tissue (Table S6) (http://eqtl.uchicago.edu). In fact, prior genome-wide eQTL studies demonstrate tissue specificity of a substantial number of eQTL/gene associations. Large-scale studies, such as the genotype tissue-expression project (https://commonfund.nih.gov/GTEx/), should provide a more comprehensive view of the tissue specificity of risk loci and of the genetics of gene expression in general.

The genetic data implicate *NUDT11*, *HNF1B*, and *SLC22A3* as being regulated by PCa risk variants. The functional data largely corroborate the genetic data; suppression of these genes influence cellular phenotypes associated with tumor-related properties. When *NUDT11* is suppressed, colony formation is significantly decreased in LNCaP and PC3 cell lines and proliferation/viability is affected in LNCaP and LHSAR. When *HNF1B* levels are knocked down, colony formation is significantly decreased in the LNCaP cell line and proliferation/viability is significantly decreased in both the LNCaP and LHSAR cell lines. Because these cell lines reflect different histological origins (LHSAR and RWPE1: normal tissue; LNCaP and PC3: tumor tissue) as well as different stages of disease (LNCaP: androgen-dependent; PC3: androgen-independent), the results may signify that the genes are operative at different disease states. In contrast, the rs9364554 risk allele is associated with decreased expression levels of *SLC22A3*; however, the functional assays reveal that suppression of this transcript leads to reduced viability. Further characterization will be necessary to elucidate the functional consequences of altering *SLC22A3* levels.

NUDT11 is a diphosphoinositol polyphosphate phosphohydrolase. Diphosphoinositol polyphosphates are involved in a variety of biologic functions, including vesicle trafficking, maintenance of cell wall integrity, and mediation of cellular responses to environmental stress in yeast (29). In addition, turnover of diphosphoinositol polyphosphates in ovarian carcinoma cells affects apoptotic processes (30). *HNF1B* is a transcription factor that plays a role in kidney and pancreas development (31, 32). Mutations of *HNF1B* have been described in renal cell carcinoma (33), and epigenetic silencing of the gene has been reported in ovarian cancer, as well as gastric, pancreatic, and colorectal cell lines (34). GWAS have implicated variants in the *HNF1B* locus in diabetes and endometrial cancer risk (35, 36). *SLC22A3* is a member of the solute carrier family 22; it functions as a cation transporter in various organs, including prostate tissue, and plays a role eliminating small organic cations and toxins. *SLC22A3* is inversely correlated with PCa progression, with reduced expression as disease advances (37). Variants in *SLC22A3* are also associated with colorectal cancer (38).

GWAS pave the way to unraveling the genetic basis of disease initiation. Compared with Mendelian disorders, where the majority of pathogenic alleles disrupt protein structure, there is a less-developed genetic code for understanding DNA changes in the nonprotein coding portion of the genome. Future work will continue to clarify the various mechanisms underlying common human traits and may present opportunities to more rationally intervene in prevention and treatment of disease.

Materials and Methods

Prostate Tissue Specimens. Prostate tissue specimens were obtained from the Brigham and Women's Hospital, Tulane University School of Medicine, Weill Cornell Medical College, and the Jikei University School of Medicine on 483 patients who signed an informed content to an Institutional Review Board approved protocol (*SI Materials and Methods*).

Expression Analysis. Each of 12 risk variants was tested for association with expression levels of transcripts in a 1-Mb interval centered on the risk variant. Normal and tumor tissues were analyzed independently. Gene expression levels of selected candidate transcripts were evaluated using the NanoString nCounter platform and TaqMan assays for a subset of genes/individuals. (*SI Materials and Methods*).

Genotyping. Genotyping for all samples was performed by mass spectrometry using the Sequenom iPLEX system (Sequenom Inc, San Diego, CA). (*SI Materials and Methods*).

