Association of germline genetic variants with *TMPRSS2-ERG* fusion status in prostate cancer

Indu Kohaar[1,2], Qiyuan Li[3], Yongmei Chen[1,2], Lakshmi Ravindranath[1,2], Denise Young[1,2], Amina Ali[1,2], Isabell A. Sesterhenn[4], Inger L. Rosner[1], Jennifer Cullen[1,2], Shiv Srivastava[1], Matthew Freedman[3] and Gyorgy Petrovics[1,2]

[1]Center for Prostate Disease Research, Department of Surgery, Uniformed Services University of the Health Sciences and the Walter Reed National Military Medical Center, Bethesda, Maryland, USA

[2]Henry Jackson Foundation for the Advancement of Military Medicine (HJF), Bethesda, Maryland, USA

[3]Department of Medicine, Harvard Medical School and Dana-Farber Cancer Institute, Massachusetts General Hospital, Boston, Massachusetts, USA

[4]Joint Pathology Center, Silver Spring, Maryland, USA

Correspondence to: *Indu Kohaar*, **email**: *ikohaar@cpdr.org*
          Gyorgy Petrovics, **email**: *gpetrovics@cpdr.org*

ABSTRACT

Introduction: Oncogenic activation of ERG resulting from *TMPRSS2-ERG* gene fusion is a key molecular genetic alteration in prostate cancer (CaP). The frequency of ERG fusion is variable by race; however, there are limited data available on germline polymorphisms associating with ERG fusion status. The goal of this study is to identify the inherited risk variants associating with ERG status of CaP.

Materials and Methods: SNP genotyping was performed on the Illumina platform using Infinium Oncoarray SNP chip on blood derived genomic DNA samples from 400 patients treated by radical prostatectomy at a single military institution. ERG status was determined in whole mounted prostate specimens by immuno-histochemistry (IHC) for ERG protein expression. Data analysis approaches included association analyses based on EMMAX and imputation by IMPUTE2. Imputed SNPs were validated by ddPCR.

Results: SNP genotyping analysis using imputation identified rs34349373 (p 4.68×10^{-8}) and rs2055272 (p 5.62×10^{-8}) in *TBC1D22B* to be significantly associated with ERG fusion status in index tumor and non-index tumor foci. Imputed SNP rs2055272 was further experimentally validated by ddPCR with 98.04% (100/102) concordance. Initial discovery analysis based on SNPs on Oncoarray SNP chip, showed significant (p 10^{-5}) association for SNPs (rs6698333, rs1889877, rs3798999, rs10215144, rs3818136, rs9380660 and rs1792695) with ERG fusion status. The study also replicated two previously known ERG fusion associated SNPs (rs11704416 in chromosome 22; rs16901979 in chromosome 8).

Conclusions: This study identified SNPs associated with ERG status of CaP.

Impact: The findings may contribute towards defining the underlying genetics of ERG positive and ERG negative CaP patients.

INTRODUCTION

Prostate cancer (CaP) is a major cause of morbidity and mortality worldwide [1]. It is the second most frequent cancer and the fifth leading cause of cancer death in men [1–3]. In the United States, CaP is the most prevalent non-

skin male cancer and ranks second in cancer-related deaths [3]. Oncogenic activation of ERG resulting from prevalent gene fusions (predominantly as *TMPRSS2-ERG*) is a key driver event in CaP pathogenesis [4–6]. Multiple studies have reported a significantly lower frequency of ERG positivity in CaP tumors of African American (AA; 23%)

compared to Caucasian American (CA; 49%) patients [4, 7–9]. A significant correlation between ERG negativity in CaP tumors with development of distant metastasis, as well as biochemical recurrence (BCR) was noted in CA men by comprehensive analysis on 930 whole mount prostate specimens in AA and CA men [7]. Additionally, AA patients with high Gleason grade tumors [8–10] exhibited primarily ERG negative index tumor type [8, 10]. However, there are also controversial reports on the association of ERG fusion status in CaP with disease aggressiveness [(BCR, pathological Gleason score and Grade Group (GG), prostate specific death) [11–16]. These studies differ by study design, as well as patient clinicopathological features and treatment. Key biological differences are evident across ERG fusion status (positive versus negative), including distinct methylation patterns, with hypermethylation more pronounced in ERG positive versus negative CaP tumors [17, 18]. In addition, the processes of tumor evolution are also different between fusion positive and negative tumors, with *TMPRSS2-ERG* tumors characterized by chromoplexy, while chromothripsis is more common in *TMPRSS2-ERG* negative tumors [19, 20]. Considering the multiclonal and heterogenous nature of CaP, it is important to examine all tumor foci for ERG fusion status as any of these may lead to aggressive CaP [9, 21]. Overall, these findings suggest that tumor etiology is variable, depending on fusion status in CaP.

Based on ERG fusion positive and fusion negative distinctness of CaP, we hypothesized that there may also be differences at the underlying germline level between these two molecular subtypes. CaP is one of the most heritable solid tumors with up to 15% of cases linked to family history [22, 23]. Additionally, inherited germline risk variants have been implicated in different stages of CaP management including screening, staging and treatment [24–26]. Genome wide association studies (GWAS) have identified about 167 common, low penetrance CaP susceptibility variants [27–42]. However, vast majority of GWAS have been performed in populations of European ancestry, only a - few studies are published in men of African-American origin [43–46]. This may have important implications for disease risk prediction across global populations [47], as implied by differences in CaP associated SNPs between AA and CA patients. A whole genome admixture mapping study in AA CaP has identified the 8q24 risk locus to be significantly associated with prostate cancer [48]. We also showed that the Broad11934905 SNP, which segregates with African ancestry, is associated with an increase in non-organ-confined CaP at time of surgery [49].

Thus, it is hypothesized that ERG gene fusion status of AA and CA patients reflects underlying biological and/or genetic differences of CaP development. Since *TMPRSS2-ERG* fusion is considered to be an early event in CaP [50], it is anticipated that SNPs associated with CaP risk may influence ERG fusion status. Therefore, the goal of the present study was to identify germline SNPs associated with ERG status of CaP.

RESULTS

The frequency of SNPs on oncoarray in 321 CaP patients was compared between fusion positive and fusion negative CaP subtypes to agnostically examine the association of the inherited variants with *TMPRSS2-ERG* status of CaP, A description of the patients in the study cohort across ERG + vs. ERG - groups is provided in Table 1. Most men had pathological Gleason Grade Group 1–3 tumors and stage pT2. The frequency of ERG positive index tumors was 37.5% (108/288), while the frequency of positive ERG staining in any tumor focus was 54.3% (158/291). Schematic representation of the study workflow is depicted in Figure 1.

Association between the SNPs and ERG status

For genetic association studies in admixture population, it is important to consider that ancestry differences among the sampled individuals can be a confounder. Failure to appropriately account for population structure due to ancestry admixture can lead to both spurious association (increased type-I error rates—false positive) as well as reduced power (inflated type-II error rates—false negative) (Supplementary Figure 1). To correct for the confounder a variance component approach called Efficient Mixed-Model Association eXpedited (EMMAX) was used (Supplementary Figure 1). This approach is based on pair wise relatedness between individuals, using high-density markers to model the phenotype distribution. EMMAX implements linear mixed model approach for association testing, accounting for global population substructures with an empirical covariance matrix.

In the EMMAX corrected dataset rs6698333, an intron variant of Kruppel-like factor 17 (*KLF17*) and two SNPs (rs1889877, rs3798999) in the intron of adhesion G protein-coupled receptor B3 (*ADGRB3*) were significantly ($< 10^{-5}$) associated with ERG fusion status of the index tumor (Figure 2; Table 2). The Krüppel-like factor (KLF) family is highly conserved zinc finger transcription factors that regulate cell proliferation, differentiation, apoptosis, and migration. Reduced KLF17 in human cancer affects TGF-β and p53 pathways. *ADGRB3* is a p53-target gene that encodes a brain-specific angiogenesis inhibitor, and is a member of the secretin receptor family.

Four SNPs (rs10215144, rs3818136, rs9380660 and rs1792695) were significantly ($< 10^{-5}$) associated with ERG positive phenotype under any tumor focus positive for the fusion (Figure 2, Table 2). rs3818136 is a synonymous variant and rs9380660 is downstream variant of *TBC1D22B*, a GTPase activating protein for Rab family. Rab GTPase proteins are aberrantly expressed in various tumors and are found to be involved in cancer progression. rs10215144 is an intron variant in *AGBL3*, an ATP/GTP binding protein-like 3 and rs1792695 is an intron variant in ncRNA LOC100505474.

