



DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Analysis of the 10q11 Cancer Risk Locus Implicates MSMB and NCOA4 in Human Prostate Tumorigenesis

The Harvard community has made this article openly available.
[Please share](#) how this access benefits you. Your story matters.

Citation	Pomerantz, Mark M., Yashaswi Shrestha, Richard J. Flavin, Meredith M. Regan, Kathryn L. Penney, Lorelei A. Mucci, Meir J. Stampfer, and et al. 2010. Analysis of the 10q11 cancer risk locus implicates MSMB and NCOA4 in human prostate tumorigenesis. PLoS Genetics 6(11): e1001204.
Published Version	doi://10.1371/journal.pgen.1001204
Accessed	February 19, 2015 7:44:07 AM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:5343425
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

Analysis of the 10q11 Cancer Risk Locus Implicates *MSMB* and *NCOA4* in Human Prostate Tumorigenesis

Mark M. Pomerantz^{1,2,9}, Yashaswi Shrestha^{1,3,4,9}, Richard J. Flavin⁵, Meredith M. Regan⁶, Kathryn L. Penney⁷, Lorelei A. Mucci^{7,8}, Meir J. Stampfer^{7,8,9}, David J. Hunter^{7,8,9}, Stephen J. Chanock¹⁰, Eric J. Schafer^{1,3,4}, Jennifer A. Chan¹¹, Josep Taberero¹², José Baselga¹², Andrea L. Richardson¹³, Massimo Loda^{1,3,4,5,13}, William K. Oh¹⁴, Philip W. Kantoff^{1,2}, William C. Hahn^{1,2,3,4}, Matthew L. Freedman^{1,3,4*}

1 Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, United States of America, **2** Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States of America, **3** Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts, United States of America, **4** Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **5** Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, **6** Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts, United States of America, **7** Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America, **8** Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States of America, **9** Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, United States of America, **10** Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **11** Department of Pathology and Laboratory Medicine, University of Calgary, Calgary, Alberta, Canada, **12** Vall d'Hebron Institute of Oncology, Vall d'Hebron University Hospital, Barcelona, Spain, **13** Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States of America, **14** Tisch Cancer Institute, Mount Sinai School of Medicine, New York, New York, United States of America

Abstract

Genome-wide association studies (GWAS) have established a variant, rs10993994, on chromosome 10q11 as being associated with prostate cancer risk. Since the variant is located outside of a protein-coding region, the target genes driving tumorigenesis are not readily apparent. Two genes nearest to this variant, *MSMB* and *NCOA4*, are strong candidates for mediating the effects of rs10993994. In a cohort of 180 individuals, we demonstrate that the rs10993994 risk allele is associated with decreased expression of two *MSMB* isoforms in histologically normal and malignant prostate tissue. In addition, the risk allele is associated with increased expression of five *NCOA4* isoforms in histologically normal prostate tissue only. No consistent association with either gene is observed in breast or colon tissue. In conjunction with these findings, suppression of *MSMB* expression or *NCOA4* overexpression promotes anchorage-independent growth of prostate epithelial cells, but not growth of breast epithelial cells. These data suggest that germline variation at chromosome 10q11 contributes to prostate cancer risk by influencing expression of at least two genes. More broadly, the findings demonstrate that disease risk alleles may influence multiple genes, and associations between genotype and expression may only be observed in the context of specific tissue and disease states.

Citation: Pomerantz MM, Shrestha Y, Flavin RJ, Regan MM, Penney KL, et al. (2010) Analysis of the 10q11 Cancer Risk Locus Implicates *MSMB* and *NCOA4* in Human Prostate Tumorigenesis. *PLoS Genet* 6(11): e1001204. doi:10.1371/journal.pgen.1001204

Editor: James M. Ford, Stanford University School of Medicine, United States of America

Received: July 14, 2010; **Accepted:** October 13, 2010; **Published:** November 11, 2010

This is an open-access article distributed under the terms of the Creative Commons Public Domain declaration which stipulates that, once placed in the public domain, this work may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose.

Funding: MLF is a Howard Hughes Medical Institute Physician-Scientist Early Career Awardee (www.hhmi.org), a recipient of a 2006 Doris Duke Clinical Scientist Development Award (www.ddcf.org), and a recipient of a 2009 Claudia Adams Barr award (www.dana-farber.org). This work was supported by grants from the US National Institutes of Health (R01 CA129435 to MLF), the Mayer Foundation (to MLF), the H.L. Snyder Medical Foundation (to MLF), the Dana-Farber/Harvard Cancer Center Prostate Cancer SPOR (National Cancer Institute Grant No. 5P50CA90381), and the Emerald Foundation, Inc. (to MLF). MMP is supported by the William L. Edwards - Prostate Cancer Foundation Young Investigator Award (www.prostatecancerfoundation.org). WCH receives support from the Prostate Cancer Foundation and the John Snyder Medical Foundation (snyderdmf.org). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: freedman@broadinstitute.org

☞ These authors contributed equally to this work.

