



Published in final edited form as:

*Cancer Res.* 2009 July 1; 69(13): 5568–5574. doi:10.1158/0008-5472.CAN-09-0387.

## Evaluation of the 8q24 prostate cancer risk locus and *MYC* expression

Mark M. Pomerantz<sup>1</sup>, Christine A. Beckwith<sup>1</sup>, Meredith M. Regan<sup>2</sup>, Stacia K. Wyman<sup>3</sup>, Gyorgy Petrovics<sup>4</sup>, Yongmei Chen, Dorota J Hawksworth<sup>4</sup>, Fredrick R Schumacher<sup>5,6</sup>, Lorelei Mucci<sup>5,6</sup>, Kathryn L. Penney<sup>5</sup>, Meir J. Stampfer<sup>5,6</sup>, Jennifer A. Chan<sup>6</sup>, Kristin G. Ardlie<sup>7</sup>, Brian R. Fritz<sup>3</sup>, Rachael K. Parkin<sup>3</sup>, Daniel W. Lin<sup>8,9,10</sup>, Michelle Dyke<sup>1</sup>, Paula Herman<sup>1</sup>, Steve Lee<sup>1</sup>, William K. Oh<sup>1</sup>, Philip W. Kantoff<sup>1</sup>, Muneesh Tewari<sup>3,11</sup>, David G. McLeod<sup>4</sup>, Shiv Srivastava<sup>4</sup>, and Matthew L. Freedman<sup>1</sup>

<sup>1</sup> Department of Medical Oncology, Dana-Farber Cancer Institute, and Harvard Medical School, Boston, Massachusetts <sup>2</sup> Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts <sup>3</sup> Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington <sup>4</sup> Center for Prostate Disease Research, Department of Surgery, Uniformed Services University of the Health Sciences, Rockville, Maryland <sup>5</sup> Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts <sup>6</sup> Departments of Pathology and Laboratory Medicine, Clinical Neurosciences, and Oncology, University of Calgary, Calgary, Alberta, Canada <sup>7</sup> Broad Institute of Harvard and MIT, Cambridge, Massachusetts <sup>8</sup> Department of Urology, University of Washington, Seattle, Washington <sup>9</sup> Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington <sup>10</sup> Department of Veterans Affairs Puget Sound Health Care System, Seattle, Washington <sup>11</sup> Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, Washington

### Abstract

Polymorphisms at 8q24 are robustly associated with prostate cancer risk. The risk variants are located in non-protein coding regions and their mechanism has not been fully elucidated. To further dissect the function of this locus, we tested two hypotheses: i) unannotated microRNAs are transcribed in the region, and that ii) this region is a cis-acting enhancer. Using next generation sequencing, 8q24 risk regions were interrogated for known and novel microRNAs (miRNAs) in histologically normal radical prostatectomy (RP) tissue. We also evaluated the association between the risk variants and transcript levels of multiple genes, focusing on the proto-oncogene, *MYC*. RNA expression was measured in histologically normal and tumor tissue from 280 RP specimens (from 234 European American and 46 African American patients), and paired germline DNA from each individual was genotyped for six 8q24 risk SNPs. No evidence was found for significant miRNA transcription within 8q24 prostate cancer risk loci. Likewise, no convincing association between distal RNA expression and risk allele status was detected in either histologically normal or tumor tissue. To our knowledge, this is one of the first and largest studies to directly assess miRNA in this region and to systematically measure *MYC* expression levels in prostate tissue in relation to inherited risk variants. These data will help to direct the future study of this risk locus.

## INTRODUCTION

Until recently, the genetic etiology of prostate cancer remained largely unknown. Difficulty in validating loci discovered by linkage mapping led to the hypothesis that the genetic basis of prostate cancer was largely due to the actions of many loci, each of modest effect. The effort to identify and to catalogue common human genetic variation resulted in dense genetic maps and facilitated the mapping of lower penetrant risk loci through genome wide association studies (1). In 2006, a prostate cancer locus on chromosome 8q24 was among the first genetic risk loci to be reproducibly validated. Initially described by two groups using different methodologies, the region confers an elevated risk across multiple ethnic groups (2,3), and multiple studies involving thousands of cases and controls have confirmed the finding (4–7). Since the original discovery in prostate cancer, risk for other cancers, such as those of the colon, breast and bladder, have also been mapped to 8q24 (8–11).

Fine mapping of the locus has revealed a set of six single nucleotide polymorphisms (SNPs) and a microsatellite variant that independently contribute to prostate cancer risk (12). These alleles cluster in 3 regions (defined by linkage disequilibrium) spanning a distance of ~490 kilobases (kb) (Supplemental figure 1). Intriguingly, the risk variants are located in a non-protein coding region of the genome and the mechanism by which they contribute to disease is unknown. Possible mechanisms of action include the presence of unannotated transcripts whose expression is influenced by inherited variation, and/or regulation of genes beyond the risk regions.

