



Published in final edited form as:

*Int J Cancer*. 2017 August 15; 141(4): 731–743. doi:10.1002/ijc.30787.

## Associations between RNA splicing regulatory variants of stemness-related genes and racial disparities in susceptibility to prostate cancer

Yanru Wang<sup>1,2</sup>, Jennifer A. Freedman<sup>1,2</sup>, Hongliang Liu<sup>1,2</sup>, Patricia G. Moorman<sup>1,3</sup>, Terry Hyslop<sup>1,4</sup>, Daniel J. George<sup>1,2</sup>, Norman H. Lee<sup>6</sup>, Steven R. Patierno<sup>1,2,3,5</sup>, and Qingyi Wei<sup>1,2,\*</sup>

<sup>1</sup>Duke Cancer Institute, Duke University Medical Center, Durham, NC 27710, USA

<sup>2</sup>Department of Medicine, Division of Medical Oncology, Duke University Medical Center, Durham, NC 27710, USA

<sup>3</sup>Department of Community and Family Medicine, Duke University Medical Center, Durham, NC 27710, USA

<sup>4</sup>Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC 27710, USA

<sup>5</sup>Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, NC 27710, USA

<sup>6</sup>Department of Pharmacology and Physiology, The George Washington University School of Medicine and Health Sciences, Washington, DC 20037, USA

### Abstract

Evidence suggests that cells with a stemness phenotype play a pivotal role in oncogenesis, and prostate cells exhibiting this phenotype have been identified. We used two genome-wide association study (GWAS) datasets of African descendants, from the Multiethnic/Minority Cohort Study of Diet and Cancer (MEC) and the Ghana Prostate Study, as well as two GWAS datasets of non-Hispanic whites, from the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial and the Breast and Prostate Cancer Cohort Consortium (BPC3), to analyze the associations between genetic variants of stemness-related genes and racial disparities in susceptibility to prostate cancer. We evaluated associations of single nucleotide polymorphisms (SNPs) in 25 stemness-related genes with prostate cancer risk in 1,609 cases and 2,550 controls of non-Hispanic whites (4,934 SNPs) and 1,144 cases and 1,116 controls of African descendants (5,448 SNPs) with correction by false discovery rate 0.2. We identified 32 SNPs in five genes (*TP63*, *ALDH1A1*, *WNT1*, *MET* and *EGFR*) that were significantly associated with prostate cancer risk, of which six SNPs in three genes (*TP63*, *ALDH1A1* and *WNT1*) and eight *EGFR* SNPs showed heterogeneity in susceptibility between these two racial groups. In addition, 13 SNPs in *MET* and one in *ALDH1A1* were found only in African descendants. The *in silico* bioinformatics analyses

Correspondence to: Qingyi Wei, M.D., Ph.D., Duke Cancer Institute, Duke University Medical Center and Department of Medicine, Duke School of Medicine, 905 S LaSalle Street, Durham, NC 27710, USA, Tel.: (919) 660-0562, qingyi.wei@duke.edu. YW and JAF contributed equally to this work. SRP and QW contributed equally as co-principal investigators.

**Conflict of interest:** The authors state no conflict of interest.

revealed that *EGFR* rs2072454 and SNPs in linkage with the identified SNPs in *MET* and *ALDH1A1* ( $r^2 > 0.6$ ) were predicted to regulate RNA splicing. These variants may serve as novel biomarkers for racial disparities in prostate cancer risk.

### Keywords

cancer disparities among racial groups; single nucleotide polymorphism; stemness; prostate cancer susceptibility; RNA splicing

---

### Introduction

In the United States, prostate cancer has surpassed lung cancer as the most common cancer and is the second leading cause of cancer-related deaths in men<sup>1</sup>. Racial and geographic disparities in prostate cancer have been observed in large population studies<sup>1-3</sup>. Notably, African Americans (AAs) have the highest rate of prostate cancer, and prostate cancer in patients of African descent often exhibits a more aggressive phenotype<sup>3</sup>. Specifically, according to the Surveillance, Epidemiology, and End Results Program (SEER) dataset, the age-adjusted rate of prostate cancer was 203.5 per 100,000 among AA men versus 121.9 per 100,000 among white men for the period between 2009 and 2013<sup>3</sup>. Previous studies indicated that racial disparities in both incidence and mortality persist even after controlling for factors associated with social determinants of health, including access to care<sup>4-6</sup>.

Though the mechanisms underlying prostate cancer disparities among racial groups remain largely unknown, high-throughput DNA sequencing studies have begun to elucidate a genomic landscape of cancer traits in recent years<sup>7,8</sup>. Racial disparities in cancer incidence and mortality are likely attributed to the diversity in the genome<sup>9,10</sup>. In recent years, genome-wide association studies (GWASs) have identified some susceptibility loci that provide partial evidence for a role for genetic factors in cancer risk in the general population, and such factors may pave the way toward precision approaches for early detection of cancer<sup>11,12</sup>. Among these studies, few have been undertaken in populations of African descendants, and the results from most validation studies of the identified susceptibility loci have not been consistent across different races<sup>13-16</sup>. Therefore, it is important to continue investigating factors underlying disparities in genetic susceptibility to prostate cancer in different racial populations to unravel further the diverse etiology of the disease. To date, most of the GWAS-identified loci are located in intronic and intergenic regions. In the post-GWAS era, re-analyzing GWAS datasets by using candidate gene or pathway-based gene-set approaches will help identify susceptibility loci that are likely to have biologically relevant functions.

For many years, clinical screening tests for prostate cancer have provided approaches to detect the disease at an early stage<sup>17,18</sup>. Despite recent challenges and controversies, the PSA blood test is still the most widely used method for early detection of prostate cancer<sup>19</sup>. PSA is expressed and secreted by the luminal cells in the prostate. In addition to being routinely used for prostate cancer detection, PSA also serves as a differentiation marker for and an indicator of recurrence following treatments<sup>20</sup>. However, prostate cancer is heterogeneous, exhibiting diverse morphological and histopathological phenotypes<sup>10,19</sup>.

Subpopulations of prostate cancer cells with a lower or negative PSA level have been identified in high-grade, recurrent or metastatic tumors<sup>21</sup>. Such subpopulations of cells related to castration-resistant prostate cancer exhibit a stemness phenotype, with preferential expression of stem cell markers, such as CD44, integrin  $\alpha_2$  and ALDH1A1<sup>21</sup>. Cells having a stemness phenotype are thought to promote oncogenesis, characterized by self-renewal and generation of heterogeneous cell populations<sup>22</sup>. For example, CD133<sup>+</sup>/ $\alpha_2\beta_1^{\text{hi}}$  cells isolated from prostate tumors have been shown to have self-renewal ability with high differentiation similar to that of embryonic stem cells<sup>23</sup>.

Given the evidence that cells exhibiting a stemness phenotype contribute to oncogenesis, we hypothesize that genetic variants in stemness-related genes are associated with prostate cancer risk. In the present study, we tested our hypothesis using four previously published GWAS datasets, including African descendants and non-Hispanic whites, and identified mechanisms underlying racial disparities in susceptibility to prostate cancer.

## Materials and methods

### Study populations

The present study included a total of 2,753 prostate cancer patients and 3,666 controls representing two racial groups (i.e., men of African descent and European descent/non-Hispanic whites) from four studies (Supporting Information Table 1 & 2; See details of study populations in Supporting Information Methods). Specifically, there were 1,144 cases and 1,116 controls from populations of African descendants, of which 670 cases and 658 controls were from African Americans in the Multiethnic Cohort Study of Diet and Cancer (MEC) and 474 cases and 458 controls were from Africans in the Ghana Prostate study. For non-Hispanic whites, the present study included 1,609 cases and 2,550 controls, with 1,150 cases and 1,101 controls from The Prostate, Lung, Colorectal and Ovary (PLCO) screening trial and 459 cases and 1,449 controls from the Breast and Prostate Cancer Cohort Consortium (BPC3). Each of the four studies was reviewed and approved by the corresponding Institutional Review Board and Duke Institutional Review Board.