Table 1: Clinico-pathological characteristics of patients in ERG+ and ERG- prostate cancer

Variable	ERG-	ERG+	*P* value
N	182	120	
Age at diagnosis (year)			
Mean (SD)	57.9 (8.4)	56.1 (8.8)	0.0785
PSA at diagnosis (ng/mL)			
Median (range)	5.0 (0.6–129.1)	5.3 (0.5–22.4)	0.6992
FU (year)			
Median (range)	7.5 (0.6–17.3)	7.9 (1.2–17.7)	0.5047
Pathological T stage			
pT2	136 (74.7)	93 (77.5)	
pT3–4	46 (25.3)	27 (22.5)	0.5815
GG			
GG1-3	104 (59.4)	82 (68.9)	
GG4-5	71 (40.6)	37 (31.1)	0.098
Missing			
Surgical margin			
Negative	139 (80.8)	83 (70.9)	
Positive	33 (19.2)	34 (29.1)	0.0509

Abbreviations: N: total numbers; SD: standard deviation; FU: follow-up; pT: pathological T stage; GG: Gleason group. Data represented here is based on index tumor.
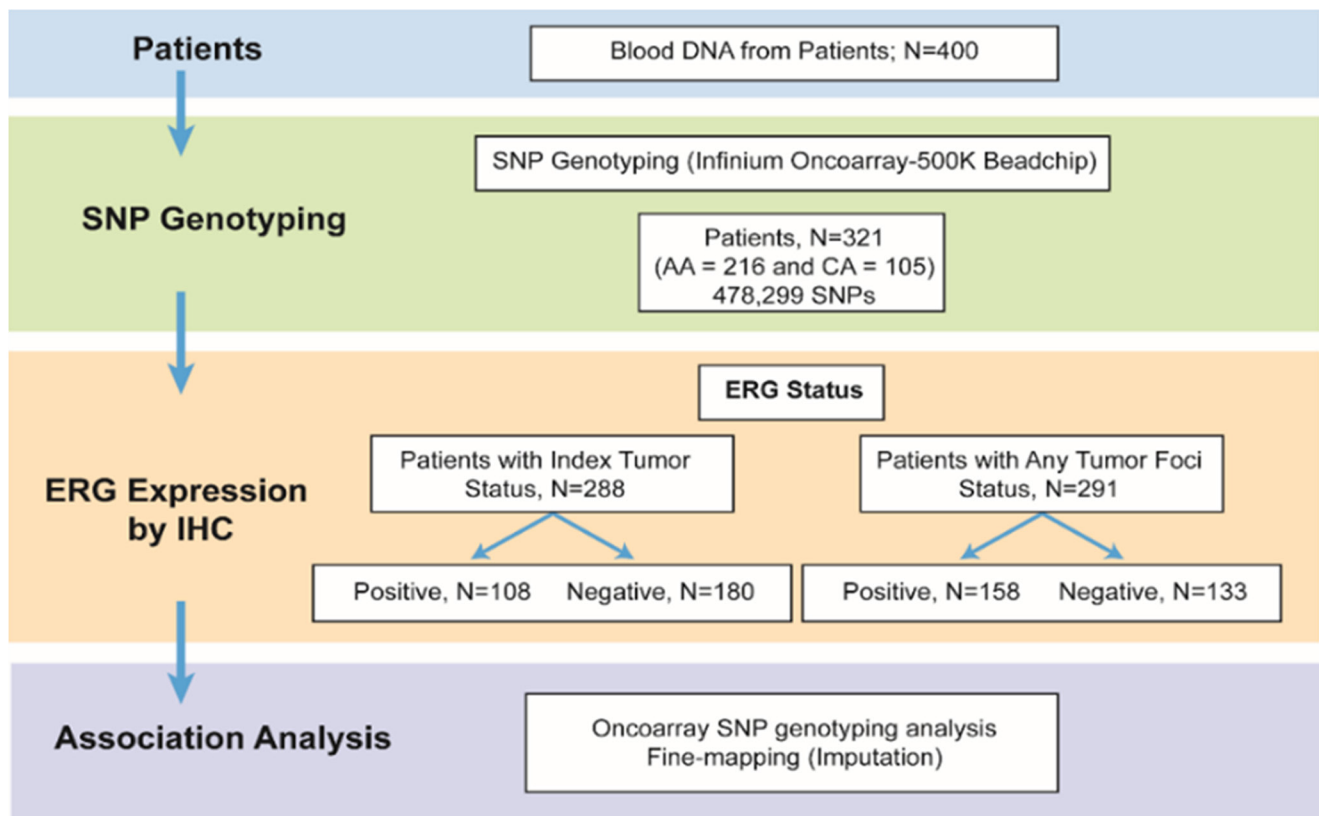

Figure 1: Schematic representation of the study.

We also performed the association analysis for the six known GWAS risk SNPs at 8q24, 5p15, 8p21, 17q24, 19q13 and 22q13 that were reported to be associated with ERG fusion status [51–53]. The present study validates rs11704416 in chromsome 22 and rs16901979 in chromosome 8q24 to be significantly associated with ERG fusion status either by index tumor (rs11704416; $p = 0.0043$; rs16901979; $p = 0.012$) or by any tumor focus positive for ERG fusion (rs11704416; $p = 0.033$; rs16901979; $p = 0.034$) (Supplementary Table 1).

Genotype imputation analysis

Imputation analysis of genome-wide Oncoarray (500,000 SNP) data was performed by the IMPUTE2 approach using the 1000 Genomes reference dataset. Imputed SNPs rs34349373 and rs2055272, two intronic variants in *TBC1D22B* (TBC1 Domain Family Member 22B), a GTPase activating protein for Rab family, were significantly ($< 10^{-6}$) associated with ERG positive phenotype in any tumor foci (Figure 3). The 2 variants are found to be in strong linkage disequilibrium in both CA and AA populations with r^2 of 1.0 and 0.91 respectively. Imputed SNP rs2055272 was further validated by TaqMan based ddPCR genotyping approach. Concordance between Taqman genotypes and imputed genotypes was 98.04% (100/102).

Association between the SNPs and clinicopathological status

The 9 SNPs (7 SNPs on Oncoarray chip and 2 imputed SNPs) were assessed for associations with clinicopathologic features, including pathological stage and grade at prostatectomy and incidence of biochemical recurrence. rs34349373 and rs2055272 were significantly ($p < 0.05$) associated with CaP ERG fusion status in both AA and CA patients, where the

variant allele is more frequent in ERG negative cases (Supplementary Table 2). However, these SNPs were not associated with pathological stage (pT stage) or Grade Group (GG). Kaplan–Meier analysis indicated no association between imputed SNPs and BCR when stratified by race or ERG status (Supplementary Figure 2). rs3798999 (intron variant in *ADGRB3*) and rs10215144 (intron variant in *AGBL3*) were significantly associated with pT stage ($p < 0.05$) and rs10215144 was also significantly associated ($p = 0.048$) with high grade CaP (GG4-5 vs. GG1-3) on a univariate analysis (Supplementary Table 3). Unadjusted univariate Kaplan–Meier analysis indicated that rs6698333 (intron variant in *KLF17*) was associated with BCR ($p = 0.032$) in both AA and CA patients, where carriers of the risk allele develop BCR significantly earlier during disease progression than the carriers of wild type (wt) allele (Figure 4A).

ERG status based unadjusted univariate analysis revealed that rs3798999 SNP was associated with the development of BCR in ERG negative patients (AA and CA combined; $p = 0.016$) and the association is also present in ERG negative AA patients ($p = 0.032$). The variant shows "protective effect" towards development of BCR (Figure 4B).