Introduction

Variation at rs10993994 on chromosome 10q11 is associated with prostate cancer risk [1–5]. The risk polymorphism is located at the telomeric end of a 50 kilobase (kb) linkage disequilibrium block and is within 60 base pairs (bp) of the transcription start site of beta-microseminoprotein (*MSMB*). *MSMB* has been characterized as a tumor suppressor [6], and lower levels of its product, PSP94, are associated with more aggressive forms of prostate cancer [7]. *MSMB* has therefore been a target of recent

investigation into the mechanism of chromosome 10q-associated risk. Reporter assays demonstrate that plasmids carrying the rs10993994 risk allele (T) significantly decrease luciferase activity compared with the wild-type allele (C) [8,9]. In addition, in 19 cancer cell lines of various tissue types expressing *MSMB*, those carrying the TT genotype have decreased *MSMB* expression relative to those carrying a C allele [9]. However, no study has definitively linked *MSMB* expression to risk allele status in human prostate tissue. A second gene, nuclear receptor co-activator 4 (*NCOA4*, also known as *ARA70*), is a ligand-dependent androgen

Author Summary

Family history has long been recognized as an important risk factor for prostate cancer. Beginning in 2006, researchers have identified several genetic variants that are associated with prostate cancer risk. Intriguingly, the majority of prostate cancer risk variants do not reside in genes. Determining the genes involved in the development of disease, therefore, has proved challenging. In this study we interrogate a known prostate cancer risk polymorphism on chromosome 10—rs10993994. We report that this variant is significantly associated with the RNA expression levels of two genes—*MSMB* and *NCOA4*. When these expression changes are modeled in a cell line, prostate cells that were previously non-tumorigenic acquire a property known as anchorage independence, a characteristic of cancer cells. Notably, the prostate risk variant is *not* associated with expression or functional changes in breast or colon cells. In addition, the effects are most pronounced in normal rather than tumor prostate tissue. Overall, these findings help define the genes driving prostate cancer risk at chromosome 10. More generally, the discoveries demonstrate the importance of considering several target genes, as well as the importance of cellular context (tissue type and histological state), in future analyses of other genetic risk regions.

receptor co-activator [10,11] and is within 16 kb telomeric of rs10993994. Given its proximity to the risk variant and its activity in the prostate gland [12], *NCOA4* has also been considered a candidate gene involved in the mechanism of disease risk [8].

Gene expression is a heritable trait [13–16] and represents a powerful avenue for connecting risk variants with their target genes. Studies have demonstrated that variation at intergenic or intronic disease-associated loci can act through gene regulatory mechanisms [17–20]. Because regulatory elements can interact with many genes [21], and since both *MSMB* and *NCOA4* are strong candidates for prostate cancer risk, we evaluated the relationship between risk allele status and transcript abundance of these genes across both normal and tumor prostate tissues. We also tested the functional consequences of altering the expression levels of these candidate genes in immortalized prostate epithelial cells.

Results

A total of 180 individuals of European ancestry were genotyped for the rs10993994 polymorphism, and *MSMB* and *NCOA4* mRNA levels were quantified in tissue isolated from radical prostatectomy surgical specimens. Samples were derived from two cohorts— the Gelb Center at Dana-Farber Cancer Institute (DFCI) (N = 121 – histologically normal and tumor prostate tissue) and the Physicians' Health Study (PHS) [22] (N = 59 – prostate tumor tissue only). In the DFCI cohort, transcript levels were measured in both normal and tumor prostate tissue using a quantitative competitive PCR strategy (Methods). Two *MSMB* and five *NCOA4* isoforms annotated in the Ensembl database (build 52) were evaluated (Figure 1A). Each isoform of *MSMB* and *NCOA4* was expressed in both normal and tumor prostatic tissue. Expression levels of *MSMB* and *NCOA4* transcripts were significantly higher in normal compared with tumor tissue ($p < 0.0001$), consistent with previously published reports [23–25]. In the PHS cohort, only tumor tissue was isolated from radical prostatectomy specimens. For this cohort the probes used to measure expression captured both *MSMB* isoforms and all but one of the *NCOA4* isoforms.

The T (risk) allele is significantly associated with transcript levels of both *MSMB* and *NCOA4* in histologically normal prostate tissue (N = 84). While the T allele is associated with decreased expression of both *MSMB* isoforms (p-value range, 0.0033–0.0042), it is associated with *increased* expression of all five isoforms of *NCOA4* (p-value range, 6.7×10^{-7} –0.0055) (Figure 1B). In tumor tissue, *MSMB* retains its association with risk allele status (p-value range, 0.016–0.053, Figure 1C, Figure S1). *NCOA4* expression levels, however, are no longer associated with genotype in tumor tissue ($p > 0.30$, Figure 1C, Figure S1). The associations are specific to the genes in this region. Expression levels of *TMM23*, the next closest gene to the risk locus and 1.4 kb telomeric to *NCOA4*, are not correlated with genotype status in either normal or tumor tissue. ($p > 0.30$, Figure S2).