### Gene and SNP selection

A list of 25 stemness-related genes (Supporting Information Table 3) in prostate cancer was collected according to the online dataset GeneCards (<http://www.genecards.org/>) using the search term “prostate cancer stem cell”. Genotyped SNPs within these genes and their  $\pm$  2-kb flanking regions were selected for association analysis. Genotype imputation was performed with IMPUTE2 for each study according to multi-population reference panels from the 1000 Genomes Project Phase 3<sup>24, 25</sup>. Imputed SNPs with info value  $\geq$  0.8 were qualified for further analysis. As a result, there were 8,609, 8,600, 13,765 and 13,845 SNPs in the aforementioned 25 genes from populations of the PLCO, BPC3, Ghana and MEC studies, respectively. After being filtered with the criteria of minor allele frequency (MAF)  $\geq$  0.05, genotyping call rate  $\geq$  95% and Hardy-Weinberg equilibrium  $\geq 10^{-5}$  in each study, there were 5,239, 5,345, 6,267 and 6,549 common SNPs from PLCO, BPC3, Ghana and MEC, respectively. In the final analysis, 4,934 common SNPs were included for both PLCO and BPC3 for non-Hispanic whites and 5,448 common SNPs were included for both Ghana

and MEC for African descendants. Pairwise linkage disequilibrium (LD) of the SNPs was further estimated by using the data from the 1000 Genomes Project Phase 3 of the matched population (<http://www.1000genomes.org/>).

### Functional annotation and eQTL analysis

We predicted functional annotations of the SNPs using three online tools: SNPinfo (<https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm>), RegulomeDB (<http://www.regulomedb.org/>) and HaploReg (<http://archive.broadinstitute.org/mammals/haploreg/haploreg.php>). In the expression quantitative trait loci (eQTL) analysis, we used additional data on genotyping and mRNA expression levels measured by the Illumina Human-6 v2 Expression BeadChip for the corresponding genes in lymphoblastoid cell lines obtained from 107 Northern Europeans from Utah and 326 Africans in the HapMap 3 project (<https://hapmap.ncbi.nlm.nih.gov/>)<sup>26</sup>.

### Statistical analysis

For each study, we calculated principal components (PCs) using the Genome-wide Complex Trait Analysis (GCTA) on the LD-pruned subset of the whole-genome typed dataset, as recommended<sup>27, 28</sup>. The first 20 PCs were analyzed for their associations with prostate cancer risk by the univariate logistic regression analysis. Those PCs with significant associations were included as covariates in further analyses of associations between SNPs and prostate cancer risk. For each SNP, we estimated odds ratios (ORs) and 95% confidence intervals (95% CIs) by unconditional logistic regression of case/control groups with adjustment for age and PCs. Cochran's Q statistics and  $I^2$  was used to assess the inter-study and inter-race heterogeneity of the SNPs. We defined an SNP with a Q-test  $P < 0.100$  or  $I^2 > 50.0\%$  as heterogeneous. We used a meta-analysis first to generate race-specific results of overall risk associated with the SNPs in fixed-effects models, if no heterogeneity between two studies, or random-effects models, when heterogeneity existed. We then generated the heterogeneity statistic to test the differences between non-Hispanic whites and African descendants by using Cochran's Q statistics and  $I^2$ . The  $I^2$  index of 25%, 50% and 75% was considered low, medium and high heterogeneity, respectively<sup>29</sup>. The false discovery rate (FDR) approach was used to correct for multiple comparisons to reduce the probability of false-positive findings<sup>30</sup>. For each identified SNP, a meta-regression analysis was conducted among results from the four studies to further analyze the heterogeneity between the two racial groups by using "metafor" package of R software. Multiplicative interactions between the SNPs and the racial group were evaluated using multivariate logistic regression analysis adjusted for age and the racial group in a pooled dataset from the four studies. In the eQTL analysis, we calculated the correlations between SNPs and corresponding mRNA expression levels by using a general linear regression model. Differences in mRNA expression levels between the two populations were examined by a two-sample *t*-test. Statistical analyses were performed using R (version 3.3.1), SAS (version 9.1.3; SAS Institute, Cary, NC, USA) and PLINK (version 1.07), unless otherwise specified.

## Results

### Basic characteristics of the study populations

The overall analysis included 2,753 prostate cancer cases and 3,666 controls of two racial groups from four studies (Supplemental Tables 1–2). The age distribution was statistically different between cases and controls ( $P < 0.001$ ), with the control group being older than the case group (> 70 years: 58.5% versus 43.8%). Additional details regarding the racial groups from the four studies are presented in Supporting Information Table 2. To control for the population stratification, the first 20 PCs in each study were included in the models for analyses of associations with prostate cancer risk (Supporting Information Table 4). Therefore, age and PCs were adjusted for their possible confounding effects in the following multivariate logistic regression analysis.

### Association analysis of SNPs and prostate cancer risk in populations of African descendants

The workflow of the current study is presented in Fig. 1. Considering that the allele frequency of each SNP varies between populations of different races, we separated our analyses into two parts. In the first part, we analyzed the associations between common SNPs (MAF  $\geq 0.05$ ) and prostate cancer risk in populations of African descendants (Fig. 1a). The imputation resulted in 6,267 and 6,549 common SNPs for the Ghana study and the MEC study, respectively (Fig. 2a–b); we then performed a meta-analysis using the 5,448 overlapped SNPs present in both studies (Fig. 2c) and found that 300 common SNPs were associated with prostate cancer risk with a  $P$ -value  $\leq 0.05$ . After multiple test correction using the FDR  $\leq 0.2$  (Table 1), 24 SNPs in four genes (i.e., eight in *TP63*, 13 in *MET*, two in *ALDH1A1* and one in *WNT1*) remained significant. The minor alleles of SNPs in *ALDH1A1* were associated with a decreased risk of prostate cancer, whereas the other 22 SNPs in three genes were all associated with an increased risk of prostate cancer.

We predicted potential functions of those 24 SNPs by using three online tools, and results are summarized in Supporting Information Table 5. Most of these SNPs are located in intronic regions of the corresponding genes, except for rs149188492 (located in the 3' untranslated region of *MET*) and rs855723 (located in the 5' upstream region of *WNT1*). All 24 SNPs were predicted to play a role in transcriptional regulation, either located within transcription factor binding sites and DNase I hypersensitive sites or promoter and enhancer histone marks. Among the 13 SNPs of *MET*, three (rs139335187, rs201395418 and rs567033632) are in high LD ( $r^2 > 0.8$ ) with rs13223756 and four (rs377420134, rs115240747, rs114707545 and rs115293079) are in moderate LD ( $r^2 = 0.63–0.72$ ) with rs13223756 based on the LD data from populations of African descendants in the 1000 Genomes Project Phase 3. The rs13223756 SNP is located in an exonic region of *MET*, which is predicted to be involved in RNA splicing regulation by SNPinfo. In *ALDH1A1*, the risk-associated SNPs rs722921 and rs13959 are in moderate LD ( $r^2 = 0.69$ ) based on populations of African descendants in the 1000 Genomes Project Phase 3. *ALDH1A1* rs13959 is located in an exonic region, which is predicted to affect RNA splicing by SNPinfo as well.