DISCUSSION

This is the first genetic epidemiological study on the association of genetic variants at a genome-wide scale, as opposed to selected SNPs, with *TMPRSS2: ERG* fusion status both by index tumor or by any tumor foci, considering the multifocal and multiclonal nature of the disease [9, 21]. It integrates data on inherited susceptibility and tumor ERG status within a well-defined cohort of men with a median longitudinal follow-up of 7.5 years. Two SNPs were identified: rs34349373 and rs2055272, in *TBC1D22B* on chromosome 6 to be associated with ERG
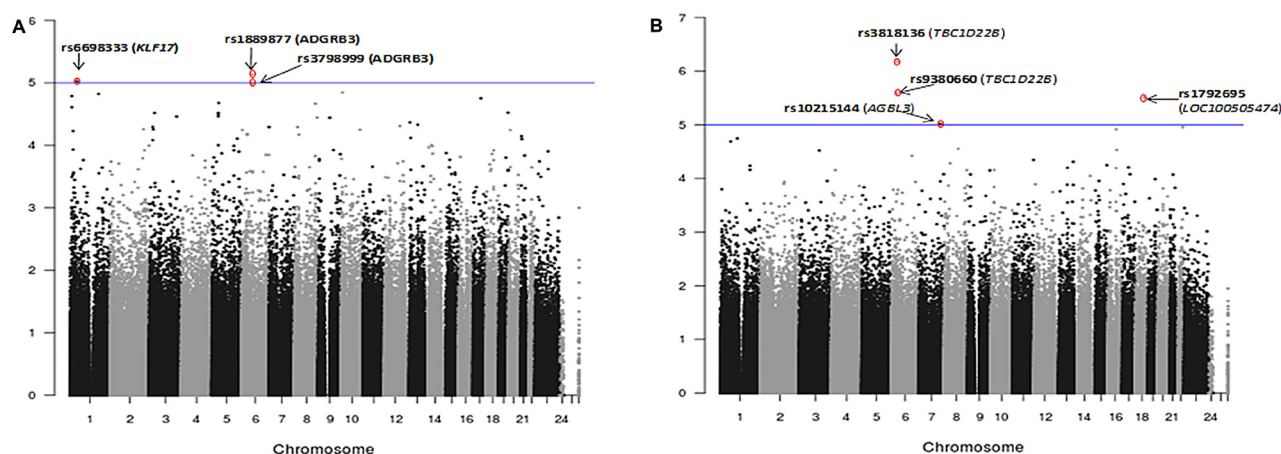

Figure 2: Manhattan plots showing association analysis (EMMAX) of SNPs with (**A**) ERG positive index tumor ($N = 108$) vs. ERG negative index tumor ($N = 180$). (**B**) Any tumor foci positive for ERG ($N = 158$) vs. ERG negative tumor ($N = 133$). A total of 478,299 SNPs are plotted against their respective positions on the chromosomes.

Table 2: Description of the 9 significant SNPs

Chr	SNP	Location	Alleles [minor allele/ common allele]	Transcript (s)	Gene (s)	*In-exon*	Mutation (s)	*P*-value	MAF
6	rs1889877	69729678	[A/G]	NM_001704.2	*ADGRB3 (BAI3)*	Intron Variant	NA	0.00549042	0.028
1	rs6698333	44554457	[T/C]	NM_173484.4	*KLF17*	Intron Variant	NA	0.0017769	0.489
6	rs3798999	69714947	[A/G]	NM_001704.2	*ADGRB3 (BAI3)*	Silent	NA	0.00017719	0.486
7	rs10215144	134765148	[A/G]	NM_178563	*AGBL3*	Silent	NA	0.00000952	0.461
6	rs3818136	37252210	[T/C]	NM_017772	*TBC1D22B*	EXON	Synonymous N257N	0.00000067	0.400
6	rs9380660	37305622	[T/G]	NA	NA	NA	NA	0.00000282	0.495
18	rs1792695	53782900	[T/G]	NR_040026 NR_040025	LOC100505474 LOC100505474	Silent, Silent	NA	0.00000319	0.273
6	rs34349373	37254109	-/T	NM_017772	*TBC1D22B*	intron variant	NA	0.00000005	0.419
6	rs2055272	37289781	A/G	NM_017772	*TBC1D22B*	intron variant	NA	0.00000006	0.396

Abbreviations: Chr: chromosome; MAF: minor allele frequency; SNP: single nucleotide polymorphism; NA: not applicable. Location Information is based on genome assembly- GRCh37 (hg19). *P* value is based on EMMAX analysis using an additive model.

status of CaP, a major driver oncogene in CaP. Minor allele frequency (MAF) reveals that del variant for rs34349373 and A allele variant for rs2055272 are significantly lower in ERG positive cases compared to wild type (wt) in both AA and CA men, thereby implying the association of the SNPs with the fusion negative subtype of CaP. These two SNPs are also in strong linkage disequilibrium (LD; $r^2 > 0.9$). After initial validation, it will be important to understand the mechanism by which SNPs influence the formation of fusion protein. rs34349373 (-/T) is an intronic deletion/insertion variation while rs2055272 (A/G) is an intronic SNV in TBC1 domain family, member 22B (*TBC1D22B*) gene with unknown clinical relevance (dbSNP). TBC1D22B is a GTPase activating protein for Rab protein family, and is found to be over-expressed in many cancers.

Comprehensive genetic and epigenetic analyses including our study suggest that tumors with *TMPRSS2-ERG* fusion exhibit different genetic etiology compared to fusion negatives. Similar associations have been observed for GWAS identified genomic loci with the risk of ER-negative disease in breast cancer subtypes [54]. Further, functional analysis based on long range chromatin interactomes analysis in CaP cells has shown strong enrichment of CaP GWAS SNPs at AR-ERG co-binding sites participating in chromatin interactions and gene regulation, suggesting potential functional role of these SNPs towards specific ERG subtype [55].

We hypothesized that SNPs may influence the generation of ERG fusion, which is an early event in CaP carcinogenesis. A genome-wide linkage analysis found that several loci located on chromosomes 9, 18, and X are associated with the development of fusion-positive prostate cancer; however, these studies were performed in familial prostate cancer [56, 57].

There are two studies which focused on candidate SNP approach, covering known GWAS risk variants. The first study, Physicians Health Study (PHS) and Health Professionals Follow-up Study (HPFS), examined 39 known risk variants in a patient cohort of 227 ERG fusion-positive and 260 ERG negative cases [51, 52]. Six SNPs at 8q24, 5p15, 8p21, 17q24, 19q13 and 22q13 were found to be significantly associated with ERG fusion status. The second study by Luedeke *et al.* examined 27 common CaP risk variants using case-case comparison approach on 296 *TMPRSS2: ERG* fusion-positive versus 256 fusion-negative cases, alongwith an independent validation of significant SNPs in a patient cohort of 669 cases. The study found that variants at 8q24 and 17q24 were significantly linked with *TMPRSS2-ERG* fusion status. Interestingly, in both studies, rs1859962 at 17q24 was identified to be significantly associated with fusion positive CaP and the variants in the risk loci of 8q24 were found to be over-represented in fusion negative CaP, implying the role of 8q24 region towards fusion negative subtype of CaP. The present study also replicated 2 of the known ERG associated SNPs (rs11704416 in chromsome 22; rs16901979 in chromosome 8) from PHS and HPFS study. The SNPs were associated with ERG fusion status either by index tumor or by any tumor foci positive for ERG. rs16901979 has been associated with increased risk for CaP in several studies including AA men [36, 58, 59]. SNP rs11704416 was found to be associated with aggressive CaP in a meta-analysis of four GWAS including 5,953 cases of aggressive CaP and 11,463 controls [60].