The annotated protein coding region closest to any risk allele is the well-known oncogene, *MYC*, >250 kb from the nearest prostate cancer risk SNP. Attempts have been made to detect an association between *MYC* expression and genotype at cancer risk SNPs. To date, no study has definitively found an association, but each has been limited by sample size or tissue type. For example, some have measured *MYC* RNA expression in lymphoblastoid cell lines rather than the tissue from which cancer formed. Expression of many genes differs markedly across tissues (13). Therefore, when evaluating the influence of a risk SNP for a particular disease such as prostate cancer, it is important to study the gene in a tissue-specific context.

In the present study, two hypotheses concerning the mechanism of inherited prostate cancer risk are evaluated: first, that 8q24 prostate cancer risk regions harbor previously unannotated miRNA species and second, that elements in the risk regions participate in the regulation of distal genes, affecting RNA expression. We searched for miRNA transcripts within risk loci in prostate tissue using a next generation sequencing approach. Association between 8q24 risk allele status and *MYC* mRNA expression levels was then measured in 401 normal and tumor prostate tissue samples derived from 280 European and African American prostate cancer patients. A subset of 176 samples had histologically normal and tumor isolated via laser capture microdissection for RNA extraction. Lastly, in a subset of 158 samples, associations between risk allele carrier status and mRNA expression of six other annotated genes in the 8q24 region were evaluated.

## MATERIALS and METHODS

### Sequencing for detection of microRNA species

Histologically normal segments of prostate tissue were obtained from individuals undergoing retropubic RP for prostate cancer. Specimens were collected under informed consent with institutional review board approved protocols. Approximately 100 mg of fresh frozen, histologically normal prostate tissue from each of two individuals was homogenized in Ambion mirVana Lysis/Binding buffer using a Qiagen Tissuelyzer. Total RNA was extracted with the Ambion miRVana miRNA isolation kit, using the manufacturers recommended protocol and

RNA samples were pooled. Three small RNA cDNA libraries were generated using 50 µg, 10 µg, and 5 µg of pooled RNA, respectively, using a protocol described previously (14) with modified linker and primer sequences to adapt for sequencing on an Illumina Genome Analyzer. Three prime cloning linker 1(5' rAppCTGTAGGCACCATCAAT/3ddC/3') was purchased from IDT and the five prime Illumina linker was synthesized by IDT as a custom oligo. The fully ligated libraries were reverse transcribed and PCR amplified with primer sequence necessary for sequencing on the Illumina Genome Analyzer. In addition, 40 non-human, synthetic miRNA oligos were added to each RNA sample before library preparation for use as normalization controls, although analysis of these is not relevant to the present study. Sequencing was carried out using the Illumina Genome Analyzer and raw image data was processed using Illumina primary analysis software for image analysis and base calling.

A total of 17,552,100 reads were obtained from the three libraries. The linker sequence CTGTAGGCACCATCAATC was trimmed from the 3' end, following which the reads were collapsed to generate nonredundant sequences. These were then mapped to the reference human genome sequence (NCBI build 36.1). As 3' end non-templated addition of nucleotides is common in miRNA deep sequencing datasets, we trimmed the final three nucleotides from reads not mapping to the human genome and then re-mapped them to the human genome sequence; no additional reads mapping to the 8q24 region were identified by this operation. Unique sequences mapping to the segment chr8:128,100,000–128,700,000 were selected for examination in this study. As described in Results, all sequences in this segment were present at 3 reads or fewer in the dataset, with the exception of one sequence present at 5 reads.

### Gene expression analysis in three RP patient populations

Fresh frozen RP specimens from 108 subjects at the Dana-Farber Cancer Institute (DFCI) and Brigham and Women's Hospital (Boston, MA) were reviewed by a pathologist (J.C.) to isolate areas prostatic adenocarcinoma and benign tissue. Areas of tumor were selected where >60% of cells consisted of tumor cells. Areas of benign tissue were selected where >50% of cells consisted of non-neoplastic epithelium and were at least 5mm away from any area of tumor focus. Two mm punch biopsy cores of frozen tissue were processed RNA extraction using a modified Qiagen Allprep DNA/RNA protocol. RNA from tumor and from normal tissue was isolated in 88 RP samples (62 European American and 26 African American) via laser capture microdissection (LCM) at the Center for Prostate Disease Research (CPDR) (Rockville, MD), a collection of databases derived from nine military hospitals (15).. Procedures for tissue processing, LCM, and RNA preparation were performed as described previously (16–18). RNA was isolated from 67 RP tumor specimens from subjects in the Physicians' Health Study (PHS),, initiated in 1982 and comprising 22,071 U.S. male physicians, aged 40–84 years (19). RNA was reverse transcribed and each sample was subjected to one of three methods of gene expression analysis-competitive RT-PCR, quantitative real time-PCR and the cDNA-mediated Annealing, Selection, Extension and Ligation (DASL) expression assay.

In competitive RT-PCR, used for DFCI samples, 7 transcripts of interest were chosen based on proximity to risk polymorphisms and previously reported oncogenic activity. Seven normalization genes were chosen based on known expression in prostate tissue. All 14 assays and their competitive oligos were plexed into a single reaction mix, using the Sequenom iPLEX mass spectrometry platform. Sequences are available upon request. Reactions were performed in quadruplicate using 8 serial dilutions of competitor, ranging from  $10^{-18}$  M to  $10^{-12}$  M. Thus, a total of 32 reactions were performed for each individual cDNA species. Resulting spectra were analyzed and the EC50-the point at which cDNA and competitor concentrations are equal-was calculated using QGE Analyzer software (Sequenom). A gene expression normalization factor was calculated for each sample using the geNorm algorithm (20).