### Association analysis of SNPs and prostate cancer risk in non-Hispanic whites

Similar to the previous analyses for populations of African descendants, we analyzed the associations between prostate cancer risk and common SNPs in two GWAS datasets of non-Hispanic whites, including 5,239 SNPs in the PLCO study and 5,345 SNPs in the BPC3 study (Fig. 2d–e). We performed a meta-analysis by using the overlapped 4,934 common SNPs in both studies (Fig. 2f) and identified 233 SNPs that were associated with prostate cancer risk with a  $P$ -value  $\leq 0.05$ , of which eight SNPs in *EGFR* remained significantly associated with an increased risk of prostate cancer after multiple test correction by FDR 0.2. One of these eight SNPs, rs2072454, is in an exonic region of *EGFR*, and the other seven SNPs of *EGFR* are located in the intronic regions. Functional prediction, using the three previously mentioned online tools, indicated that rs2072454 may play a role in RNA splicing and that the other intronic variants may play roles in transcriptional regulation (Supporting Information Table 5).

### Heterogeneity of the SNPs between two racial groups of African descendants and non-Hispanic whites

We further evaluated potential differences in the SNP-associated of prostate cancer between the two racial groups of African descendants and non-Hispanic whites. Of the top 24 SNPs initially identified in African descendants, only 10 were found in non-Hispanic whites from both PLCO and BPC3, and the other 14 SNPs (13 in *MET* and one in *ALDH1A1*) were significantly associated with cancer risk only in African descendants (Table 1). According to European data from the 1000 Genomes Project and the two non-Hispanic white population datasets used in the present study, we found only one allele for these 14 SNPs in European descendants, which appeared to be African-specific variants. Of the other 10 SNPs found in both racial groups, six SNPs in three genes (four in *TP63*, one in *ALDH1A1* and one in *WNT1*) showed moderate to high heterogeneity between African descendants and non-Hispanic whites ( $I^2=51.2-81.7$  between two racial groups, Table 2 and Supporting Information Table 6). In contrast, the top eight *EGFR* SNPs initially identified in non-Hispanic whites were but not in African descendants; however, these eight SNPs showed high heterogeneity between populations of African descendants and non-Hispanic whites (all Q-test  $P < 0.100$  and  $I^2 > 75.0$  between two racial groups, Table 2 and Supporting Information Table 6).

Therefore, 14 SNPs in four genes (*TP63*, *ALDH1A1*, *WNT1* and *EGFR*) showed different effects on prostate cancer risk between the two racial groups. To further evaluate these differences, we applied the meta-regression analysis of the four studies as independent datasets and included the racial group (i.e., African descendant and non-Hispanic white) as a moderator (Table 2). The results of the heterogeneity of the two racial groups in meta-regression analysis were similar to the Q-test  $P$  value for testing the heterogeneity between the two racial groups (Table 2). In addition, we analyzed multiplicative interactions between the 14 SNPs and the racial group in a pooled dataset, including 2,753 cases and 3,666 controls from the four studies. The 14 SNPs showed significant interaction effects with race (all  $P < 0.05$ , Table 2).



**LD analysis**—We further analyzed the LD map for the top SNPs that were identified to be associated with prostate cancer risk in the two racial groups (i.e. 24 SNPs in four genes of African descendants and eight SNPs in *EGFR* of non-Hispanic whites). Based on the 1000 Genome Project Phase 3 African dataset, the eight SNPs in *TP63* all located in the intronic regions are in moderate to high LD (all  $r^2 > 0.7$ ) (Fig. 3a). In the *MET* gene, the 12 SNPs also share moderate to high LD ( $r^2=0.6-1.00$ ), except for the intronic SNP rs116458171 of African descendants (Fig. 3b). The two SNPs in *ALDH1A1* of African descendants were in low LD ( $r^2=0.44$ ) (Fig. 3c). The eight SNPs in *EGFR* were all in high LD according to the 1000 Genomes Project Phase 3 European dataset (all  $r^2 > 0.8$ ) (Fig. 3d). Therefore, we chose the SNPs, with predicted functions and exhibiting heterogeneity between racial groups, in each gene based on a threshold of  $r^2 = 0.6$  in the LD analysis. As a result, seven SNPs in five genes (*TP63* rs7616437, *MET* rs114707545, *MET* rs116458171, *ALDH1A1* rs8187942, *ALDH1A1* rs72291, *WNT1* rs855723 and *EGFR* rs2072454) were selected for further analyses.

### Stratified analysis for tumor aggressiveness

Prostate cancer with a Gleason score  $> 7$  or stage III was defined as aggressive, and other cases with a Gleason score  $\leq 7$  and stage  $< III$  were defined as non-aggressive (Supporting Information Table 2). We then conducted a subgroup analysis by tumor aggressiveness in the available datasets, including MEC for AAs and PLCO for non-Hispanic whites (Supporting Information Table 7). In six selected SNPs in four genes of African descendants, SNPs in *TP63*, *MET* and *WNT1* showed risk effects on both non-aggressive and aggressive disease, and two SNPs in *ALDH1A1* showed protective effects on both disease groups in the MEC study. In the PLCO study of non-Hispanic whites, rs2072454 was associated with an increased risk of prostate cancer in both disease groups. However, no heterogeneity was identified between subgroups of aggressiveness.

**eQTL analysis**—To substantiate biological mechanisms underlying the findings of risk associations, we conducted additional eQTL analysis to explore the correlation between identified SNPs and the mRNA expression levels of the corresponding genes using the HapMap 3 project (Table 1). Among these, the results of the selected SNPs in five genes (i.e., *TP63*, *MET*, *ALDH1A1*, *WNT1* and *EGFR*) are shown in Fig. 4. Four of the five genes were differentially expressed between African and European populations. Compared to European populations, African populations had a significantly lower mRNA expression level of *TP63* and *MET* (Fig. 4a and c), but a significantly higher expression level of *WNT1* and *EGFR* (Fig. 4d and e). Among the 24 top SNPs of African descendants, the risk A allele of rs116458171 was significantly associated with lower levels of *MET* mRNA expression in Africans ( $P = 0.024$ , Beta =  $-0.107$ , Table 1 and Fig. 4c). In the eight top SNPs of non-Hispanic whites, we observed significant associations between these eight SNPs (all in high LD) and *EGFR* mRNA expression in Europeans (all  $P = 0.031$  and Beta =  $0.036$ , Table 1). We visualized the eQTL results for the selected SNP, rs2072454 in *EGFR* (Fig. 4e). Europeans with the rs2072454 risk C allele had significantly higher levels of expression of *EGFR* mRNA ( $P = 0.031$ , Fig. 4e). In Africans, although there was a slight trend toward correlation of the rs2072454 C allele with a lower level of expression of *EGFR* mRNA, this correlation was not statistically significant ( $P = 0.361$ , Beta =  $-0.005$ , Table 1 and Fig. 4e).

## Discussion

To determine whether genetic variants in prostate cancer stemness-related genes contribute to prostate cancer susceptibility among racial groups, we used the genotyping data from published GWASs from the Ghana and MEC studies of African descendants and the PLCO and BPC3 studies of non-Hispanic whites. We found a greater than 1.50-fold increased risk of prostate cancer associated with 13 SNPs in *MET* and a 40% decreased risk of prostate cancer associated with rs8187942 in *ALDH1A1* in African descendants. In contrast, SNPs in *EGFR* were found to be associated with an increased risk of prostate cancer in non-Hispanic whites. Moreover, we found that several SNPs in *TP63*, *ALDH1A1*, *WNT* and *EGFR* were associated with risk of prostate cancer in populations of African descent, but not in non-Hispanic whites or vice versa. Regarding functional prediction of the SNPs associated with prostate cancer risk, rs2072454 in *EGFR* was predicted to be involved in RNA splicing regulation and its risk allele was also associated with higher levels of expression of *EGFR* mRNA. In addition, seven SNPs in *MET* and one in *ALDH1A1* were in moderate to high LD ( $r^2 > 0.6$ ) with other variants in the same genes and were also predicted to regulate RNA splicing. These functional SNPs in five genes found to be associated with prostate cancer risk and with racial differences in prostate cancer risk have not been previously reported.