An earlier study by FitzGerald [16] *et al.* based on association of 5 candidate SNPs in *ERG* and *TMPRSS2* in a cohort of 127 patients, showed a positive association for rs12329760 in chromosome 21 in *TMPRSS2* for fusion

C
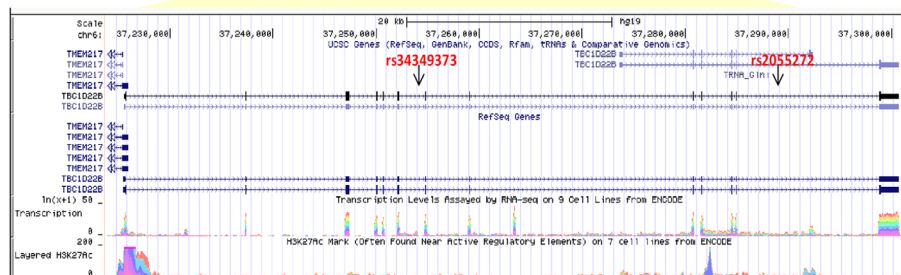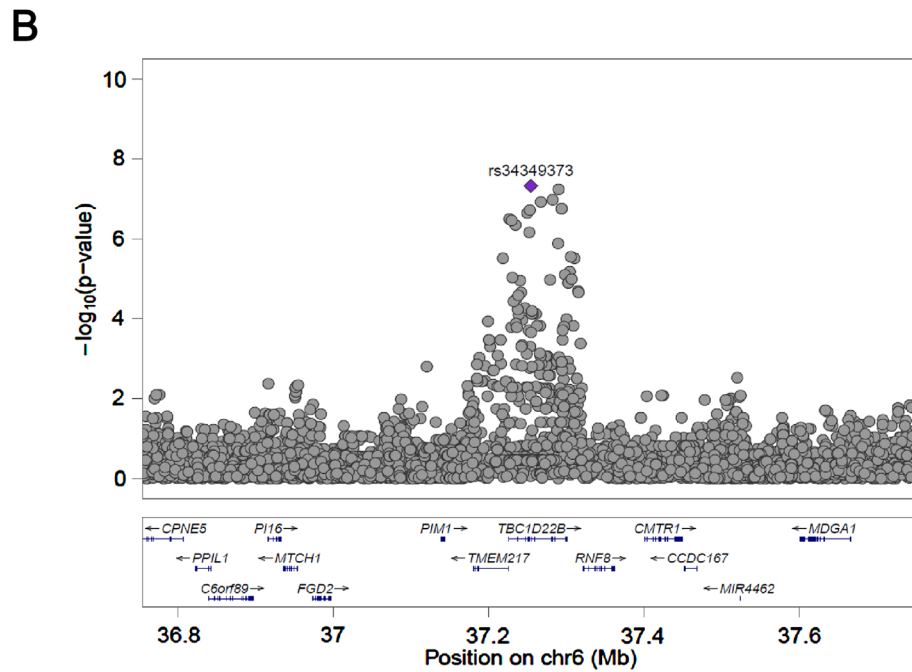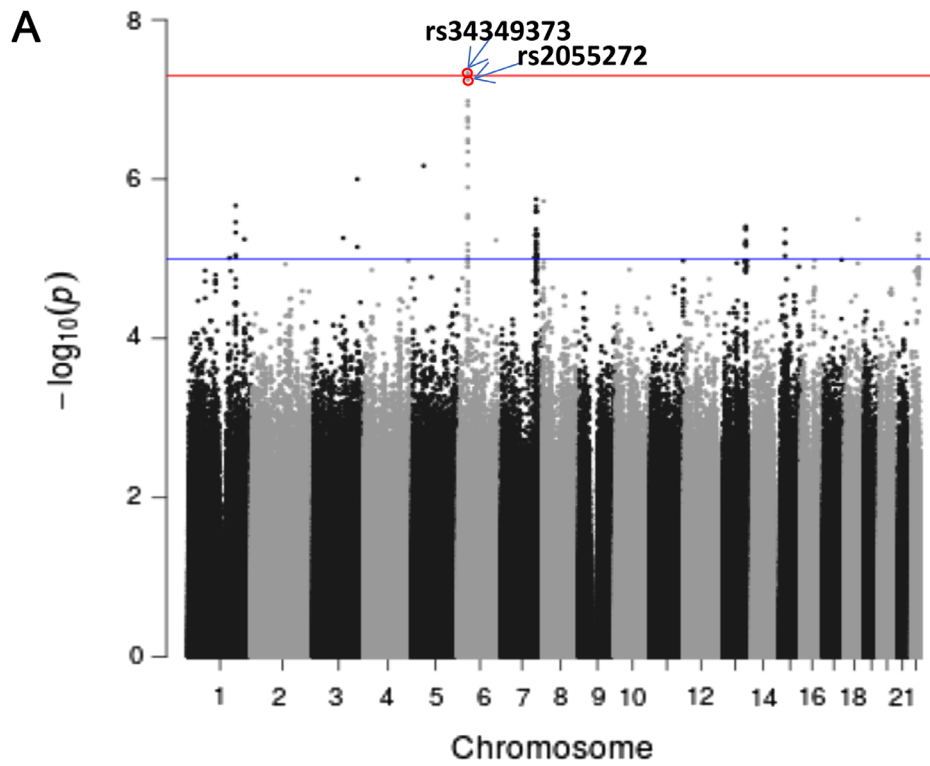
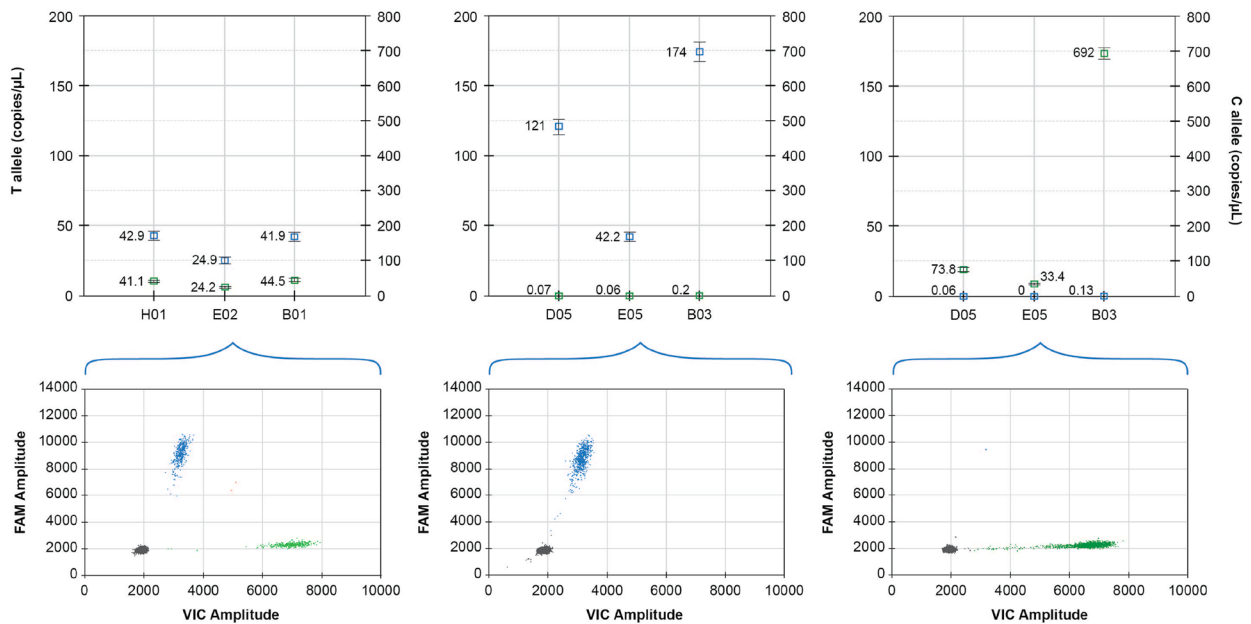| rs34349373	del/del	del/T	TT	del	T
		Genotype Frequency		**Allele Frequency**	
ERG+	10	71	67	0.31	0.69
ERG-	30	71	23	0.53	0.47
rs2055272	**A/A**	**A/G**	**G/G**	**A**	**G**
ERG+	9	72	72	0.29	0.71
ERG-	28	74	27	0.5	0.49

D

Figure 3: Fine-mapping of genetic associations by imputation analysis (**A**) Imputation analysis (IMPUTE2) of SNPs with any tumor foci positive for ERG phenotype in a total of 158 ERG positive and 133 ERG negative cases. A total of 13 million imputed SNPs with MAF > 1% SNPs are plotted against their respective positions on the chromosomes. (**B**) Plots show association results of imputed SNPs. -log10 P values (y axis) of the SNPs are shown according to their chromosomal positions (x axis). The Genome Browser annotation track page zoomed in to display the rs34349373 and rs2055272 (intron variants) in TBC1D22B gene on human chromosome 6, Feb 2009 assembly (hg19) (**C**) Genotype and allele frequencies of the rs34349373 and rs2055272 polymorphisms in ERG positive (by any tumor foci) vs. ERG negative CaP (**D**) Representative graph for SNP genotyping for rs2055272 (C/T) using droplet digital PCR (ddPCR) approach. Upper panel shows concentration (copies/ul) of FAM allele (T; Channel 1) and VIC allele (C; Channel 2) in set of representative samples with 3 genotypes (CT, TT, CC). Lower panel is 2-D Amplitude view where each axis represents the amplitude of fluorescence for either FAM (vertical axis) or VIC (horizontal axis). The FAM probe can hybridize only to the alternate allele (T allele), while the VIC probe hybridizes only to reference allele (C allele).

subtype resulting from translocation. Another study showed that shorter germline CAG repeat length in *AR* [61] and rare variants in DNA repair genes, *ESCO1* N191S and *POLI* F532S [51] were associated with higher risk of ERG-positive CaP.