CPDR samples were analyzed via TaqMan-based qRT-PCR on ABI 7700 (Applied Biosystems), as previously described (16,18). *MYC* gene expression in each sample was normalized to GAPDH expression levels and results were plotted as average CT (threshold cycle) values of duplicate samples.

Analyses across DFCI and CPDR cohorts were performed separately by ancestry. To compare samples based on total number of risk alleles, each cohort was split into 5 risk groups. The European Americans were grouped as 0–1, 2, 3, 4, and 5–6 risk allele carriers. The African Americans were grouped as carrying 2–6, 7–8, and 9–10 risk alleles. Expression levels were also conditioned on number of risk SNPs carried within a given risk region in a separate analysis. Transcript levels for each message were regressed on risk allele carrier status. The analysis used Kruskal-Wallis tests for a global test of difference among compared samples; p-values were adjusted for multiple comparisons using permutation testing. To compare samples based on genotype status at an individual polymorphism, each individual was classified as carrying 0, 1, or 2 risk alleles (reflecting homozygote wild type, heterozygote, and homozygote risk genotypes). For SNPs at low frequency in the population, only two groups were feasible—0 vs. 1 or 2 alleles.

PHS samples were analyzed via the DASL expression assay (Illumina, Inc.), as described previously (19). Molecular data generated using the DASL approach were compared with prostate cancer expression array data generated using frozen tissue samples on conventional microarray platforms to verify results (19). Samples were compared based on number of risk alleles at an individual polymorphism (0, 1, or 2 risk alleles or 0 vs. 1–2 for low frequency risk alleles). Anova analysis (F-test) was used to detect associations between *MYC* expression and risk allele status.

Genomic DNA was prepared from peripheral blood using QIAamp DNA Blood mini kit (QIAGEN Inc, Valencia, CA). Subjects were genotyped for 6 SNPs identified in a fine mapping analysis of 8q24-associated prostate cancer risk and 8 SNPs in high or complete linkage disequilibrium with the risk SNPs. Genotyping was carried out using Sequenom iPLEX mass spectrometry platform. The error rate on this platform is estimated to be less than 0.03%.

## RESULTS

### Assessment of microRNA transcripts in the 8q24 prostate risk regions

MicroRNAs (miRNAs) are an important layer of regulatory control and therefore we investigated the possibility that the 8q24 locus risk allele may be related to a known or novel miRNA encoded in the region.

To identify potential novel miRNAs expressed in the prostate that may have not yet been discovered, we characterized known and novel miRNAs in histologically normal prostate tissue using a next generation sequencing approach. We pooled RNA from histologically normal prostate tissue samples from two individuals collected at the time of RP. We isolated small RNA (18–24 nt) in triplicate by polyacrylamide gel electrophoresis and generated three cDNA libraries that we subjected to ultra-high-throughput sequencing using an Illumina Genome Analyzer. A total of 17,552,100 reads were generated.

We first analyzed seven computationally defined miRNAs that had been reported previously as 8q24 miRNAs and confirmed by RT-PCR to determine whether they were present in our dataset (21). None of the seven miRNAs perfectly matched any of our reads. Given that it is difficult to computationally identify the 5' and 3' ends of miRNAs, we trimmed the sequence of these seven miRNAs by 1, 2 and then 3 nucleotides from the 5' and 3' ends and searched

for a match. Even by that analysis, none of the computationally identified miRNAs were found in our expression data.

We proceeded to search for potential novel miRNAs in the 8q24 region that might be relevant to the risk allele. We processed our sequence reads by trimming linker sequences, combining redundant reads and mapping to the reference human genome sequence 8q24 region at chr8:128,100,000–128,700,000 (NCBI build 36.1) (Supplemental figure 1). Removing sequences that mapped to 10 or more loci (i.e., those that correspond to repetitive sequence and are difficult to ascribe to a particular locus), gave us 459 unique sequences with 487 reads mapping to 1,283 loci in the specific region of interest. All the matching sequences were present at extremely low abundance. For example, 443 of the 471 were singleton reads, 24 had two reads, three had three reads, and one had five reads. Minimal criteria for annotation of novel miRNAs generally require multiple reads and origin from a hairpin precursor. We examined each of the non-singleton sequences for origin from a putative hairpin precursor by folding each read with 100 nt of up- and down-stream genomic sequence at a given locus (as described previously in Bar *et al*) (14). None of the non-singleton reads met criteria for origin from a hairpin precursor.

Taken together, our analyses did not find evidence for a significant miRNA transcript in the 8q24 risk region studied.

### Genotype and *MYC* expression in three prostate cancer cohorts

We sought to identify a gene or set of genes whose expression is influenced by risk variants. We focused on the proto-oncogene *MYC*, testing the hypothesis that *MYC* transcript abundance is associated with risk allele status. A total of 280 individuals were evaluated from three independent prostate cancer populations (Table 1).