In the present study, we found eight SNPs in *TP63* to be significantly associated with prostate cancer risk in men of African descent, of which four showed differences in prostate cancer risk between the two racial groups. The *TP63* gene is a homolog of *TP53*, which is a member of the tumor suppressor gene p53 family. Unlike the role of *TP53* as solely a tumor suppressor, *TP63* encodes two isoforms generated from alternative RNA splicing, including TAp63, which contains an N-terminal transactivation (TA) domain and functions as a tumor suppressor, and DNp63, which lacks the TA domain and functions as a proto-oncogene. DNp63 plays a role in maintaining the proliferative potential of epidermal progenitor cells, including bladder and prostate epithelial cells<sup>31</sup>. It has been reported that overexpression of TP63 can change the ratio of TAp63/DNp63, resulting in relatively higher expression of DNp63 in tumor tissue versus normal tissue<sup>32</sup>. Several SNPs in 3q28 within the *TP63* gene region have been previously reported to be associated with risk of lung cancer and bladder cancer<sup>33, 34</sup>, but have not been previously reported to be associated with risk of prostate cancer.

We also identified SNPs in *WNT1* and *ALDH1A1* that were differentially associated with prostate cancer risk between the two racial groups. *WNT1* is part of the WNT signaling pathway that plays crucial roles in cell proliferation, cell migration, cell fate and stem cell renewal<sup>35</sup>. SNP rs855723 was previously reported to be correlated with expression of *WNT1* mRNA in lymphoblastoid cells from Europeans<sup>36</sup>. In the present study, we demonstrated that the rs855723G allele was significantly associated with lower *WNT1* mRNA expression levels in the HapMap 3 project European population. This variant was predicted to affect transcription factor binding, residing within the CCCTC-Binding Factor (CTCF) binding site based on chromatin immunoprecipitation sequencing (CHIP-seq) data from HaploReg. Distinct epigenetic patterns, including DNA and histone methylation of the CTCF binding sites have been documented in benign prostate hyperplasia versus prostate cancer<sup>37</sup>.



*ALDH1A1* is part of the aldehyde dehydrogenase family, which is involved in intracellular retinoic acid production<sup>38</sup>. Aldehyde dehydrogenase serves as a binding protein for metabolic molecules and potentially functions as an antioxidant, which links it to a role in maintaining stemness<sup>38</sup>. In particular, one study has shown that *ALDH1A1*(+) prostate cancer cells exhibit high clonogenic and tumorigenic capacities<sup>39</sup>. Isoforms of *ALDH1A1*, generated by alternative RNA splicing, are associated with different expression levels of the gene in endometrial adenocarcinoma patients<sup>39</sup>. In the present study, one SNP in the *ALDH1A1* exon region, which is predicted to affect alternative RNA splicing, is in LD with the identified SNP associated with prostate cancer risk in African population.

*MET*, also known as the hepatocyte growth factor receptor, is a proto-oncogenic receptor tyrosine kinase that transduces signals from the extracellular matrix to the cytoplasm<sup>40</sup>. A recent study reported that a splice site alteration involving exon 14 in *MET* was found in multiple cancers<sup>41</sup>. The altered *MET* is constitutively active and plays a role in oncogenic transformation<sup>41</sup>. In the present study, the identified prostate cancer risk-associated SNPs in *MET* were in LD with a SNP predicted to regulate RNA splicing. The SNPs identified in *MET* here are novel loci that contribute to prostate cancer risk and are enriched in populations of African descent. In addition, the risk allele of rs116458171 was associated with lower levels of expression of *MET* mRNA, is located in an intronic region and is predicted to affect transcriptional regulation. Taken together, these results indicate that further studies are warranted to elucidate the biological mechanisms underlying the observed associations.

rs2072454, located in the *EGFR* coding region, is predicted to have an effect on the regulation of RNA splicing by SNPinfo. *EGFR*, also known as *ERBB1*, encodes a well-known transmembrane glycoprotein that is a member of the receptor tyrosine kinase superfamily<sup>42</sup>. Particular growth factors selectively bind to *EGFR* and trigger intracellular signaling, which ultimately results in cell proliferation<sup>42</sup>. Increased expression of and somatic mutations in *EGFR* and related growth factors are frequently observed in cancer and promote proliferation of tumor cells<sup>42</sup>. In non-small cell lung cancer and prostate cancer, somatic mutations in exons 20 and 21 of *EGFR* have been observed in such tumors with a highly proliferative and invasive phenotype<sup>43</sup>. It has been reported that *EGFR* overexpression in prostate cancer tissues is significantly more common in men of African ancestry than in white men<sup>44</sup>.

Genome-wide analyses of transcriptomes have revealed extensive alternative RNA splicing, which generates enormous biological diversity<sup>45</sup>. Specifically, next-generation sequencing data suggest that approximately 95% of genes in the human genome undergo alternative RNA splicing<sup>45</sup>. During this process, the pre-mRNA generates multiple messenger RNAs that are translated into distinct proteins with divergent biological functions<sup>45</sup>. The connection between alternative RNA splicing and cancer risk and progression has become increasingly appreciated because of a considerable number of recent studies indicating that alternative and aberrant isoforms can dysregulate signaling pathways, thus contributing to oncogenesis<sup>45</sup>. The *cis*-acting RNA splicing elements, a part of the regulatory system controlling RNA splicing, consist of exonic and intronic splicing enhancers (ESEs and ISEs) and silencers (ESSs and ISSs). In the present study, *EGFR* rs2072454, *MET* rs13223756 and

*ALDH1A1* rs13959 (the latter two are in LD with the identified SNPs) were predicted to be located within ESE regions by SNPinfo. Specifically, we found that the rs2072454C allele was associated with an increased risk of prostate cancer and higher level of expression of *EGFR* mRNA in Europeans.

Genetic patterns vary across different populations, such as the allele frequencies of SNPs and LD by haplotype structures<sup>46</sup>. Therefore, SNPs, haplotypes and gene-gene interactions could provide broad aspects of genetic diversity in multiple dimensions. In the present study, we have shown that the number of common SNPs in stemness-related genes was greater in African descendants than in non-Hispanic whites. Furthermore, we have found that prostate cancer risk-associated SNPs in *MET* and *ALDH1A1* were more likely to be confined to populations of African descent.

However, SNPs with a MAF less than 0.05 were not all included in the analyses of the present study. Several studies have shown that rare variants may be responsible for substantial portions of inherited prostate cancer susceptibility<sup>47</sup>. In searching evidence for racial disparity, we attempted to assess the interactions between SNPs and two racial groups in the available GWAS datasets. However, interactions between SNPs and four GWAS studies were nested inside the racial groups, and thus it was impossible to completely adjust or detect such an interaction in present models. In the future, large studies including more racial groups, rare variants and environmental factors may provide sufficient statistical power to detect possible interactions among races, rare functional SNPs and environmental factors in prostate cancer risk. Besides, we found that four genes (*TP63*, *MET*, *WNT1* and *EGFR*) with identified prostate cancer risk-associated SNPs may function differently by racial group, because they were differentially expressed in lymphoblastoid cell lines derived from Africans and Europeans in the HapMap 3 project. Subsequent molecular studies to investigate biological differences by racial backgrounds are needed to substantiate the functional impact of these risk-associated SNPs on racial disparities in prostate cancer risk.