However, these studies are mainly based on a few candidate SNPs, in a population predominantly represented by Caucasians. Therefore, it is important to explore the SNP association on a genomewide scale in context of race, as ERG is expressed almost 2 times more frequent in CA than AA CaP. Additionally, in these

studies, only one tumor focus was evaluated per specimen which does not take into account focal heterogeneity of the disease. Limitations of the present study includes: a) lack of independent validation b) limited generalizability of the CPDR cohort to other US longitudinal cohorts, as well as RP patients in other nations [7].

In summary, this study identified the association of 2 SNPs (rs34349373 and rs2055272) in *TBC1D22B* with ERG fusion status where the minor alleles are associated with an ERG negative subtype of CaP. Additionally, rs3798999 SNP (*ADGRB3*) was significantly associated

with the development of BCR in ERG negative patient cohort. Validation study in independent large patient cohort with race stratified analysis, and functional understanding of the biology of these SNPs in relation to ERG phenotype are warranted. Overall, this study may contribute toward defining the underlying biology and genetics of ERG positive and ERG negative CaP in AA and CA patients.

MATERIALS AND METHODS

DNA specimens

In this retrospective cohort-based study, 400 archived genomic DNA specimens were used from blood of CaP patients undergoing radical prostatectomy treatment at Walter Reed National Military Medical Center (WRNMMC) under an IRB approved protocol. DNA was extracted from peripheral blood lymphocytes using Qiagen DNeasy Blood kit. Archived clinicopathological data were evaluated from the 400 patients who self-identified their race as AA or CA.

TMPRSS2-ERG fusion status

ERG status was determined by immuno-histochemistry (IHC) for ERG protein expression, as a surrogate for the *TMPRSS2-ERG* fusion. For ERG IHC analysis, representative whole-mount 4-um cross section encompassing tumor foci with highest grade and/or stage from each prostatectomy specimen were processed and stained with a highly specific anti-ERG monoclonal antibody (clone 9FY; Biocare Medical Inc., Concord, CA, USA) as previously described [8, 62]. The index tumor was identified as the tumor with the largest volume if all foci have the same grade, or with the highest Grade Group. Multiple tumor foci (average no. 5) in representative whole-mount prostate sections were evaluated, per patient, for the presence or absence of the ERG oncoprotein. The patient was called ERG positive when any of the tumor foci was positive. Slide selection, tumor grading, and staining interpretation were performed by a single pathologist.

SNP genotyping (infinium oncoarray, a 500K BeadChip)

SNP genotyping was performed on the Illumina Golden Gate genotyping platform using Infinium Oncoarray, a 500K BeadChip on a genome wide scale in 400 CaP patients. The oncoarray contains approximately 500,000 SNPs including 275,000 tagSNPs. It includes SNPs covering common ancestry (1,500 SNPs), genetic variants associated with 5 common cancers (breast,
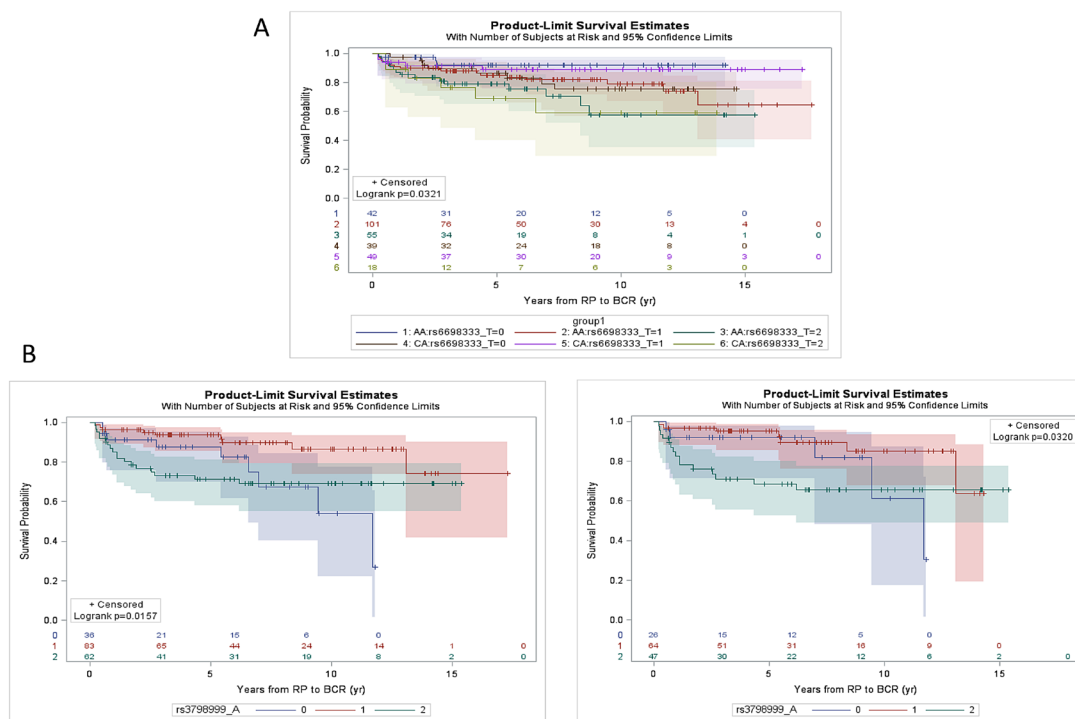

Figure 4: Kaplan-Meier estimation curve of time to BCR as a function of SNPs. (**A**) rs6698333 SNP [Wt (CC), 0 vs Heterozygous (CT), 1 vs Polymorphic homozygous (TT), 2] for CaP patients in AA (*n* = 198) and CA patients (*n* = 106). The log-rank *p* value (*p* = 0.0321) indicates that there is an association between SNP and BCR over time. (**B**) rs3798999 SNP [Wt (GG), 0 vs Heterozygous (GA), 1 vs Polymorphic homozygous (AA), 2] for CaP patients in ERG negative patients (AA and CA combined; *n* = 181; left panel) and for ERG negative AA patients only (*n* = 137; right panel). The SNP was found to be associated with BCR in ERG negative patient cohort (*p* = 0.016) specifically in AA patients (*p* = 0.032).

colorectal, lung, ovarian and prostate) as well as SNPs covering quantitative traits, pharmacogenetics, and fine mapping of common cancer susceptibility loci. It includes 80,000 prostate cancer specific genetic variants. A total of 496654 SNPs genotype calls were generated from the 400 specimens. Data quality control (QC) was performed by applying sample and SNP QC using PLINK [63]. 14,729 SNPs were excluded due to SNP call rate < 0.90. Additional 1,482 SNPs were excluded due to deviation from Hardy–Weinberg Equilibrium (HWE) test (*P*-value $< 1 \times 10^{-8}$). 48 samples were excluded due to sample call rate < 0.95, and 33 samples were further excluded based on plink proportion identity-by-descent (PI_HAT) > 0.15) [64]. The final QCed genotype data included 478,299 SNPs from a total of 321 patients (AA = 216 and CA = 105).

Statistical analysis

Association analysis was based on Efficient Mixed-Model Association eXpedited (EMMAX) which accounts for population structure, including relatedness between cases [65]. The analysis was performed based on the additive model. Genotype imputation analysis was performed on oncoarray SNP array dataset using IMPUTE2 based on the 1000 Genomes Project data (phase 3) [66, 67]. Imputation analysis was based on a total of 13 million SNPs with MAF > 1%, where we used only well-imputed variants (IMPUTE2 info score ≥ 0.9; MAF >1%). For the initial discovery analysis, we have used lower significant threshold (1×10^{-5}) to identify possible genetic loci associated ERG fusion. However, for the imputation approach, we have used a higher significant threshold (1×10^{-8}) to discover probable causal variants associated with ERG fusion. Chi-square testing or Fisher exact test were used to evaluate the associations of genotypes with clinical-pathological variables. Unadjusted Kaplan-Meier survival analysis and log-rank testing were used to show the probability of BCR-free survival stratified by genotypes. All statistical analysis was performed using SAS version 9.4 (North Carolina) and statistical significance was set at *p* < 0.05.