Given the range of plausible biologic models and the genetic complexity of the 8q24 locus, *MYC* expression was analyzed across a range of scenarios (Table 1). First, histologically normal prostate tissue and prostate tumor tissue were each examined since risk alleles may exert their effect more profoundly in a particular tissue type. To minimize confounding from tissue cell admixture, a subset of RP samples whose normal and tumor epithelial cells were isolated by laser capture microdissection were analyzed separately (N=88). Further, European American and African American prostate tissues were distinguished from one another, since risk allele frequencies differ significantly across the two populations and one of the risk alleles is expressed only in men of African ancestry.

To evaluate risk allele status, each subject was genotyped for six prostate cancer risk SNPs at chromosome 8q24 (Table 1). The SNPs chosen for analysis independently influence risk and reside across three distinct linkage disequilibrium regions (Supplemental figure 1). Given 6 risk SNPs, an individual can carry a maximum of 12 risk alleles (2 alleles per locus  $\times$  6 loci). Since the Broad1193405 risk allele is present only in individuals of African ancestry, European Americans carry a maximum of 10 risk alleles. Across all populations genotyped, European American subjects (N=234) carried a total of 0 to 6 alleles, and the total number of risk alleles carried by African American subjects (N=46) ranged from 2 to 10.

Three models for the relationship between inherited variation and gene expression were tested since the mechanism by which 8q24 variants confer risk is unknown. One hypothesis is that the risk alleles act collectively to influence *MYC* expression. In this model, European American and African American subjects were categorized into subgroups based on the number (out of a possible total of 12) of risk alleles carried (see Methods). No significant association exists between *MYC* expression and total risk allele status in normal prostate tissue or tumor tissue ( $p \geq 0.05$ ). Another hypothesis is that each of the three 8q24 linkage disequilibrium risk regions



(as described by Haiman et al. (12)) behaves as a unit that influences a target gene. In this model, risk alleles within a given risk region are summed and analyzed for association with *MYC* expression. Under this model, no statistically significant associations ( $p \geq 0.05$ ) were found between steady state *MYC* mRNA expression and genotype within any risk region. This included normal and tumor tissue in both ancestral populations

A third model is that each risk SNP acts independently. Based on the SNP allele frequencies, subjects were categorized as carrying 0, 1, or 2 risk alleles, or as carriers or non-carriers of the risk allele. (Table 1 and Figure 1) Only one assay demonstrated a nominally statistically significant association between risk status at an individual SNP and *MYC* expression: rs13254738 in laser capture microdissected European American tumor tissue ( $p=0.02$ ). The statistical significance was driven by increased expression among the seven individuals homozygous for the rs13254738 risk allele. No difference in expression was identified between homozygotes for the wild type allele and heterozygotes. The association was not observed in two other European American tumor tissue sample sets or in normal prostate tissue from the same population.

### Association between risk alleles and six candidate 8q24 transcripts

To detect associations between risk allele status and expression at other candidate genes, six transcripts of interest in the 8q24 region were selected for analysis. (Supplemental figure 1) Annotated genes located within 1 Mb for the nearest risk region, such as *FAM84B* and *PVT1*, were included. *PVT1* appears to encompass a wide area of transcription and exists in multiple splice forms. Three annotated transcripts in the region were selected to survey *PVT1* expression-TMEM75, M34330 and BC033263. Also, the genes *MTSSI* and *KIAA0196* were analyzed. Though greater than 1 Mb from any single risk SNP, these genes do fall within the originally described 3.8 Mb admixture peak identifying 8q24 as a risk locus (3) and were added to the analysis since they have been implicated in prostate carcinogenesis (22, 23).

Associations between risk allele status and transcript abundance at these six additional 8q24 transcripts (Supplemental figure 1) were assessed in 105 histologically normal European American and 20 histologically normal African American RP tissues. The same analysis was performed in paired tumor tissue specimens in a 33-person subset of the European American samples. As with *MYC*, associations were sought between total risk alleles, the number of risk alleles within a risk region and the number of risk alleles at an individual SNP. Among the European Americans no significant association was seen between expression of any of the six transcripts and risk allele status in normal and tumor tissue. Among African Americans, a statistically significant association ( $p=.0068$ ) was detected with expression levels of *PVT1* transcript BC033263 and risk allele status across region 2 (Supplemental figure 1). In this instance, expression increased when comparing those with two risk alleles in this region to those with four. However, expression decreased when comparing carriers of four risk alleles to carriers of five. Thus, no consistent trend in gene expression correlated with risk allele status.

## DISCUSSION

Inherited variants at chromosome 8q24 are associated with prostate cancer risk, a finding validated in multiple cohorts and several ethnic groups (2–7). This discovery, made possible by the sequencing of the human genome, the International HapMap Project, and technological advances, provides a unique opportunity to gain insight into the pathogenesis of prostate cancer. However, the observation that many risk variants reside in non-protein coding regions presents a formidable challenge to identifying the mechanism by which these variants cause disease. Since there are no known protein coding genes at the risk loci, we sought to identify whether one or more miRNAs are transcribed in the region using a high throughput sequencing

approach. No robust evidence of miRNA activity was found. This observation further supports the notion that the risk variants may be regulatory elements.