Because rs10993994 is *not* a risk allele in colon or breast cancers, we reasoned that the association between genotype and disease-relevant genes may be specific to prostate tissue and not observed in other tissue types. If an association between genotype and expression of *MSMB* or *NCOA4* were observed in tissue other than prostate, then that gene may be less likely to be involved in prostate cancer risk. *MSMB* and *NCOA4* expression levels were measured in histologically normal colon (N = 72) and breast (N = 56) tissue samples. While breast tissue expresses both genes, colon tissue only expresses *NCOA4*. Unlike prostate tissue, neither breast nor colon tissue demonstrates convincing or consistent associations with genotype across isoforms (Figure S3).

To evaluate the functional implications of the genetic findings, we tested the effect of increasing *NCOA4* and suppressing *MSMB* expression levels in immortalized prostate epithelial cells (LHSAR) [26]. Specifically, we assessed the ability of *NCOA4* and *MSMB* to promote anchorage-independent growth, a phenotype strongly associated with cell transformation [27–30]. Suppression of *MSMB* expression in LHSAR cells led to a significant increase in anchorage-independent colony growth (p-values 0.0023–0.0001; Figure 2A, Figure S4). Overexpression of *NCOA4* in LHSAR cells also resulted in robust anchorage-independent colony growth (p-value 0.0074; Figure 2B, Figure S4). To assess whether these alterations were specific for prostate epithelial cells, similar functional studies were performed in immortalized human mammary epithelial cells [31]. In contrast to what was observed in prostate epithelial cells, manipulating expression levels of *MSMB* or *NCOA4* did not result in any consistently significant change in the anchorage-independent growth in mammary epithelial cells (Figure S5). Together, these observations implicate a role for both *NCOA4* and *MSMB* in the transformation of prostate epithelial cells.

Discussion

Genetic data presented here demonstrate that the chromosome 10q11 prostate cancer risk locus is associated with decreased levels of *MSMB* and increased levels of *NCOA4* RNA expression. Strikingly, the functional data fully corroborate the genetic data. When *MSMB* is knocked down or *NCOA4* is overexpressed in immortalized prostate epithelial cells, the cells become anchorage independent. Our data suggest that both *MSMB* and *NCOA4* mediate prostate tumorigenesis, and this study is the first, to our knowledge, to implicate these genes in actual human prostate tissue. As expected in a cohort comprised of subjects who underwent radical prostatectomy, a large majority of individuals included in the analysis were diagnosed with low- or intermediate-risk prostate cancer. Despite a relatively homogenous cohort, the results presented here are likely generalizable to most prostate cancer cases since rs10993994 appears to confer risk for all levels of prostate cancer aggressiveness [1,3,32,33,34].