TaqMan SNP genotyping

Imputation results for rs2055272 were further validated by TaqMan genotyping (assay C_3025729_20, ThermoFisher Scientific) in 102 patients (98.03% concordance). Validation was performed by droplet digital PCR (ddPCR) approach using a QX200 Droplet Generator (BioRad) and the data was analyzed by QuantaSoft software (BioRad). Briefly, a ddPCR mastermix was prepared containing 11 μl 2× ddPCR Supermix (Bio-Rad), 1.1 μl 20× Taqman SNP Genotyping Assay (Applied Biosystems), and 7.9 μl nuclease-free water (Qiagen) per sample. The mastermix was prepared at room temperature

and 20 μl was added to 2 μl (5 ng) of each DNA sample. Samples were loaded into individual wells of DG8TM cartridges (BioRad), and droplets were generated using a QX200 Droplet Generator (BioRad). For each sample, 40 μl of droplet mix was then transferred to a 96-well plate, and PCR was performed in a thermal cycler using the following cycling conditions: 95°C × 10 min; 40 cycles of [94°C × 30s, 60°C × 60s]; 98°C × 10s; 40°C × 10 min. The Bio-Rad QX200 Droplet Reader was then used to assess droplets as positive or negative based on fluorescence amplitude. The QuantaSoft software (BioRad) was used to analyze droplet data.

Abbreviations

CaP: prostate cancer; RP: radical prostatectomy; DNA: deoxy-ribonucleic acid; SNP: single nucleotide polymorphism; ddPCR: droplet digital PCR; RT-PCR: real-time polymerase chain reaction; LD: linkage disequilibrium; MAF: minor allele frequency; BCR: biochemical recurrence; FFPE: formalin-fixed-paraffin embedded; H&E: hematoxylin and eosin; LCM: laser-capture-microdissection; MAb: monoclonal antibody.

Author contributions

LR and DY carried out sample preparation, IK performed SNP genotyping and RT-PCR experiments, or involved in the acquisition of data. IK, QL, JC, YC, MF and GP analyzed and interpreted the data. GP, MF and SS conceived the study and participated in the design and coordination. JC, ILR and IAS selected the patient cohort. AA and ILR obtained informed consent from the patients. IAS performed pathologic examination of the patient specimens, analyzed and interpreted the ERG expression status. IK, JC, QL, MF and GP contributed to the discussion and preparation of the manuscript. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no potential conflicts of interest.

FUNDING

Disclaimer

The contents of this publication are the sole responsibility of the author(s) and do not necessarily reflect the views, opinions or policies of the Henry M.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018; 68:394–424. https://doi.org/10.3322/caac.21492. [PubMed]

2. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015; 136:E359–E386. https://doi.org/10.1002/ijc.29210. [PubMed]

3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018; 68:7–30. https://doi.org/10.3322/caac.21442. [PubMed]

4. Sedarsky J, Degon M, Srivastava S, Dobi A. Ethnicity and ERG frequency in prostate cancer. Nat Rev Urol. 2018; 15:125–131. https://doi.org/10.1038/nrurol.2017.140. [PubMed]

5. Petrovics G, Liu A, Shaheduzzaman S, Furusato B, Sun C, Chen Y, Nau M, Ravindranath L, Dobi A, Srikantan V, Sesterhenn IA, McLeod DG, Vahey M, et al. Frequent overexpression of ETS-related gene-1 (ERG1) in prostate cancer transcriptome. Oncogene. 2005; 24:3847–3852. [PubMed]

6. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science. 2005; 310:644–648. https://doi.org/10.1126/science.1117679. [PubMed]

7. Cullen J, Young D, Chen Y, Degon M, Farrell J, Sedarsky J, Baptiste W, Rosen P, Tolstikov V, Kiebish M, Kagan J, Srivastava S, Kuo HC, et al. Predicting Prostate Cancer Progression as a Function of ETS-related Gene Status, Race, and Obesity in a Longitudinal Patient Cohort. Eur Urol Focus. 2018; 4:818–824. https://doi.org/10.1016/j.euf.2017.02.016. [PubMed]

8. Rosen P, Pfister D, Young D, Petrovics G, Chen Y, Cullen J, Böhm D, Perner S, Dobi A, McLeod DG, Sesterhenn IA, Srivastava S. Differences in frequency of ERG oncoprotein expression between index tumors of Caucasian and African American patients with prostate cancer. Urology. 2012; 80:749–753. https://doi.org/10.1016/j.urology.2012.07.001. [PubMed]

9. Farrell J, Petrovics G, McLeod DG, Srivastava S. Genetic and molecular differences in prostate carcinogenesis between African American and Caucasian American men. Int J Mol Sci. 2013; 14:15510–15531. https://doi.org/10.3390/ijms140815510. [PubMed]

10. Farrell J, Young D, Chen Y, Cullen J, Rosner IL, Kagan J, Srivastava S, McLeod DG, Sesterhenn IA, Petrovics G. Predominance of ERG-negative high-grade prostate cancers in African American men. Mol Clin Oncol. 2014; 2:982–986. https://doi.org/10.3892/mco.2014.378. [PubMed]

11. Nam RK, Sugar L, Yang W, Srivastava S, Klotz LH, Yang LY, Stanimirovic A, Encioiu E, Neill M, Loblaw DA, Trachtenberg J, Narod SA, Seth A. Expression of the TMPRSS2:ERG fusion gene predicts cancer recurrence after surgery for localised prostate cancer. Br J Cancer. 2007; 97:1690–1695. https://doi.org/10.1038/sj.bjc.6604054. [PubMed]

12. Wang J, Cai Y, Ren C, Ittmann M. Expression of variant TMPRSS2/ERG fusion messenger RNAs is associated with aggressive prostate cancer. Cancer Res. 2006; 66:8347–8351. https://doi.org/10.1158/0008-5472.CAN-06-1966. [PubMed]

13. Yoshimoto M, Joshua AM, Cunha IW, Coudry RA, Fonseca FP, Ludkovski O, Zielenska M, Soares FA, Squire JA. Absence of TMPRSS2:ERG fusions and PTEN losses in prostate cancer is associated with a favorable outcome. Mod Pathol. 2008; 21:1451–1460. https://doi.org/10.1038/modpathol.2008.96. [PubMed]

14. Attard G, Clark J, Ambroisine L, Fisher G, Kovacs G, Flohr P, Berney D, Foster CS, Fletcher A, Gerald WL, Moller H, Reuter V, De Bono JS, et al; Transatlantic Prostate Group. Duplication of the fusion of TMPRSS2 to ERG sequences identifies fatal human prostate cancer. Oncogene. 2008; 27:253–263. https://doi.org/10.1038/sj.onc.1210640. [PubMed]

15. Demichelis F, Fall K, Perner S, Andren O, Schmidt F, Setlur SR, Hoshida Y, Mosquera JM, Pawitan Y, Lee C, Adami HO, Mucci LA, Kantoff PW, et al. TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. Oncogene. 2007; 26:4596–4599. https://doi.org/10.1038/sj.onc.1210237. [PubMed]

16. FitzGerald LM, Agalliu I, Johnson K, Miller MA, Kwon EM, Hurtado-Coll A, Fazli L, Rajput AB, Gleave ME, Cox ME, Ostrander EA, Stanford JL, Huntsman DG. Association of TMPRSS2-ERG gene fusion with clinical characteristics and outcomes: results from a population-based study of prostate cancer. BMC Cancer. 2008; 8:230. https://doi.org/10.1186/1471-2407-8-230. [PubMed]

17. Börno ST, Fischer A, Kerick M, Fälth M, Laible M, Brase JC, Kuner R, Dahl A, Grimm C, Sayanjali B, Isau M, Röhr C, Wunderlich A, et al. Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. Cancer Discov. 2012; 2:1024–1035. https://doi.org/10.1158/2159-8290.CD-12-0041. [PubMed]

18. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;

163:1011–1025. https://doi.org/10.1016/j.cell.2015.10.025. [PubMed]

19. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N, MacDonald TY, Ghandi M, Van Allen E, Kryukov GV, Sboner A, et al. Punctuated evolution of prostate cancer genomes. Cell. 2013; 153:666–677. https://doi.org/10.1016/j.cell.2013.03.021. [PubMed]

20. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, McLaren S, Lin ML, McBride DJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 2011; 144:27–40. https://doi.org/10.1016/j.cell.2010.11.055. [PubMed]

21. Haffner MC, Mosbruger T, Esopi DM, Fedor H, Heaphy CM, Walker DA, Adejola N, Gurel M, Hicks J, Meeker AK, Halushka MK, Simons JW, Isaacs WB, et al. Tracking the clonal origin of lethal prostate cancer. J Clin Invest. 2013; 123:4918–4922. https://doi.org/10.1172/JCI70354. [PubMed]

22. MacInnis RJ, Antoniou AC, Eeles RA, Severi G, Guy M, McGuffog L, Hall AL, O'Brien LT, Wilkinson RA, Dearnaley DP, Ardern-Jones AT, Horwich A, Khoo VS, et al. Prostate cancer segregation analyses using 4390 families from UK and Australian population-based studies. Genet Epidemiol. 2010; 34:42–50. https://doi.org/10.1002/gepi.20433. [PubMed]

23. Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, Graff RE, Holst K, Moller S, Unger RH, McIntosh C, Nuttall E, Brandt I, et al; Nordic Twin Study of Cancer (NorTwinCan) Collaboration. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. JAMA. 2016; 315:68–76. https://doi.org/10.1001/jama.2015.17703. [PubMed]

24. Castro E, Goh C, Olmos D, Saunders E, Leongamornlert D, Tymrakiewicz M, Mahmud N, Dadaev T, Govindasami K, Guy M, Sawyer E, Wilkinson R, Ardern-Jones A, et al. Germline BRCA mutations are associated with higher risk of nodal involvement, distant metastasis, and poor survival outcomes in prostate cancer. J Clin Oncol. 2013; 31:1748–1757. https://doi.org/10.1200/JCO.2012.43.1882. [PubMed]

25. Bancroft EK, Page EC, Castro E, Lilja H, Vickers A, Sjoberg D, Assel M, Foster CS, Mitchell G, Drew K, Maehle L, Axcrona K, Evans DG, et al. Targeted prostate cancer screening in BRCA1 and BRCA2 mutation carriers: results from the initial screening round of the IMPACT study. Eur Urol. 2014; 66:489–499. https://doi.org/10.1016/j.eururo.2014.01.003. [PubMed]

26. Pomerantz MM, Spisák S, Jia L, Cronin AM, Csabai I, Ledet E, Sartor AO, Rainville I, O'Connor EP, Herbert ZT, Szállási Z, Oh WK, Kantoff PW, et al. The association between germline BRCA2 variants and sensitivity to platinum-based chemotherapy among men with metastatic prostate cancer. Cancer. 2017; 123:3532–3539. https://doi.org/10.1002/cncr.30808. [PubMed]

27. Al Olama AA, Kote-Jarai Z, Berndt SI, Conti DV, Schumacher F, Han Y, Benlloch S, Hazelett DJ, Wang Z, Saunders E, Leongamornlert D, Lindstrom S, Jugurnauth-Little S, et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. Nat Genet. 2014; 46:1103–1109. https://doi.org/10.1038/ng.3094. [PubMed]

28. Al Olama AA, Kote-Jarai Z, Giles GG, Guy M, Morrison J, Severi G, Leongamornlert DA, Tymrakiewicz M, Jhavar S, Saunders E, Hopper JL, Southey MC, Muir KR, et al; UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK Prostate testing for cancer and Treatment study (ProtecT Study) Collaborators. Multiple loci on 8q24 associated with prostate cancer susceptibility. Nat Genet. 2009; 41:1058–1060. https://doi.org/10.1038/ng.452. [PubMed]

29. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet. 2007; 39:631–637. https://doi.org/10.1038/ng1999. [PubMed]

30. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, Wang Z, Welch R, Staats BJ, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet. 2007; 39:645–649. https://doi.org/10.1038/ng2022. [PubMed]

31. Yeager M, Chatterjee N, Ciampa J, Jacobs KB, Gonzalez-Bosquet J, Hayes RB, Kraft P, Wacholder S, Orr N, Berndt S, Yu K, Hutchinson A, Wang Z, et al. Identification of a new prostate cancer susceptibility locus on chromosome 8q24. Nat Genet. 2009; 41:1055–1057. https://doi.org/10.1038/ng.444. [PubMed]

32. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Cazier JB, Sainz J, Jakobsdottir M, Kostic J, Magnusdottir DN, et al. A common variant associated with prostate cancer in European and African populations. Nat Genet. 2006; 38:652–658. https://doi.org/10.1038/ng1808. [PubMed]

33. Ahmed M, Eeles R. Germline genetic profiling in prostate cancer: latest developments and potential clinical applications. Future Sci OA. 2016; 2:FSO87. https://doi.org/10.4155/fso.15.87. [PubMed]

34. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet. 2007; 39:977–983. https://doi.org/10.1038/ng2062. [PubMed]

35. Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, Gudbjartsson D, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Blondal T, Jakobsdottir M, Stacey SN, Kostic J, et al. Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. Nat Genet. 2008; 40:281–283. https://doi.org/10.1038/ng.89. [PubMed]

36. Gudmundsson J, Sulem P, Gudbjartsson DF, Blondal T, Gylfason A, Agnarsson BA, Benediktsdottir KR, Magnusdottir DN, Orlygsdottir G, Jakobsdottir M, Stacey SN, Sigurdsson A, Wahlfors T, et al. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. Nat Genet. 2009; 41:1122–1126. https://doi.org/10.1038/ng.448. [PubMed]

37. Gudmundsson J, Besenbacher S, Sulem P, Gudbjartsson DF, Olafsson I, Arinbjarnarson S, Agnarsson BA, Benediktsdottir KR, Isaksson HJ, Kostic JP, Gudjonsson SA, Stacey SN, Gylfason A, et al. Genetic correction of PSA values using sequence variants associated with PSA levels. Sci Transl Med. 2010; 2:62ra92. https://doi.org/10.1126/scitranslmed.3001513. [PubMed]

38. Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, Field HI, Southey MC, Severi G, et al. Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet. 2008; 40:316–321. https://doi.org/10.1038/ng.90. [PubMed]

39. Eeles RA, Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, Ghoussaini M, Luccarini C, Dennis J, Jugurnauth-Little S, Dadaev T, Neal DE, Hamdy FC, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat Genet. 2013; 45:385–91, 91e1–2. https://doi.org/10.1038/ng.2560. [PubMed]

40. Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, Neubauer J, Tandon A, Schirmer C, McDonald GJ, Greenway SC, Stram DO, Le Marchand L, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet. 2007; 39:638–644. https://doi.org/10.1038/ng2015. [PubMed]

41. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, Dadaev T, Leongamornlert D, Anokian E, Cieza-Borrella C, Goh C, Brook MN, Sheng X, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet. 2018; 50:928–936. https://doi.org/10.1038/s41588-018-0142-8. [PubMed]

42. Benafif S, Kote-Jarai Z, Eeles RA, and PRACTICAL Consortium. A Review of Prostate Cancer Genome-Wide Association Studies (GWAS). Cancer Epidemiol Biomarkers Prev. 2018; 27:845–857. https://doi.org/10.1158/1055-9965.EPI-16-1046. [PubMed]

43. Conti DV, Wang K, Sheng X, Bensen JT, Hazelett DJ, Cook MB, Ingles SA, Kittles RA, Strom SS, Rybicki BA, Nemesure B, Isaacs WB, Stanford JL, et al, and PRACTICAL/ELLIPSE Consortium. Two Novel Susceptibility Loci for Prostate Cancer in Men of African Ancestry. J Natl Cancer Inst. 2017; 109. https://doi.org/10.1093/jnci/djx084. [PubMed]

44. Han Y, Rand KA, Hazelett DJ, Ingles SA, Kittles RA, Strom SS, Rybicki BA, Nemesure B, Isaacs WB, Stanford JL, Zheng W, Schumacher FR, Berndt SI, et al. Prostate Cancer Susceptibility in Men of African Ancestry at 8q24. J Natl Cancer Inst. 2016; 108:djv431. https://doi.org/10.1093/jnci/djv431. [PubMed]