Previous studies have utilized gene expression as an intermediate trait to understand the mechanism through which risk alleles are acting (24,25). Several factors point to *MYC* as a prime candidate for being the transcript under regulation by risk variants. It is the closest annotated gene to the risk loci and is a well-established oncogene. *MYC* is instrumental in several critical cell processes, such as embryogenesis and regulation of cell growth, cell differentiation and apoptosis. Coding mutations do not appear to be a common mechanism for *MYC*-associated carcinogenesis (26). Overexpression of the gene, on the other hand, has been shown to lead to prostate cancer formation and progression (27–29). In addition, overexpression of *MYC* protein in prostate cancer relative to matched normal tissue has been shown (30). Finally, recent work demonstrates that risk loci may act as enhancer elements and these elements come into contact with *MYC*<sup>12</sup>. Influence on *MYC* expression therefore is a likely mechanism by which the 8q24 risk polymorphisms exert their effects.

Yet, no study has comprehensively quantified *MYC* expression levels in prostate tissue of men across all of the known risk alleles. A previous study genotyped 32 individuals for the SNP rs1447295 and reported a statistically significant difference in histologically normal prostate *MYC* expression between individuals homozygous for the wild type allele and those heterozygous for the risk SNP. However, the analysis was based only on 6 heterozygotes (31). Upon discovery of an 8q24 risk variant for colon cancer (rs6983267, in prostate cancer risk region 3), cytoplasmic and nuclear immunohistochemical staining of *MYC* in 86 colon cancer samples were analyzed based on risk allele status (8). No significant differences were identified. In a recent study identifying 8q24 variants conferring risk for bladder cancer-located only 30 kb upstream of *MYC*- *MYC* expression was measured as a function of risk allele status; RNA expression levels in blood and adipose tissue were examined, and no significant association with risk allele status was found (11).

The data in the present analysis strongly substantiate that steady state levels of *MYC* in normal and tumor prostate tissues are not associated with risk allele status. One nominally statistically significant association was observed for *MYC* expression in a microdissected tumor tissue sample set. The association was detected between *MYC* expression in European American patients carrying the risk allele at SNP rs13254738. This finding, however, must be carefully interpreted. The signal seen in this subgroup is primarily driven by an increase in *MYC* expression among homozygotes for the risk allele, however only seven individuals were homozygous for this variant. Expression levels of heterozygotes for the risk SNP in this subgroup actually decreased slightly relative to those homozygous for the non-risk allele. Moreover, two other tumor datasets presented here did not demonstrate this association, nor did laser capture microdissected African American tumor samples. If the rs13254738 finding is a true positive, it suggests that the other risk polymorphisms are associated with risk via mechanisms other than through *MYC*. While this is a possibility, and one that was modeled in our analysis, it seems an unlikely scenario. Given the number of tests performed, the possibility that this observation is due to chance must be considered.

Our study had 80% power to detect a minimal difference of 2–2.5 fold in mean expression levels, depending on allele frequencies, among European American subjects at an alpha level of 0.05. Among African American subjects, the study had 80% power to detect a minimal difference of 2.4–3.3 fold in mean expression levels, depending on allele frequencies.

---

<sup>12</sup>M.M. Pomerantz, N. Ahmadiyeh, L. Jia, P. Herman, M.P. Verzi, C.A. Beckwith, J.A. Chan, C.A. Haiman, C. Yan, B.E. Henderson, B. Frenkel, J. Barretina, A. Bass, J. Tabernero, J. Baselga, R. Shivdasani, G.A. Coetzee, M.L. Freedman, unpublished results.

Expression of six other candidate transcripts at 8q24 in a total of 125 RP specimens was also evaluated for association with risk allele status. These included three transcripts within *PVT1*. Recent evidence suggests that *PVT1*, a noncoding RNA complex, plays a significant role in cancer pathogenesis as an activator of *MYC* and/or as an oncogene itself (32,33). No consistently significant trend in gene expression for these transcripts or for genes *FAM84B*, *KIAA0196* and *MTSS1* were found to be associated with risk SNP genotype. Power considerations for this aspect of the study were similar to those for *MYC*.

Despite these findings, the 8q24 locus, and *MYC* in particular, should continue to be the target of further investigation. Risk alleles may influence the *rate* of *MYC* mRNA expression, rather than total abundance, with steady state RNA remaining relatively constant throughout the cell. Risk alleles may increase *MYC* expression in a non-epithelial component, such as stromal cells; previous work has suggested that certain signals can dramatically shift *MYC* expression from one cell type to another despite little change in overall mRNA expression (34). Additionally, steady state levels may remain consistent across genetic risk groups but response to cellular insult or stimulation may differ. Risk alleles also may influence transcription at an earlier stage in development and the effect may no longer be present at a later stage of life when prostate cancer is diagnosed (35). In addition, it is possible that the risk alleles impact the transcription of *MYC* isoforms that remain to be annotated. The ENCODE project has revealed a complex landscape of transcription and that many loci have a surprising number of previously unannotated exons (13). Finally, risk alleles may influence *MYC* expression so subtly that it is very difficult given current technology to detect the causative changes. There may be selective pressure keeping expression of *MYC* at relatively low levels to avoid activation of intrinsic tumor suppression (36).