In conclusion, we investigated the associations between genetic variants in 25 prostate cancer stemness-related genes and prostate cancer risk in African descendants and non-Hispanic whites, including 2,753 cases and 3,666 controls. We found several SNPs in *TP63*, *MET*, *WNT1* and *ALDH1A1* to be associated with prostate cancer risk in populations of African descent and SNPs in *EGFR* to be associated with increased risk of prostate cancer in non-Hispanic whites. Several SNPs in the aforementioned genes showed differences in prostate cancer susceptibility between the two racial groups and were predicted to play roles in regulation of RNA splicing and transcription. Our findings identify some evidence for RNA splicing-regulatory SNPs as possible novel biomarkers that may provide new insights into the molecular mechanisms underlying racial disparities in prostate cancer risk.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

**Grant sponsor:** Duke Cancer Institute's P30 Cancer Center Support Grant; **Grant number:** NIH CA014236.  
**Grant sponsor:** DoD Prostate Cancer Research Program Health Disparity Research Award; **Grant number:** PC131972.

The authors thank all of the participants of the PLCO Cancer Screening Trial, BPC3 GWAS Study, MEC Prostate Cancer GWAS Study and Ghana Prostate Study. The authors would also like to acknowledge dbGaP repository for providing cancer genotyping datasets. The accession numbers for the datasets used in this manuscript are phs000207.v1.p1 for CGEMS prostate cancer scan - PLCO GWAS, phs000812.v1.p1 for BPC3 GWAS, phs000838.v1.p1 for Ghana Prostate Study and phs000306.v3.p1 for MEC Prostate Cancer GWAS. Funding sources can be found in the Supporting Information Acknowledgments. JAF, PGM, DJG, TH, QW and SRP acknowledge support from the Duke Cancer Institute's P30 Cancer Center Support Grant (Grant ID: NIH CA014236). SRP and JAF acknowledge support from the DoD Prostate Cancer Research Program Health Disparity Research Award PC131972 to SRP PI and JAF Co-I.

## Abbreviations

<b>GWAS</b>	genome-wide association study
<b>SNP</b>	single nucleotide polymorphism
<b>SEER</b>	Surveillance, Epidemiology, and End Results program
<b>MEC</b>	Multiethnic/Minority Cohort Study of Diet and Cancer
<b>PLCO</b>	Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial
<b>BPC3</b>	Breast and Prostate Cancer Cohort Consortium
<b>AA</b>	African American
<b>PC</b>	principal component
<b>LD</b>	linkage disequilibrium
<b>MAF</b>	minor allele frequency
<b>OR</b>	odds ratio
<b>CI</b>	confidence interval

## References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA: a cancer journal for clinicians*. 2016; 66:7–30. [PubMed: 26742998]
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. *International journal of cancer*. 2015; 136:E359–86. [PubMed: 25220842]
3. Howlader, NNA.Krapcho, M.Miller, D.Bishop, K.Altekruse, SF.Kosary, CL.Yu, M.Ruhl, J.Tatalovich, Z.Mariotto, A.Lewis, DR.Chen, HS.Feuer, EJ., Cronin, KA., editors. *SEER Cancer Statistics Review, 1975–2013*. National Cancer Institute; Bethesda, MD: [http://seercancer.gov/csr/1975\\_2013/](http://seercancer.gov/csr/1975_2013/)
4. Cook MB, Rosenberg PS, McCarty FA, Wu M, King J, Ehemann C, Anderson WF. Racial disparities in prostate cancer incidence rates by census division in the United States, 1999–2008. *The Prostate*. 2015; 75:758–63. [PubMed: 25619191]

5. Godley PA, Schenck AP, Amamoo MA, Schoenbach VJ, Peacock S, Manning M, Symons M, Talcott JA. Racial differences in mortality among Medicare recipients after treatment for localized prostate cancer. *Journal of the National Cancer Institute*. 2003; 95:1702–10. [PubMed: 14625261]
6. Powell IJ, Banerjee M, Bianco FJ, Wood DP Jr, Dey J, Lai Z, Heath M, Pontes EJ. The effect of race/ethnicity on prostate cancer treatment outcome is conditional: a review of Wayne State University data. *The Journal of urology*. 2004; 171:1508–12. [PubMed: 15017209]
7. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. Genomes Project C. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491:56–65. [PubMed: 23128226]
8. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74. [PubMed: 22955616]
9. Jorde LB, Wooding SP. Genetic variation, classification and 'race'. *Nature genetics*. 2004; 36:S28–33. [PubMed: 15508000]
10. Tan DS, Mok TS, Rebbeck TR. Cancer Genomics: Diversity and Disparity Across Ethnicity and Geography. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2016; 34:91–101. [PubMed: 26578615]
11. Amin AI, Olama A, Benlloch S, Antoniou AC, Giles GG, Severi G, Neal DE, Hamdy FC, Donovan JL, Muir K, Schleutker J, Henderson BE, Haiman CA, et al. Risk Analysis of Prostate Cancer in PRACTICAL, a Multinational Consortium, Using 25 Known Prostate Cancer Susceptibility Loci. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2015; 24:1121–9.
12. Choudhury AD, Eeles R, Freedland SJ, Isaacs WB, Pomerantz MM, Schalken JA, Tammela TL, Visakorpi T. The role of genetic markers in the management of prostate cancer. *European urology*. 2012; 62:577–87. [PubMed: 22695242]
13. Haiman CA, Chen GK, Blot WJ, Strom SS, Berndt SI, Kittles RA, Rybicki BA, Isaacs WB, Ingles SA, Stanford JL, Diver WR, Witte JS, et al. Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. *Nature genetics*. 2011; 43:570–3. [PubMed: 21602798]
14. Chang BL, Spangler E, Gallagher S, Haiman CA, Henderson B, Isaacs W, Benford ML, Kidd LR, Cooney K, Strom S, Ingles SA, Stern MC, et al. Validation of genome-wide prostate cancer associations in men of African descent. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2011; 20:23–32.
15. Cook MB, Wang Z, Yeboah ED, Tettey Y, Biritwum RB, Adjei AA, Tay E, Truelove A, Niwa S, Chung CC, Chokkalingam AP, Chu LW, et al. A genome-wide association study of prostate cancer in West African men. *Human genetics*. 2014; 133:509–21. [PubMed: 24185611]
16. Haiman CA, Chen GK, Blot WJ, Strom SS, Berndt SI, Kittles RA, Rybicki BA, Isaacs WB, Ingles SA, Stanford JL, Diver WR, Witte JS, et al. Characterizing genetic risk at known prostate cancer susceptibility loci in African Americans. *PLoS genetics*. 2011; 7:e1001387. [PubMed: 21637779]
17. Andriole GL, Crawford ED, Grubb RL 3rd, Buys SS, Chia D, Church TR, Fouad MN, Isaacs C, Kvale PA, Reding DJ, Weissfeld JL, Yokochi LA, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *Journal of the National Cancer Institute*. 2012; 104:125–32. [PubMed: 22228146]
18. Kolonel LN, Altshuler D, Henderson BE. The multiethnic cohort study: exploring genes, lifestyle and cancer risk. *Nature reviews Cancer*. 2004; 4:519–27. [PubMed: 15229477]
19. National Comprehensive Cancer Network. Prostate cancer ( Version 2.2016).
20. Lilja H, Ulmert D, Vickers AJ. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. *Nature reviews Cancer*. 2008; 8:268–78. [PubMed: 18337732]
21. Qin J, Liu X, Laffin B, Chen X, Choy G, Jeter CR, Calhoun-Davis T, Li H, Palapattu GS, Pang S, Lin K, Huang J, et al. The PSA(-/lo) prostate cancer cell population harbors self-renewing long-term tumor-propagating cells that resist castration. *Cell stem cell*. 2012; 10:556–69. [PubMed: 22560078]