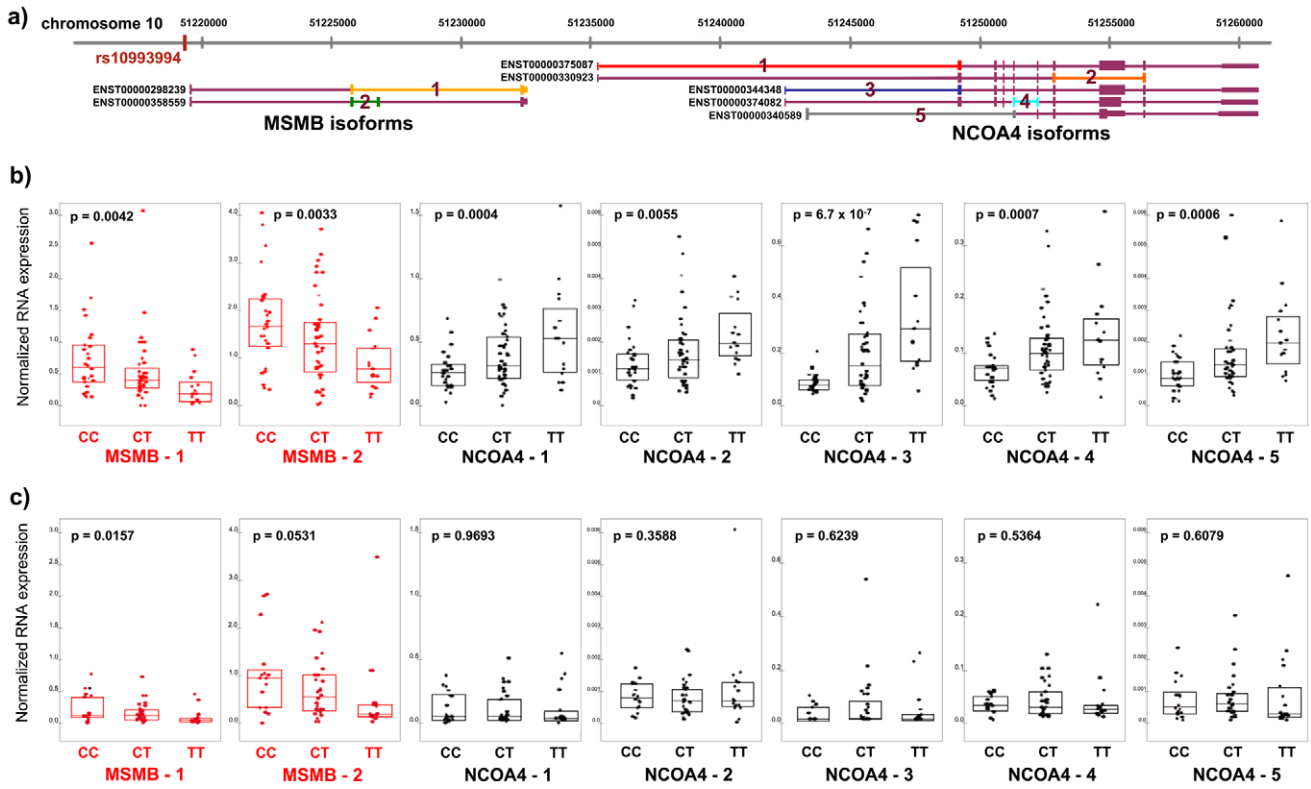


Figure 1. RNA expression of *MSMB* and *NCOA4* in normal and tumor prostate tissue by rs10994994 genotype. A. Chromosome 10q11 with isoforms of *MSMB* and *NCOA4* (Ensembl build 52). Primers for competitive PCR were designed to cross exon-exon boundaries depicted by colored lines 1–2 (*MSMB*) and 1–5 (*NCOA4*). B. Expression in histologically normal prostate tissue (n = 84). Each point represents absolute RNA expression for one individual, normalized to three housekeeping genes. The top and bottom of the boxes within each graph represent the upper and lower quartiles for expression at each genotype. The band inside each box marks the median value. P-value for each graph denotes the significance for association between expression and genotype. C. Expression in prostate tumor tissue in the Dana-Farber Cancer Institute series (n = 61). doi:10.1371/journal.pgen.1001204.g001

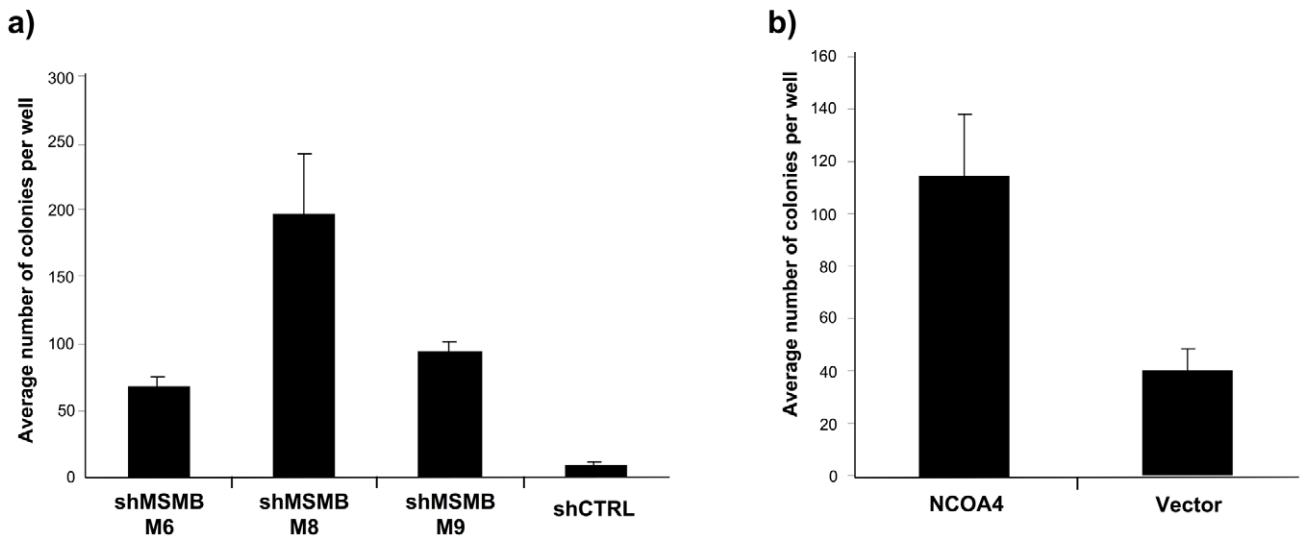


Figure 2. Suppressing *MSMB* or overexpressing *NCOA4* is associated with increased anchorage-independent growth of prostate epithelial cells. A. Effects of suppressing *MSMB* with three independent shRNAs in LHSAR cells (p-values 0.0001; M6, 0.0023; M8, and 0.0001; M9). The increase in anchorage-independent growth inversely correlates with the degree of *MSMB* suppression (Figure S4). B. Anchorage-independent growth of LHSAR cells overexpressing *NCOA4* and a control vector (p = 0.0074). doi:10.1371/journal.pgen.1001204.g002

Similar to the 10q11 risk allele, other disease risk loci have been shown to affect expression of more than one gene [17,34]. A variant associated with systemic lupus erythematosus at chromosome 8p23, for example, is associated with increased expression of one gene (*BLK*) and decreased expression of another (*C8orf13*) in B cell lines [34]. As more genes underlying complex traits are discovered, it may be that certain risk alleles mediate their effects through multiple genes, or alternatively, that two risk variants in tight linkage disequilibrium influence separate genes. The findings at 10q11 highlight the importance of considering multiple genes when analyzing GWAS results.