In summary, one of the first reproducible and robust genetic associations for prostate cancer has been identified at chromosome 8q24 (2,3,8,9,11,37). Data presented here suggest that transcription of miRNA is not present within risk regions. Steady-state expression levels of multiple transcripts at 8q24, including *MYC*, across many different conditions are not associated with the risk alleles. Identification of the gene(s) involved in this process will lend critical insight into the pathways that, when deregulated, result in prostate cancer. In fact, the majority of risk alleles discovered to date by genetic association studies are located in non-protein coding regions. Establishing a framework for understanding the functional consequences of inheriting these alleles will become increasingly important.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Grant support from the NIH (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 SPORE (National Cancer Institute Grant No. 5P50CA90381), the American Society of Clinical Oncology (to MMP), the Prostate Cancer Foundation (to MMP), the Fred Hutchinson Cancer Research Center New Development funds, including support from the Canary Foundation (to MT), Pilot Grant from the Pacific Northwest Prostate Cancer Specialized Program of Research Excellence Grant (P50 CA97186 to MT), and Chromosome Metabolism Training Grant (5 T32 CA09657-16 to SKW).

We thank David Reich for his guidance in designing and analyzing ancestry-informative markers. We thank Oliver Sartor for expert advice in project design.

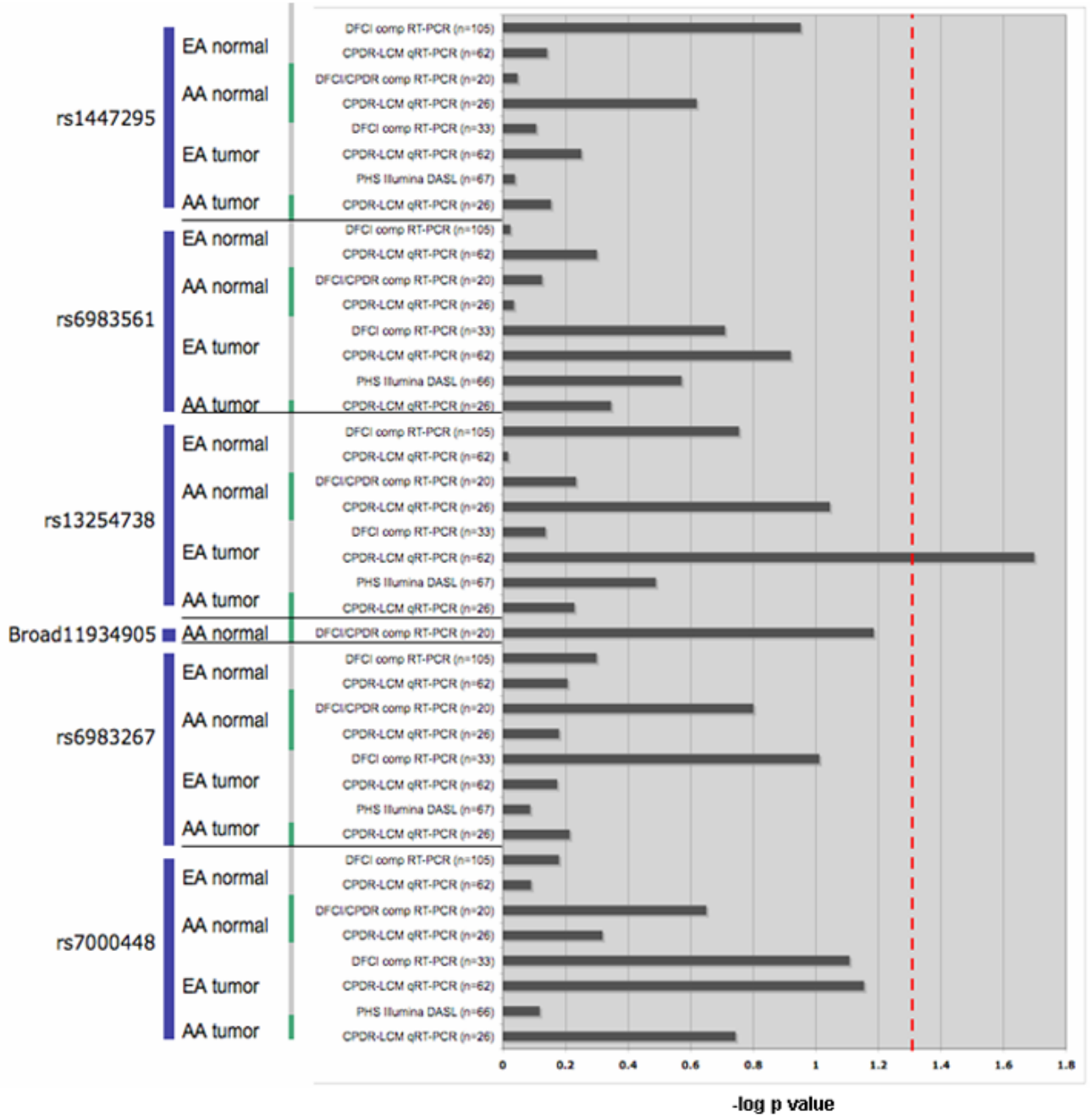
## References

1. A haplotype map of the human genome. *Nature* 2005;437:1299–1320. [PubMed: 16255080]



2. Amundadottir LT, Sulem P, Gudmundsson J, et al. A common variant associated with prostate cancer in European and African populations. *Nat Genet* 2006;38:652–8. [PubMed: 16682969]
3. Freedman ML, Haiman CA, Patterson N, et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci U S A* 2006;103:14068–73. [PubMed: 16945910]
4. Schumacher FR, Feigelson HS, Cox DG, et al. A common 8q24 variant in prostate and breast cancer from a large nested case-control study. *Cancer Res* 2007;67:2951–6. [PubMed: 17409400]
5. Severi G, Hayes VM, Padilla EJ, et al. The common variant rs1447295 on chromosome 8q24 and prostate cancer risk: results from an Australian population-based case-control study. *Cancer Epidemiol Biomarkers Prev* 2007;16:610–2. [PubMed: 17372260]
6. Suuriniemi M, Agalliu I, Schaid DJ, et al. Confirmation of a positive association between prostate cancer risk and a locus at chromosome 8q24. *Cancer Epidemiol Biomarkers Prev* 2007;16:809–14. [PubMed: 17416775]