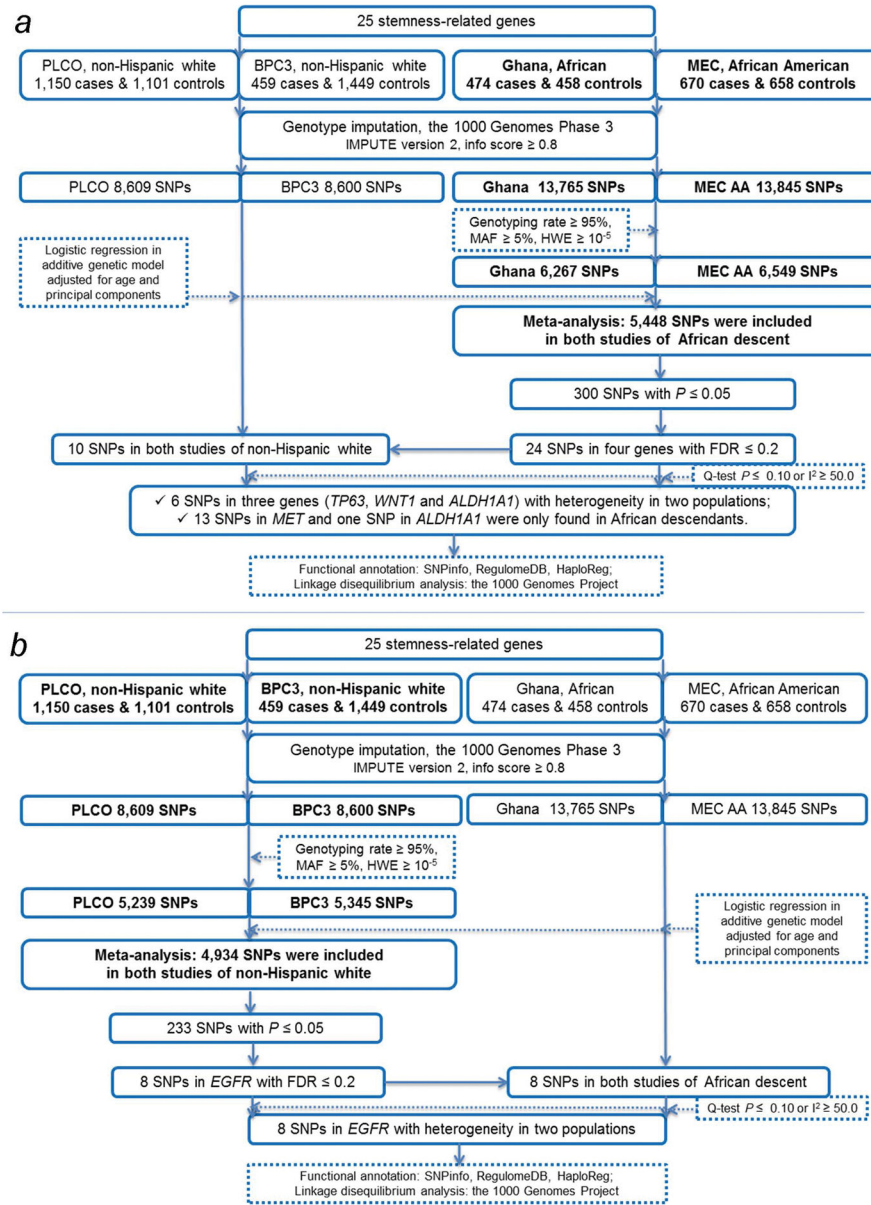
22. Medema JP. Cancer stem cells: the challenges ahead. *Nature cell biology*. 2013; 15:338–44. [PubMed: 23548926]
23. Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ. Prospective identification of tumorigenic prostate cancer stem cells. *Cancer research*. 2005; 65:10946–51. [PubMed: 16322242]
24. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics*. 2009; 5:e1000529. [PubMed: 19543373]
25. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3*. 2011; 1:457–70. [PubMed: 22384356]
26. Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, Sekowska M, Smith GD, Evans D, Gutierrez-Arcelus M, Price A, Raj T, et al. Patterns of cis regulatory variation in diverse human populations. *PLoS genetics*. 2012; 8:e1002639. [PubMed: 22532805]
27. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics*. 2011; 88:76–82. [PubMed: 21167468]
28. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annual review of genomics and human genetics*. 2009; 10:387–406.
29. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002; 21:1539–58. [PubMed: 12111919]
30. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behavioural brain research*. 2001; 125:279–84. [PubMed: 11682119]
31. Pignon JC, Grisanzio C, Geng Y, Song J, Shivdasani RA, Signoretti S. p63-expressing cells are the stem cells of developing prostate, bladder, and colorectal epithelia. *Proc Natl Acad Sci U S A*. 2013; 110:8105–10. [PubMed: 23620512]
32. Candi E, Dinsdale D, Rufini A, Salomoni P, Knight RA, Mueller M, Krammer PH, Melino G. TAp63 and DeltaNp63 in cancer and epidermal development. *Cell Cycle*. 2007; 6:274–85. [PubMed: 17264681]
33. Hu Z, Wu C, Shi Y, Guo H, Zhao X, Yin Z, Yang L, Dai J, Hu L, Tan W, Li Z, Deng Q, et al. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nature genetics*. 2011; 43:792–6. [PubMed: 21725308]
34. Kiemeny LA, Thorlacius S, Sulem P, Geller F, Aben KK, Stacey SN, Gudmundsson J, Jakobsdottir M, Bergthorsson JT, Sigurdsson A, Blondal T, Witjes JA, et al. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nature genetics*. 2008; 40:1307–12. [PubMed: 18794855]
35. Anastas JN, Moon RT. WNT signalling pathways as therapeutic targets in cancer. *Nature reviews Cancer*. 2013; 13:11–26. [PubMed: 23258168]
36. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG, Barann M, Wieland T, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013; 501:506–11. [PubMed: 24037378]
37. Paradowska A, Fenic I, Konrad L, Sturm K, Wagenlehner F, Weidner W, Steger K. Aberrant epigenetic modifications in the CTCF binding domain of the IGF2/H19 gene in prostate cancer compared with benign prostate hyperplasia. *International journal of oncology*. 2009; 35:87–96. [PubMed: 19513555]
38. Marcato P, Dean CA, Giacomantonio CA, Lee PW. Aldehyde dehydrogenase: its role as a cancer stem cell marker comes down to the specific isoform. *Cell Cycle*. 2011; 10:1378–84. [PubMed: 21552008]
39. Mamat S, Ikeda J, Tian T, Wang Y, Luo W, Aozasa K, Morii E. Transcriptional Regulation of Aldehyde Dehydrogenase 1A1 Gene by Alternative Spliced Forms of Nuclear Factor Y in Tumorigenic Population of Endometrial Adenocarcinoma. *Genes & cancer*. 2011; 2:979–84. [PubMed: 22701763]
40. Jeffers M, Schmidt L, Nakaigawa N, Webb CP, Weirich G, Kishida T, Zbar B, Vande Woude GF. Activating mutations for the met tyrosine kinase receptor in human cancer. *Proc Natl Acad Sci U S A*. 1997; 94:11445–50. [PubMed: 9326629]

41. Frampton GM, Ali SM, Rosenzweig M, Chmielecki J, Lu X, Bauer TM, Akimov M, Bufill JA, Lee C, Jentz D, Hoover R, Ou SH, et al. Activation of MET via diverse exon 14 splicing alterations occurs in multiple tumor types and confers clinical sensitivity to MET inhibitors. *Cancer Discov.* 2015; 5:850–9. [PubMed: 25971938]
42. Ciardiello F, Tortora G. EGFR antagonists in cancer treatment. *The New England journal of medicine.* 2008; 358:1160–74. [PubMed: 18337605]
43. Peraldo-Neia C, Migliardi G, Mello-Grand M, Montemurro F, Segir R, Pignochino Y, Cavalloni G, Torchio B, Mosso L, Chiorino G, Aglietta M. Epidermal Growth Factor Receptor (EGFR) mutation analysis, gene expression profiling and EGFR protein expression in primary prostate cancer. *BMC cancer.* 2011; 11:31. [PubMed: 21266046]
44. Shuch B, Mikhail M, Satagopan J, Lee P, Yee H, Chang C, Cordon-Cardo C, Taneja SS, Osman I. Racial disparity of epidermal growth factor receptor expression in prostate cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2004; 22:4725–9. [PubMed: 15570072]
45. Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, Munoz MJ. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature reviews Molecular cell biology.* 2013; 14:153–65. [PubMed: 23385723]
46. Rosenberg NA, Kang JT. Genetic Diversity and Societally Important Disparities. *Genetics.* 2015; 201:1–12. [PubMed: 26354973]
47. Mancuso N, Rohland N, Rand KA, Tandon A, Allen A, Quinque D, Mallick S, Li H, Stram A, Sheng X, Kote-Jarai Z, Easton DF, et al. The contribution of rare variation to prostate cancer heritability. *Nature genetics.* 2016; 48:30–5. [PubMed: 26569126]