The findings at 10q11 also underscore the importance of evaluating risk loci in a tissue-specific context [35]. It is hypothesized that a fraction of non-protein coding risk alleles will alter disease risk by regulating gene activity, and these variants may exert their effects in a specific genetic and epigenetic context [21,36]. In the present study, the 10q11 risk variant is associated with transcript levels of *MSMB* and *NCOA4* in primary prostate tissue. In contrast, no convincing or consistent association is observed in colon or breast tissue. Similarly, alteration of *MSMB* and *NCOA4* expression significantly affects anchorage-independence of prostate but not breast epithelial cells. These findings may have implications for future studies attempting to connect risk alleles with target gene(s). Evaluation of GWAS findings will focus, in part, on identifying the genes targeted by risk alleles, as these are the genes likely to drive the trait under study. Our findings suggest that this type of analysis should include evaluation of the tissue at risk for disease, although it is entirely plausible that variants associated with a particular disease may manifest their effects in tissues other than target tissue.

Notably, associations between rs10993994 genotype and expression of *MSMB* and *NCOA4* are observed in histologically normal prostate tissue, whereas in tumor tissue an association is detected with only *MSMB* (albeit at an attenuated level relative to the normal tissue). It is conceivable that increased expression of *NCOA4* is associated with tumor initiation, as reflected by its association with risk in solely normal tissue, while decreased expression of *MSMB* is associated with both tumor initiation and maintenance or progression. More studies, however, will be necessary before a general principle emerges.

Cellular context also appears to be an important determinant in the functional analysis of candidate risk genes. This is illustrated by comparing data in the present study to previous work involving *NCOA4*. The characteristics of two *NCOA4* isoforms, alpha and beta, have been studied in functional assays. Upregulation of *NCOA4*beta increases anchorage-independent growth [11], while overexpression of *NCOA4*alpha inhibits growth in LNCaP cells [37]. Functional analysis of immortalized prostate epithelial cells presented here demonstrates *increased* colony growth in the setting of an overexpressed alpha isoform (Figure 2). Distinctions between the cell lines used in these studies may account for these divergent results. LNCaP cells are derived from metastatic prostate lesions. Immortalized prostate epithelial cells, on the other hand, are not tumorigenic and differentiate in the presence of androgen [26], suggesting that these cells are more closely related to normal prostate epithelial cells.

The data presented here suggest that tissue type (in this case, prostate versus non-prostate tissue) and cellular states (i.e., normal versus tumor) are likely important factors in the evaluation of complex trait loci. Chromatin context and differential use of gene regulatory elements across tissues and disease states may be the basis for expression effects specific to normal prostate tissue [36]. It can be difficult to accurately model these effects outside of the particular genetic and epigenetic context of specific tissues. Luciferase reporter

assays, for example, are often utilized to define a relationship between a polymorphism and a gene. Reporter assays cannot, however, detect situations where a risk variant is associated with opposing transcriptional effects on two loci, as observed with rs10993994. An alternative explanation for the effects on the two transcripts is that two separate variants in linkage disequilibrium are responsible for the different transcriptional effects.

In contrast to Mendelian diseases, where resequencing protein-coding regions often reveals the causal gene, common complex trait alleles often occur outside of protein-coding regions. As is the case at 10q11 and other loci [38,39], these alleles may be associated with expression of nearby and/or distal candidate genes. There are also situations in which associations with strong candidate genes, however, cannot be established by measuring steady-state expression at a single point in time [19]. In order to better understand the gene(s) involved in complex trait pathogenesis, experiments will need to integrate and to account for the genetic and epigenetic contexts of the particular tissue type and cellular state.

Materials and Methods

Ethics statement

This study was conducted according to the principles expressed in the Declaration of Helsinki. The study was approved by the Institutional Review Board of Dana-Farber Cancer Institute. All patients provided written informed consent for the collection of samples and subsequent analysis.

Cohorts and RNA isolation

A total of 180 patients treated with radical prostatectomy (RP) for prostate cancer and 92 patients treated for colon cancer consented to provide tissue. Additionally, histologically normal breast tissue from 56 from women undergoing cosmetic reduction surgery was analyzed in the present study.

Fresh frozen radical prostatectomy specimens were available from 121 subjects at the Dana-Farber Cancer Institute (DFCI) and Brigham and Women's Hospital (Boston, MA) and were reviewed by a pathologist (J.C. or R.F.). Over 95% of the patients in the cohort were diagnosed with Gleason 6 or Gleason 7 disease and median PSA was 5.1 ng/ml. Areas of tumor consisted of >60% tumor cells and areas of benign tissue consisted of >80% non-neoplastic epithelial cells at least 5 mm away from any tumor focus. Biopsy cores of fresh frozen tissue were processed for RNA extraction using a modified Qiagen Allprep DNA/RNA protocol.