Statistical Method. For assessing the associations of risk SNPs and transcript abundance of genes we used a semi-parametric approach (generalized estimating equations) to estimate linear regression in both univariate as well multivariable analyses adjusting for ancestry. Criteria for calling an eQTL/taget gene association significant: in individual populations, $P < 0.05$; in the pooled analysis across populations, $P < 0.001$ (*SI Materials and Methods*).

Functional Studies. ***RNAi knockdown.*** Arrayed format RNAi screens we performed as described (39). Cellular viability was assessed using Cell Titer Glo (Promega) and the effect of the candidate short hairpin RNAs (shRNA) on viability are presented as percentage of control shRNAs. Gene knockdown validation was performed 72 h postinfection. Validation experiment was performed on three biological replicates (*SI Materials and Methods*).
Anchorage independent colony formation. Cells were infected with lentivirus encoding shRNAs of interest and selected with puromycin. After 72 h of selection, cells were trypsinized and seeded. Plates were incubated for 3–6 wk to allow colonies to form. Percentage of proliferating cells was calculated taking the average of replicate experiments. Significance was determined using two-tailed *t* test with an alpha level of 0.05 (*SI Materials and Methods*).

Evaluation of Copy Number Alteration in Regions Containing Significant eQTL/Gene Association. We evaluated in the Tumorscape database (http://www.broadinstitute.org/tumorscape/pages/portalHome.jsf), the possible presence of eQTL/gene associations within areas of known somatic copy-number alterations in prostate tumors (*SI Materials and Methods*).

Evaluation of Tissue Specificity of eQTL/Gene Association. We evaluated the eQTL browser from the Pritchard laboratory (http://eqtl.uchicago.edu.) to understand if the four significant associations between risk alleles and transcripts are prostate tissue specific. (*SI Materials and Methods*).

ACKNOWLEDGMENTS. We thank the Center for Cancer Genome Discovery at the Dana-Farber Cancer Institute for technical assistance with genotyping. This work was supported by Grants R01 CA129435 and U19 CA148537 (GAME-ON) from the US National Institutes of Health; the Mayer Foundation (M.L.F.); the H. L. Snyder Medical Foundation (M.L.F.); the Dana-Farber/Harvard Cancer Center Prostate Cancer Specialized Program of Research Excellence (National Cancer Institute Grant 5P50CA90381); the Kohlberg Foundation; the A. David Mazzone Awards Program (M.L.F.); and the Emerald Foundation (M.L.F.). M.L.F. is a Howard Hughes Medical Institute Physician-Scientist Early Career Awardee (www.hhmi.org) and a recipient of a 2009 Claudia Adams Barr award (www.dana-farber.org).