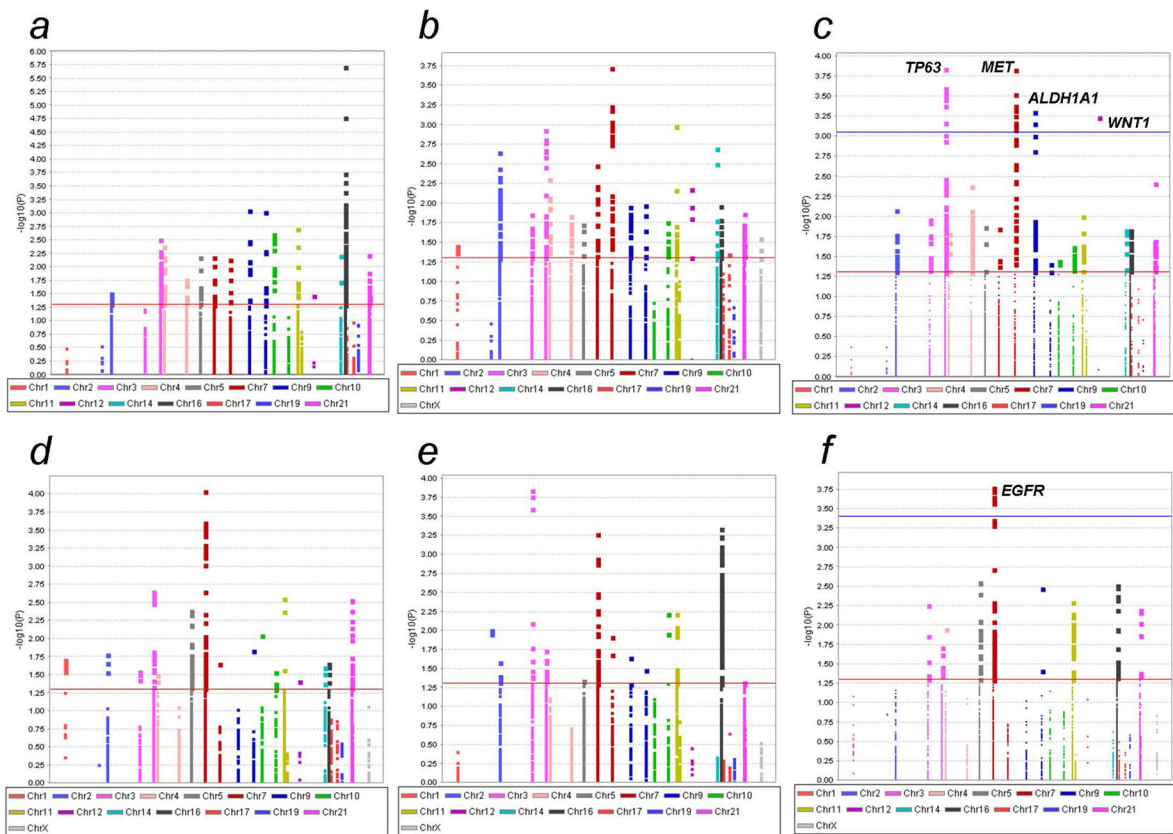


### What's new?

We assessed the role of stemness-related genetic variants in racial differences in prostate cancer risk by re-analyzing published datasets from four genome-wide association studies of African descendants and non-Hispanic whites. We found that SNPs in *TP63*, *ALDH1A1*, *WNT1*, *MET* and *EGFR* were significantly associated with prostate cancer risk. Specifically, SNPs in *TP63*, *ALDH1A1*, *WNT1* and *EGFR* showed differences in prostate cancer risk between these two racial groups. It is likely that this racial disparity in prostate cancer risk may be associated with RNA splicing-related SNPs in *EGFR*, *MET* and *ALDH1A1*.

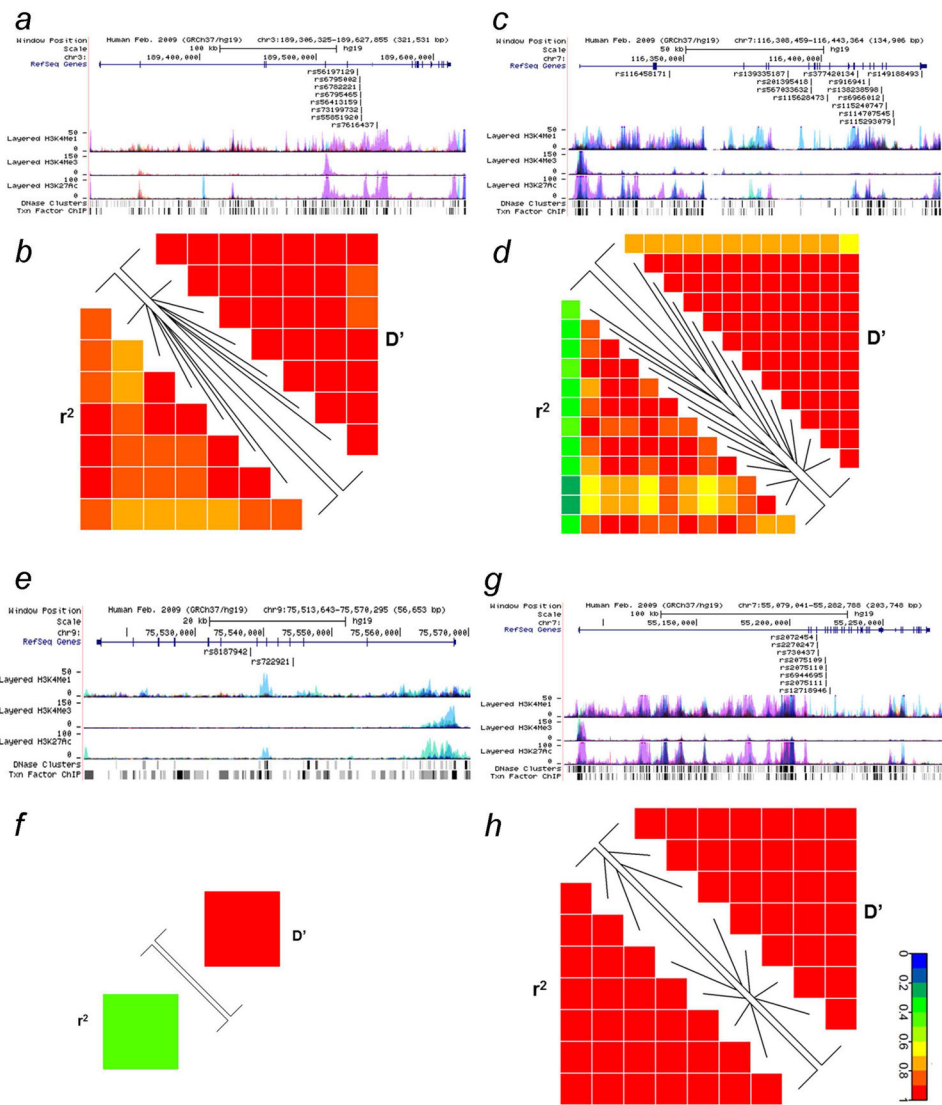


**Figure 1.** Research flowchart to identify (a) top SNPs in African descendants, (b) top SNPs in non-Hispanic whites and differences between the two racial populations.