Archival FFPE blocks were available for 59 men with prostate cancer enrolled in the Physicians' Health Study (PHS) [22,40]. These men were diagnosed with prostate cancer between 1983 and 2003 and treated by radical prostatectomy. RNA were extracted from paraffin-embedded tumor tissue as described previously [40]. Areas of tumor consisted of >90% tumor cells.

Fresh frozen colorectal cancer tissue samples were reviewed by a pathologist (J.C.) and areas of benign tissue were selected where 80% of cells consisted of non-neoplastic epithelium. RNA was extracted using a modified Qiagen Allprep DNA/RNA protocol. Fresh frozen normal breast tissue samples were reviewed to identify tissue blocks containing >40% normal epithelial cells. RNA was extracted using a modified Qiagen RNeasy protocol.

Ethnicity was self-reported by most, but not all, subjects. Subjects in the DFCI cohort of unknown ancestry were genotyped for 59 ancestry-informative SNPs in order to ascertain ethnicity. The marker set primarily captured ancestral differences between European and African ancestries (D. Reich, personal communication). Five samples found to be from subjects of African ancestry were excluded from analysis.

The human tissues analyzed in this study were from patients treated at Brigham and Women's Hospital, Dana-Farber Cancer Institute or Vall d'Hebron University Hospital in Barcelona, Spain, all of whom provided informed consent. The study was approved by the Institutional Review Board at Dana-Farber Cancer Institute.

Expression analysis

cDNA was prepared for expression analysis using Invitrogen SuperScript III Reverse Transcription kit. DFCI prostate samples, colon samples and breast samples were analyzed via competitive RT-PCR using Sequenom iPLEX matrix-assisted laser desorption/ionization (MALDI)-time of flight mass spectrometry technology. Expression levels of two *MSMB* isoforms and five *NCOA4* isoforms were measured. These splice variants represent all isoforms reported in Ensembl genebuild 52. RNA expression of *TMM23* and three housekeeping genes (*ACTB*, *MYL6* and *RPL13A*) were also measured. Primer, probe and competitor oligo sequences are available upon request. Reactions were performed in quadruplicate using 8 serial dilutions of competitor, and the EC50 was calculated using QGE Analyzer software (Sequenom). The PHS subgroup was analyzed using Illumina cDNA-mediated Annealing, Selection, Extension and Ligation (DASL) expression assay (Illumina).

Expression data analysis

A gene expression normalization factor was calculated using the geometric mean of expression level of the three housekeeping genes. Linear regression was used to assess whether expression levels increased (or decreased) as the number of T-alleles of rs10993994 increased; for prostate tissue, differences between prostate tumor and normal tissue levels were also assessed and random effects linear regression was used to account for within-sample correlation of tumor/normal pairs.

Genotyping

Genotyping of DNA from each subject was carried out using Sequenom iPLEX matrix-assisted laser desorption/ionization (MALDI)-time of flight mass spectrometry technology.

Cell culture

LHSAR: Prostate epithelial cells (PrECs) immortalized with hTERT, SV40 Large T and small t antigens and overexpressing androgen receptor were grown in PREGM (Lonza CC-3166). HMLE: Human mammary epithelial cells immortalized with hTERT, Large T and small t antigens were grown in MEGM (Lonza CC-3150). All growth media were supplemented with 100 ug/ml Penicillin/Streptomycin.

Soft agar colony formation assay

The bottom layer contained 0.6% agar (Sigma A5431) in DMEM and 8% FBS. The top layer contained 0.3% agar in PREGM or MEGM for LHSAR or HMLE respectively. Fifteen thousand cells were seeded in the top agar layer in triplicate wells of a 6 well plate. Colonies were counted from 2 to 6 weeks post-seeding. Image of each well was taken at a 6× magnification and analyzed with Image J software. Colonies that were 50 sq. pixels or larger were counted.

Quantitative RT-PCR

Qiagen RNeasy kit was used for RNA extraction. Reverse transcription was carried out with Clontech Advantage RT-to-PCR kit while the quantitative PCR was carried out using SYBR Green Master Mix (Applied Biosystems). Two sets of *NCOA4* and one set of *MSMB* primers were used:

- 1) *NCOA4* exon 3-4 - Forward CAGCAGCTCTACTCGT-TATTGG
Reverse TCTCCAGGCACACAGAGACT
- 2) *NCOA4* exon 5-6 - Forward CTCTCAAAACCATTCAA-ATTCCT
Reverse CTCTGGCATGGAGAT-ACAGC
- 3) *MSMB* exon 1-2 - Forward GCTTATCACAAATGAATG-TTCTCCT
Reverse AATCTCCTGGAACCTCC-TCA

Expression constructs

NCOA4 expression constructs were received from the human ORFeome V5.1 and cloned into pWZL-Neo retroviral expression vector. Retrovirus production, infection and selection were carried out as described previously [41].