7. Wang L, McDonnell SK, Slusser JP, et al. *Cancer Res* 2007;67:2944–50. [PubMed: 17409399]
8. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 2007;39:989–94. [PubMed: 17618283]
9. Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet* 2008;40:623–30. [PubMed: 18372905]
10. Haiman CA, Le Marchand L, Yamamoto J, et al. *Nat Genet* 2007;39:954–6. [PubMed: 17618282]
11. Kiemenev LA, Thorlacius S, Sulem P, et al. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat Genet* 2008;40:1307–12. [PubMed: 18794855]
12. Haiman CA, Patterson N, Freedman ML, et al. *Nat Genet* 2007;39:638–44. [PubMed: 17401364]
13. Birney E, Stamatoyannopoulos JA, Dutta A, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007;447:799–816. [PubMed: 17571346]
14. Bar M, Wyman SK, Fritz BR, et al. MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. *Stem Cells* 2008;26:2496–505. [PubMed: 18583537]
15. Amling CL, Riffenburgh RH, Sun L, et al. Pathologic variables and recurrence rates as related to obesity and race in men with prostate cancer undergoing radical prostatectomy. *J Clin Oncol* 2004;22:439–45. [PubMed: 14691120]
16. Petrovics G, Zhang W, Makarem M, et al. Elevated expression of PCGEM1, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients. *Oncogene* 2004;23:605–11. [PubMed: 14724589]
17. Petrovics G, Liu A, Shaheduzzaman S, et al. Frequent overexpression of ETS-related gene-1 (ERG1) in prostate cancer transcriptome. *Oncogene* 2005;24:3847–52. [PubMed: 15750627]
18. Rosner IL, Ravindranath L, Furusato B, et al. Higher tumor to benign ratio of the androgen receptor mRNA expression associates with prostate cancer progression after radical prostatectomy. *Urology* 2007;70:1225–9. [PubMed: 18158066]
19. Setlur SR, Mertz KD, Hoshida Y, et al. Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. *J Natl Cancer Inst* 2008;100:815–25. [PubMed: 18505969]
20. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;3:R34.
21. Huppi K, Volfovsky N, Runfola T, et al. The identification of microRNAs in a genomically unstable region of human chromosome 8q24. *Mol Cancer Res* 2008;6:212–21. [PubMed: 18314482]
22. Hofer MD, Kuefer R, Varambally S, et al. The role of metastasis-associated protein 1 in prostate cancer progression. *Cancer Res* 2004;64:825–9. [PubMed: 14871807]
23. Porkka KP, Tammela TL, Vessella RL, Visakorpi T. RAD21 and KIAA0196 at 8q24 are amplified and overexpressed in prostate cancer. *Genes Chromosomes Cancer* 2004;39:1–10. [PubMed: 14603436]
24. Moffatt MF, Kabesch M, Liang L, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007;448:470–3. [PubMed: 17611496]