1. Eeles RA, et al.; UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators; PRACTICAL Consortium (2009) Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet* 41:1116–1121.
2. Eeles RA, et al.; UK Genetic Prostate Cancer Study Collaborators; British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators (2008) Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* 40:316–321.
3. Gudmundsson J, et al. (2009) Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet* 41:1122–1126.
4. Gudmundsson J, et al. (2007) Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 39:631–637.
5. Gudmundsson J, et al. (2008) Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet* 40:281–283.
6. Thomas G, et al. (2008) Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* 40:310–315.
7. Yeager M, et al. (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 39:645–649.
8. Takata R, et al. (2010) Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat Genet* 42:751–754.
9. Haiman CA, et al. (2011) Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. *Nat Genet* 43:570–573.
10. Kote-Jarai Z, et al.; UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators, The Australian Prostate Cancer BioResource; PRACTICAL Consortium (2011) Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat Genet* 43:785–791.
11. Cheung VG, Spielman RS (2009) Genetics of human gene expression: Mapping DNA variants that influence gene expression. *Nat Rev Genet* 10:595–604.
12. Nicolae DL, et al. (2010) Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS Genet* 6:e1000888.
13. Moffatt MF, et al. (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 448:470–473.
14. Musunuru K, et al. (2010) From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* 466:714–719.
15. Pomerantz MM, et al. (2010) Analysis of the 10q11 cancer risk locus implicates MSMB and NCOA4 in human prostate tumorigenesis. *PLoS Genet* 6:e1001204.
16. Zhong H, et al. (2010) Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet* 6:e1000932.
17. Harismendy O, et al. (2011) 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. *Nature* 470:264–268.
18. Dimas AS, et al. (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 325:1246–1250.
19. Nica AC, et al.; MuTHER Consortium (2011) The architecture of gene regulatory variation across multiple human tissues: The MuTHER study. *PLoS Genet* 7:e1002003.
20. Emilsson V, et al. (2008) Genetics of gene expression and its effect on disease. *Nature* 452:423–428.
21. Stranger BE, et al. (2005) Genome-wide associations of gene expression variation in humans. *PLoS Genet* 1:e78.
22. Waters KM, et al. (2009) Generalizability of associations from prostate cancer genome-wide association studies in multiple populations. *Cancer Epidemiol Biomarkers Prev* 18:1285–1289.
23. Yamada H, et al. (2009) Replication of prostate cancer risk loci in a Japanese case-control association study. *J Natl Cancer Inst* 101:1330–1336.
24. Chang BL, et al. (2011) Validation of genome-wide prostate cancer associations in men of African descent. *Cancer Epidemiol Biomarkers Prev* 20:23–32.
25. Hooker S, et al. (2010) Replication of prostate cancer risk loci on 8q24, 11q13, 17q12, 19q33, and Xp11 in African Americans. *Prostate* 70:270–275.
26. Okobia MN, Zmuda JM, Ferrell RE, Patrick AL, Bunker CH (2011) Chromosome 8q24 variants are associated with prostate cancer risk in a high risk population of African ancestry. *Prostate* 71:1054–1063.
27. Xu Z, Bensen JT, Smith GJ, Mohler JL, Taylor JA (2011) GWAS SNP Replication among African American and European American men in the North Carolina-Louisiana prostate cancer project (PCaP). *Prostate* 71:881–891.
28. Xu J, et al. (2009) Prostate cancer risk associated loci in African Americans. *Cancer Epidemiol Biomarkers Prev* 18:2145–2149.
29. Dubois E, et al. (2002) In *Saccharomyces cerevisiae*, the inositol polyphosphate kinase activity of Kcs1p is required for resistance to salt stress, cell wall integrity, and vacuolar morphogenesis. *J Biol Chem* 277:23755–23763.
30. Morrison BH, Bauer JA, Kalvakolanu DV, Lindner DJ (2001) Inositol hexakisphosphate kinase 2 mediates growth suppressive and apoptotic effects of interferon-beta in ovarian carcinoma cells. *J Biol Chem* 276:24965–24970.
31. Edghill EL, et al. (2006) Hepatocyte nuclear factor-1 beta mutations cause neonatal diabetes and intrauterine growth retardation: Support for a critical role of HNF-1beta in human pancreatic development. *Diabet Med* 23:1301–1306.
32. Wu G, Bohn S, Ryffel GU (2004) The HNF1beta transcription factor has several domains involved in nephrogenesis and partially rescues Pax8/lim1-induced kidney malformations. *Eur J Biochem* 271:3715–3728.
33. Rebouissou S, et al. (2005) Germline hepatocyte nuclear factor 1alpha and 1beta mutations in renal cell carcinomas. *Hum Mol Genet* 14:603–614.
34. Terasawa K, et al. (2006) Epigenetic inactivation of TCF2 in ovarian cancer and various cancer cell lines. *Br J Cancer* 94:914–921.
35. Gudmundsson J, et al. (2007) Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 39:977–983.
36. Spurdle AB, et al.; Australian National Endometrial Cancer Study Group; National Study of Endometrial Cancer Genetics Group (2011) Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nat Genet* 43:451–454.
37. Tomlins SA, et al. (2007) Integrative molecular concept modeling of prostate cancer progression. *Nat Genet* 39:41–51.
38. Cui R, et al. (2011) Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. *Gut* 60:799–805.
39. Moffat J, et al. (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* 124:1283–1298.