RNA interference

Short hairpins in pLKO.1 lentiviral constructs were received from the RNAi Consortium (TRC). Lentivirus production and infection were carried out as described previously [42].

- 1) Hairpin shMSMB M6 TRC Clone ID TRCN0000147237
Target sequence GTTCTGTCAGTGAATGGATAA
- 2) Hairpin shMSMB M8 TRC Clone ID TRCN0000146396
Target sequence CACCTTCGTGACTTTATGCAA
- 3) Hairpin shMSMB M9 TRC Clone ID TRCN0000146343
Target sequence CAAAGGAAACAAACACCCAAT

Supporting Information

Figure S1 RNA expression of *MSMB* and *NCOA4* in tumor prostate tissue by genotype at rs10994994. Expression in prostate tumor tissue in the Physicians' Health Study series (n = 59). Each point represents absolute RNA expression for one individual. The top and bottom of the boxes within each graph represent the upper and lower quartiles for expression at each genotype. The band inside each box marks the median value. P-value for association with *MSMB* expression, 0.0073. P-value for association with *NCOA4* expression, 0.2671. Found at: doi:10.1371/journal.pgen.1001204.s001 (0.50 MB TIF)

Figure S2 Expression of *TMM23* at chromosome 10q11 is not associated with genotype at rs10993994 in prostate tissue. A. Expression in histologically normal prostate tissue (n = 84, p = 0.3459). B. Expression in prostate tumor tissue in the Dana-Farber Cancer Institute series (n = 61, p = 0.9939). Found at: doi:10.1371/journal.pgen.1001204.s002 (9.46 MB TIF)

Figure S3 *MSMB* and *NCOA4* expression in breast and colon epithelial tissue. Expression of *MSMB* and *NCOA4* isoforms are not consistently associated with genotype at rs10993994 in breast or colon epithelial tissue. A. Expression in histologically normal breast tissue (n = 56). P-values for association between genotype and expression included within each graph. B. Expression in histologically normal colon tissue (n = 72). Found at: doi:10.1371/journal.pgen.1001204.s003 (9.47 MB TIF)

Figure S4 Quantitative RT-PCR for the expression of *MSMB* and *NCOA4* in LHSAR cells. A. *MSMB* suppression with three

individual shRNAs (shMSMB M6, M8 and M9). B. NCOA4 overexpression.

Found at: doi:10.1371/journal.pgen.1001204.s004 (9.46 MB TIF)

Figure S5 Anchorage-independent growth in human mammary epithelial cells. MSMB and NCOA4 expression levels do not consistently affect anchorage-independent growth of human mammary epithelial cells (HMLE). A. Suppression of MSMB by three independent shRNAs (p -values 0.0949, 0.7482, and 0.0197). B. NCOA4 overexpression ($p = 0.1012$).