45. Cook MB, Wang Z, Yeboah ED, Tettey Y, Biritwum RB, Adjei AA, Tay E, Truelove A, Niwa S, Chung CC, Chokkalingam AP, Chu LW, Yeager M, et al, and African Ancestry Prostate Cancer GWAS Consortium. A genome-wide association study of prostate cancer in West African men. Hum Genet. 2014; 133:509–521. https://doi.org/10.1007/s00439-013-1387-z. [PubMed]

46. Hoffmann TJ, Van Den Eeden SK, Sakoda LC, Jorgenson E, Habel LA, Graff RE, Passarelli MN, Cario CL, Emami NC, Chao CR, Ghai NR, Shan J, Ranatunga DK, et al. A large multiethnic genome-wide association study of prostate cancer identifies novel risk variants and substantial ethnic differences. Cancer Discov. 2015; 5:878–891. https://doi.org/10.1158/2159-8290.CD-15-0315. [PubMed]

47. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. Cell. 2019; 177:26–31. https://doi.org/10.1016/j.cell.2019.02.048. [PubMed]

48. Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A, Penney K, Steen RG, Ardlie K, John EM, Oakley-Girvan I, Whittemore AS, Cooney KA, et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc Natl Acad Sci U S A. 2006; 103:14068–14073. https://doi.org/10.1073/pnas.0605832103. [PubMed]

49. Whitman EJ, Pomerantz M, Chen Y, Chamberlin MM, Furusato B, Gao C, Ali A, Ravindranath L, Dobi A, Sesterhenn IA, McLeod DG, Srivastava S, Freedman M, et al. Prostate cancer risk allele specific for African descent associates with pathologic stage at prostatectomy. Cancer Epidemiol Biomarkers Prev. 2010; 19:1–8. https://doi.org/10.1158/1055-9965.EPI-08-1111. [PubMed]

50. Perner S, Mosquera JM, Demichelis F, Hofer MD, Paris PL, Simko J, Collins C, Bismar TA, Chinnaiyan AM, De Marzo AM, Rubin MA. TMPRSS2-ERG fusion prostate cancer: an early molecular event associated with invasion. Am J Surg Pathol. 2007; 31:882–888. https://doi.org/10.1097/01.pas.0000213424.38503.aa. [PubMed]

51. Luedeke M, Linnert CM, Hofer MD, Surowy HM, Rinckleb AE, Hoegel J, Kuefer R, Rubin MA, Vogel W, Maier C. Predisposition for TMPRSS2-ERG fusion in prostate cancer by variants in DNA repair genes. Cancer Epidemiol Biomarkers Prev. 2009; 18:3030–3035. https://doi.org/10.1158/1055-9965.EPI-09-0772. [PubMed]

52. Penney KL, Pettersson A, Shui IM, Graff RE, Kraft P, Lis RT, Sesso HD, Loda M, Mucci LA. Association of Prostate Cancer Risk Variants with TMPRSS2:ERG Status: Evidence for Distinct Molecular Subtypes. Cancer Epidemiol Biomarkers Prev. 2016; 25:745–749. https://doi.org/10.1158/1055-9965.EPI-15-1078. [PubMed]

53. Luedeke M, Rinckleb AE, FitzGerald LM, Geybels MS, Schleutker J, Eeles RA, Teixeira MR, Cannon-Albright L, Ostrander EA, Weikert S, Herkommer K, Wahlfors T, Visakorpi T, et al. Prostate cancer risk regions at 8q24 and 17q24 are

differentially associated with somatic TMPRSS2:ERG fusion status. Hum Mol Genet. 2016; 25:5490–5499. https://doi.org/10.1093/hmg/ddw349. [PubMed]

54. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, Orr N, Rhie SK, Riboli E, Feigelson HS, Le Marchand L, Buring JE, Eccles D, et al, and Gene ENvironmental Interaction and breast CAncer (GENICA) Network, and kConFab Investigators, and Familial Breast Cancer Study (FBCS), and Australian Breast Cancer Tissue Bank (ABCTB) Investigators. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet. 2013; 45:392–8, 8e1–2. https://doi.org/10.1038/ng.2561. [PubMed]

55. Zhang Z, Chng KR, Lingadahalli S, Chen Z, Liu MH, Do HH, Cai S, Rinaldi N, Poh HM, Li G, Sung YY, Heng CL, Core LJ, et al. An AR-ERG transcriptional signature defined by long-range chromatin interactomes in prostate cancer cells. Genome Res. 2019; 29:223–235. https://doi.org/10.1101/gr.230243.117. [PubMed]

56. Hofer MD, Kuefer R, Maier C, Herkommer K, Perner S, Demichelis F, Paiss T, Vogel W, Rubin MA, Hoegel J. Genome-wide linkage analysis of TMPRSS2-ERG fusion in familial prostate cancer. Cancer Res. 2009; 69:640–646. https://doi.org/10.1158/0008-5472.CAN-08-2008. [PubMed]

57. Edwards S, Meitz J, Eles R, Evans C, Easton D, Hopper J, Giles G, Foulkes WD, Narod S, Simard J, Badzioch M, Mahle L; International ACTANE Consortium. Results of a genome-wide linkage analysis in prostate cancer families ascertained through the ACTANE consortium. Prostate. 2003; 57:270–279. https://doi.org/10.1002/pros.10301. [PubMed]

58. Xu J, Kibel AS, Hu JJ, Turner AR, Pruett K, Zheng SL, Sun J, Isaacs SD, Wiley KE, Kim ST, Hsu FC, Wu W, Torti FM, et al. Prostate cancer risk associated loci in African Americans. Cancer Epidemiol Biomarkers Prev. 2009; 18:2145–2149. https://doi.org/10.1158/1055-9965.EPI-09-0091. [PubMed]

59. Zheng SL, Sun J, Wiklund F, Smith S, Stattin P, Li G, Adami HO, Hsu FC, Zhu Y, Bälter K, Kader AK, Turner AR, Liu W, et al. Cumulative association of five genetic variants with prostate cancer. N Engl J Med. 2008; 358:910–919. https://doi.org/10.1056/NEJMoa075819. [PubMed]

60. Amin Al Olama A, Kote-Jarai Z, Schumacher FR, Wiklund F, Berndt SI, Benlloch S, Giles GG, Severi G, Neal DE, Hamdy FC, Donovan JL, Hunter DJ, Henderson BE, et al, and UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology, and UK ProtecT Study Collaborators, and Australian Prostate Cancer Bioresource, and PRACTICAL Consortium. A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease. Hum Mol Genet. 2013; 22:408–415. https://doi.org/10.1093/hmg/dds425. [PubMed]

61. Yoo S, Pettersson A, Jordahl KM, Lis RT, Lindstrom S, Meisner A, Nuttall EJ, Stack EC, Stampfer MJ, Kraft P, Brown M, Loda M, Giovannucci EL, et al. Androgen receptor CAG repeat polymorphism and risk of TMPRSS2:ERG-positive prostate cancer. Cancer Epidemiol Biomarkers Prev. 2014; 23:2027–2031. https://doi.org/10.1158/1055-9965.EPI-14-0020. [PubMed]

62. Furusato B, Tan SH, Young D, Dobi A, Sun C, Mohamed AA, Thangapazham R, Chen Y, McMaster G, Sreenath T, Petrovics G, McLeod DG, Srivastava S, et al. ERG oncoprotein expression in prostate cancer: clonal progression of ERG-positive tumor cells and potential for ERG-based stratification. Prostate Cancer Prostatic Dis. 2010; 13:228–237. https://doi.org/10.1038/pcan.2010.23. [PubMed]

63. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. https://doi.org/10.1086/519795. [PubMed]

64. International Parkinson's Disease Genomics Consortium (IPDGC), and Wellcome Trust Case Control Consortium 2 (WTCCC2). A two-stage meta-analysis identifies several new loci for Parkinson's disease. PLoS Genet. 2011; 7:e1002142. https://doi.org/10.1371/journal.pgen.1002142. [PubMed]

65. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. 2010; 42:348–354. https://doi.org/10.1038/ng.548. [PubMed]

66. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA, and 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. https://doi.org/10.1038/nature11632. [PubMed]

67. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5:e1000529. https://doi.org/10.1371/journal.pgen.1000529. [PubMed]