25. Emilsson V, Thorleifsson G, Zhang B, et al. Genetics of gene expression and its effect on disease. *Nature* 2008;452:423–8. [PubMed: 18344981]
26. Levens D. How the c-myc promoter works and why it sometimes does not. *J Natl Cancer Inst Monogr* 2008;39:41–3. [PubMed: 18648001]
27. Buttyan R, Sawczuk IS, Benson MC, Siegal JD, Olsson CA. Enhanced expression of the c-myc protooncogene in high-grade human prostate cancers. *Prostate* 1987;11:327–37. [PubMed: 2446300]
28. Gil J, Kerai P, Leonart M, Bernard D, et al. Immortalization of primary human prostate epithelial cells by c-Myc. *Cancer Res* 2005;65:2179–85. [PubMed: 15781629]
29. Williams K, Fernandez S, Stien X, et al. Unopposed c-MYC expression in benign prostatic epithelium causes a cancer phenotype. *Prostate* 2005;63:369–84. [PubMed: 15937962]
30. Gurel B, Iwata T, Koh CM, et al. Nuclear MYC protein overexpression is an early alteration in human prostate carcinogenesis. *Mod Pathol* 2008;21:1156–67. [PubMed: 18567993]
31. Sole X, Hernandez P, de Heredia ML, et al. Genetic and genomic analysis modeling of germline c-MYC overexpression and cancer susceptibility. *BMC Genomics* 2008;9:12. [PubMed: 18190704]
32. Shtivelman E, Bishop JM. The PVT gene frequently amplifies with MYC in tumor cells. *Mol Cell Biol* 1989;9:1148–54. [PubMed: 2725491]
33. Guan Y, Kuo WL, Stilwell JL, et al. Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. *Clin Cancer Res* 2007;13:5745–55. [PubMed: 17908964]
34. Weber A, Liu J, Collins I, Levens D. TFIIH operates through an expanded proximal promoter to fine-tune c-myc expression. *Mol Cell Biol* 2005;25:147–61. [PubMed: 15601838]
35. Beer S, Zetterberg A, Ihrie RA, et al. Developmental context determines latency of MYC-induced tumorigenesis. *PLoS Biol* 2004;2:e332. [PubMed: 15455033]
36. Murphy DJ, Junttila MR, Pouyet L, et al. Distinct thresholds govern Myc's biological output in vivo. *Cancer Cell* 2008;14:447–57. [PubMed: 19061836]
37. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;447:1087–93. [PubMed: 17529967]



**Figure 1. Statistical significance of association between MYC expression and 8q24 risk allele genotype**

Each bar represents the  $-\log$  of p-value for association between MYC expression and genotype at each individual risk polymorphism (far left). Each population is subdivided by ancestry (European American or African American), RP tissue type (normal or tumor tissue), prostate cancer cohort (DFCI, CPDR or PHS), and quantitative gene expression platform. Red dashed line represents threshold for statistical significance ( $p=0.05$ ). EA, European American; AA, African American; comp RT-PCR, competitive RT-PCR; LCM, laser capture microdissection.

Table 1

## Summary of study populations and genotyping results

Study subjects were first categorized by ancestry and normal prostate versus prostate tumor tissue in RP specimens. Subjects were further characterized by institution, method of tissue isolation (laser capture microdissection versus macrodissection), and method of quantifying gene expression. Genotyping results for six risk SNPs are presented in the columns to the right. The Broad11934905 SNP was not tested in all subjects. For each SNP, listed are the number of men carrying no risk alleles, the number carrying one copy, and the number carrying two copies of the risk allele.

Subjects		Risk SNPs: no. homozygote no risk (%) / no. heterozygote (%) / no. homozygote risk (%)													
Population	Source	Method	No.	rs1447295	rs13254738	rs6983561	Broad11934905	rs6983267	rs7000448						
EA normal tissue	DFCI	comp PCR	105	1 (1)	15 (14)	11 (10)	1 (1)	105 (100)	0 (0)	0 (0)	28 (27)	41 (39)	36 (34)	45 (43)	13 (12)
	CPDR LCM	qRT-PCR	62	0 (0)	7 (11)	8 (13)	0 (0)	- (-)	- (-)	16 (26)	27 (44)	19 (31)	20 (32)	34 (55)	8 (13)
AA normal tissue	DFCI/CPDR	comp PCR	20	1 (5)	10 (50)	8 (40)	8 (40)	7 (35)	13 (65)	0 (0)	0 (0)	5 (25)	15 (75)	2 (10)	10 (50)
	CPDR LCM	qRT-PCR	26	3 (12)	12 (46)	11 (42)	5 (19)	- (-)	- (-)	0 (0)	0 (0)	6 (23)	19 (73)	5 (19)	9 (35)
EA tumor tissue	DFCI	comp PCR	33	0 (0)	5 (15)	5 (15)	0 (0)	33 (100)	0 (0)	0 (0)	9 (27)	13 (39)	11 (33)	14 (42)	4 (12)
	CPDR LCM	qRT-PCR	62	0 (0)	7 (11)	8 (13)	0 (0)	- (-)	- (-)	16 (26)	27 (44)	19 (31)	20 (32)	34 (55)	8 (13)
	PHS	DASL	67	0 (0)	9 (13)	9 (13)	0 (0)	- (-)	- (-)	18 (27)	37 (55)	12 (18)	26 (39)	32 (48)	8 (12)
AA tumor tissue	CPDR LCM	qRT-PCR	26	3 (12)	12 (46)	11 (42)	5 (19)	- (-)	- (-)	0 (0)	6 (23)	19 (73)	5 (19)	9 (35)	12 (46)

Abbreviations: EA, European Americans; AA, African Americans; DFCI, Dana-Farber Cancer Institute; CPDR, Center for Prostate Disease Research; PHS, Physicians' Health Study; LCM, laser capture microdissection; comp PCR, competitive PCR; qRT-PCR, quantitative real time PCR; DASL, cDNA-mediated Annealing, Selection, Extension and Ligation; HMZ, homozygote; HET heterozygote.