References

- Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, et al. (2008) Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* 40: 316–321.
- Kote-Jarai Z, Easton DF, Stanford JL, Ostrander EA, Schleutker J, et al. (2008) Multiple novel prostate cancer predisposition loci confirmed by an international study: the PRACTICAL Consortium. *Cancer Epidemiol Biomarkers Prev* 17: 2052–2061.
- Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, et al. (2008) Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* 40: 310–315.
- Waters KM, Le Marchand L, Kolonel LN, Monroe KR, Stram DO, et al. (2009) Generalizability of associations from prostate cancer genome-wide association studies in multiple populations. *Cancer Epidemiol Biomarkers Prev* 18: 1285–1289.
- Zheng SL, Sun J, Wiklund F, Gao Z, Stattin P, et al. (2009) Genetic variants and family history predict prostate cancer similar to prostate-specific antigen. *Clin Cancer Res* 15: 1105–1111.
- Reeves JR, Dulude H, Panchal C, Daigneault L, Ramnani DM (2006) Prognostic value of prostate secretory protein of 94 amino acids and its binding protein after radical prostatectomy. *Clin Cancer Res* 12: 6018–6022.
- Shukeir N, Arakelian A, Kadhim S, Garde S, Rabbani SA (2003) Prostate secretory protein PSP-94 decreases tumor growth and hypercalcemia of malignancy in a syngenic in vivo model of prostate cancer. *Cancer Res* 63: 2072–2078.
- Chang BL, Cramer SD, Wiklund F, Isaacs SD, Stevens VL, et al. (2009) Fine mapping association study and functional analysis implicate a SNP in MSMB at 10q11 as a causal variant for prostate cancer risk. *Hum Mol Genet* 18: 1368–1375.
- Lou H, Yeager M, Li H, Bosquet JG, Hayes RB, et al. (2009) Fine mapping and functional analysis of a common variant in MSMB on chromosome 10q11.2 associated with prostate cancer susceptibility. *Proc Natl Acad Sci U S A* 106: 7933–7938.
- Niu Y, Yeh S, Miyamoto H, Li G, Altuwajri S, et al. (2008) Tissue prostate-specific antigen facilitates refractory prostate tumor progression via enhancing ARA70-regulated androgen receptor transactivation. *Cancer Res* 68: 7110–7119.
- Peng Y, Li CX, Chen F, Wang Z, Ligr M, et al. (2008) Stimulation of prostate cancer cellular proliferation and invasion by the androgen receptor co-activator ARA70. *Amer J Pathol* 172: 225–235.
- Hu YC, Yeh S, Yeh SD, Sampson ER, Huang J, et al. (2004) Functional domain and motif analyses of androgen receptor coregulator ARA70 and its differential expression in prostate cancer. *J Biol Chem* 279: 33438–33446.
- Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, et al. (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet* 33: 422–425.
- Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, et al. (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422: 297–302.
- Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, et al. (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature* 430: 743–747.
- Jia Z, Xu S (2007) Mapping quantitative trait loci for expression abundance. *Genetics* 176: 611–623.
- Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, et al. (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 448: 470–473.
- Meyer KB, Maia AT, O'Reilly M, Teschendorff AE, Chin SF, et al. (2008) Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol* 6: e108. doi:10.1371/journal.pbio.0060108.
- Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, et al. (2009) The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet* 41: 882–884.
- Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, et al. (2009) The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 41: 885–890.
- Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447: 799–816.
- Gann PH, Hennekens CH, Stampfer MJ (1995) A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. *Jama* 273: 289–294.
- Li P, Yu X, Ge K, Melamed J, Roeder RG, et al. (2002) Heterogeneous expression and functions of androgen receptor co-factors in primary prostate cancer. *Amer J Pathol* 161: 1467–1474.
- Vanaja DK, Chevillet JC, Iturria SJ, Young CY (2003) Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. *Cancer Res* 63: 3877–3882.
- Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, et al. (2004) Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A* 101: 811–816.
- Berger R, Febbo PG, Majumder PK, Zhao JJ, Mukherjee S, et al. (2004) Androgen-induced differentiation and tumorigenicity of human prostate epithelial cells. *Cancer Res* 64: 8867–8875.
- Freedman VH, Shin SI (1974) Cellular tumorigenicity in nude mice: correlation with cell growth in semi-solid medium. *Cell* 3: 355–359.
- Shin SI, Freedman VH, Rissler R, Pollack R (1975) Tumorigenicity of virus-transformed cells in nude mice is correlated specifically with anchorage independent growth in vitro. *Proc Natl Acad Sci U S A* 72: 4435–4439.
- Bouck N, di Mayorca G (1976) Somatic mutation as the basis for malignant transformation of BHK cells by chemical carcinogens. *Nature* 264: 722–727.
- Spandidos DA, Siminovich L (1977) Transfer of anchorage independence by isolated metaphase chromosomes in hamster cells. *Cell* 12: 675–682.
- Elenbaas B, Spirio L, Koerner F, Fleming MD, Zimonjic DB, et al. (2001) Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev* 15: 50–65.
- Camp NJ, Farnham JM, Wong J, Christensen GB, Thomas A, et al. (2009) Replication of the 10q11 and Xp11 prostate cancer risk variants: results from a Utah pedigree-based study. *Cancer Epidemiol Biomarkers Prev* 18: 1290–1294.
- Fitzgerald LM, Kwon EM, Koopmeiners JS, Salinas CA, Stanford JL, et al. (2009) Analysis of recently identified prostate cancer susceptibility loci in a population-based study: associations with family history and clinical features. *Clin Cancer Res* 15: 3231–3237.
- Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, et al. (2008) Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med* 358: 900–909.
- Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, et al. (2008) Genetics of gene expression and its effect on disease. *Nature* 452: 423–428.
- Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, et al. (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 325: 1246–1250.
- Ligr M, Li Y, Zou X, Daniels G, Melamed J, et al. (2010) Tumor suppressor function of androgen receptor coactivator ARA70 alpha in prostate cancer. *Am J Pathol* 176: 1891–900.
- Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, et al. (2007) Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet* 3: e58. doi:10.1371/journal.pgen.0030058.
- Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M (2009) Mapping complex disease traits with global gene expression. *Nat Rev Genet* 10: 184–194.
- Setlur SR, Mertz KD, Hoshida Y, Demicheli F, Lupien M, et al. (2008) Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. *J Natl Cancer Inst* 100: 815–825.
- Boehm JS, Hession MT, Bulmer SE, Hahn WC (2005) Transformation of human and murine fibroblasts without viral oncoproteins. *Mol Cell Biol* 25: 6464–6474.
- Moffatt J, Gruenberg DA, Yang X, Kim SY, Kloepfer AM, et al. (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* 124: 1283–1298.

Found at: doi:10.1371/journal.pgen.1001204.s005 (9.46 MB TIF)

Author Contributions

Conceived and designed the experiments: MMP YS WCH MLF. Performed the experiments: MMP YS RJF EJS JAC. Analyzed the data: MMP YS RJF MMR MLF. Contributed reagents/materials/analysis tools: KLP LAM MJS DJH SJC JT JB ALR ML WKO PWK. Wrote the paper: MMP YS MLF.