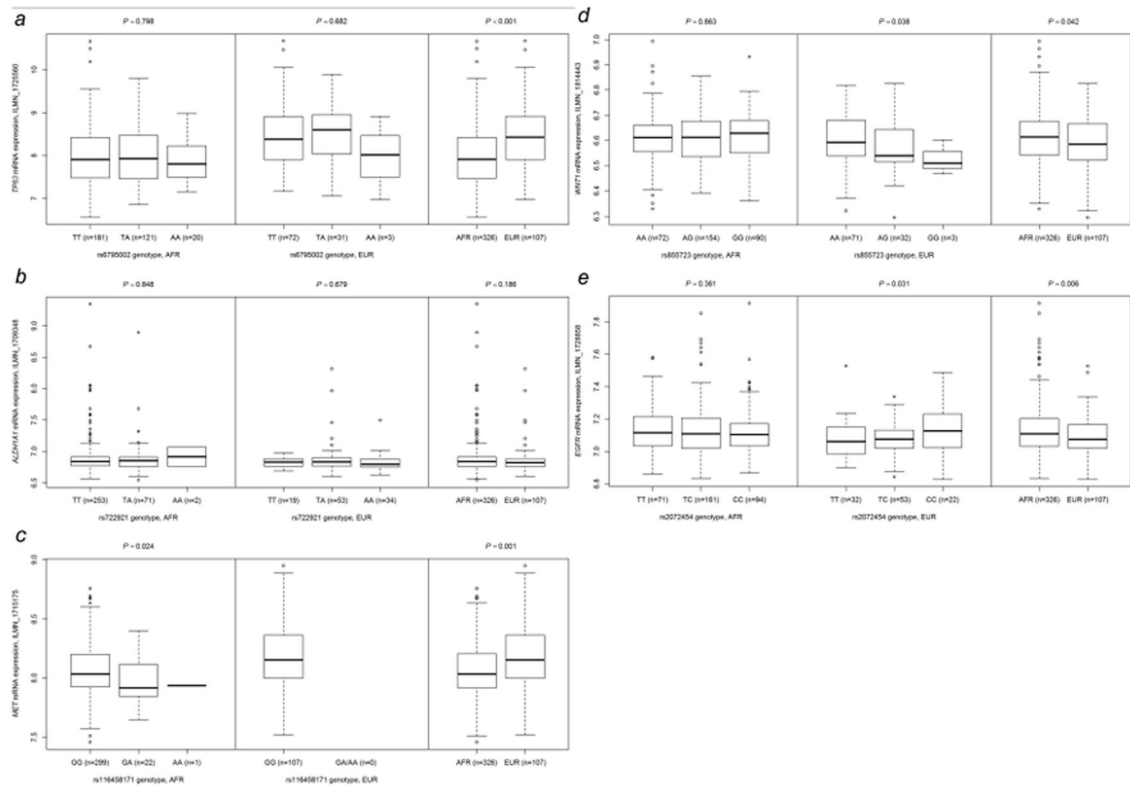


**Figure 2.**

Manhattan plots of the four studies and the meta-analysis results of the two racial populations. The red horizontal line indicates  $P = 0.05$  and the blue line indicates FDR = 0.2. (a) 6,549 common SNPs from Africans of the Ghana study. (b) 6,267 common SNPs from African descendants of the MEC AA study. (c) The meta-analysis of 5,448 SNPs in two studies of African descendants. (d) 5,239 common SNPs from non-Hispanic whites of the PLCO study. (e) 5,345 common SNPs from non-Hispanic whites of the BPC3 study. (f) The meta-analysis of 4,934 SNPs in two studies of non-Hispanic whites.



**Figure 3.** Overview of the top SNPs in the two racial populations and linkage disequilibrium (LD) analysis based on the 1000 Genomes Project Phase 3 database. Gene regions from the UCSC browser (NCBI37/hg19) (a) *TP63*, (c) *MET*, (e) *ALDH1A1* and (g) *EGFR*. LD analysis in Africans: (b) eight SNPs in *TP63*, (d) 13 SNPs in *MET* and (f) two SNPs in *ALDH1A1*. LD analysis in Europeans: (h) eight SNPs in *EGFR*.

**Figure 4.**

Correlation between SNPs and the mRNA expression level of corresponding genes in lymphoblastoid cell lines from the HapMap 3 project. Each figure depicts results of each SNP in 326 Africans, 107 Europeans and the comparison of overall expression levels between the two populations: (a) rs6795002 and *TP63*, (b) rs722921 and *ALDH1A1*, (c) rs116458171 and *MET*, (d) rs855723 and *WNT1* and (e) rs2072454 and *EGFR*. Abbreviations: AFR, African; EUR, European.

**Table 1**

The top SNPs associated with prostate cancer risk by FDR 0.2 in two racial groups

SNP	CHR	Position	Gene	Allele <sup>a</sup>	EAF <sup>b</sup>	OR (95%CI) <sup>c</sup>	P <sup>c</sup>	FDR <sup>d</sup>	eQTL in Africans <sup>e</sup>			eQTL in Europeans <sup>e</sup>		
									P	Beta	SE	P	Beta	SE
<b>24 Top SNPs in African Descendants</b>														
rs56197129	3	189534843	<i>TP63</i>	C/A	0.17	1.35 (1.15–1.60)	3.16E-04	0.170	0.801	-0.016	0.064	0.766	-0.041	0.139
rs6795002	3	189537082	<i>TP63</i>	T/A	0.18	1.35 (1.15–1.59)	2.51E-04	0.170	0.798	-0.016	0.064	0.682	-0.056	0.138
rs6782221	3	189537156	<i>TP63</i>	G/A	0.16	1.35 (1.14–1.60)	4.17E-04	0.170	0.887	-0.009	0.062	0.766	-0.041	0.139
rs6795465	3	189537521	<i>TP63</i>	T/C	0.16	1.36 (1.15–1.61)	3.46E-04	0.170	0.981	-0.001	0.064	0.959	0.007	0.144
rs56413159	3	189537609	<i>TP63</i>	G/C	0.16	1.34 (1.13–1.59)	6.72E-04	0.170	0.995	0.0004	0.063	0.766	-0.041	0.139
rs73199732	3	189538240	<i>TP63</i>	C/G	0.17	1.36 (1.15–1.60)	2.88E-04	0.170	0.883	-0.009	0.063	0.766	-0.041	0.139
rs55851920	3	189538560	<i>TP63</i>	T/C	0.17	1.36 (1.15–1.60)	2.77E-04	0.170	0.883	-0.009	0.063	0.671	-0.059	0.138
rs7616437	3	189551420	<i>TP63</i>	A/G	0.18	1.37 (1.16–1.60)	1.44E-04	0.170	0.597	-0.032	0.061	0.912	-0.013	0.114
rs116458171*	7	116345048	<i>MET</i>	G/A	0.04	1.65 (1.23–2.21)	7.18E-04	0.170	0.024	-0.107	0.047	--	--	--
rs139335187*	7	116387806	<i>MET</i>	AACTCGGCTCTGG/A	0.05	1.66 (1.26–2.19)	2.96E-04	0.170	0.339	-0.039	0.040	--	--	--
rs201395418*	7	116396348	<i>MET</i>	ATAAT/A	0.05	1.58 (1.21–2.05)	6.66E-04	0.170	0.445	-0.029	0.039	--	--	--
rs567033632*	7	116396353	<i>MET</i>	A/C	0.05	1.58 (1.21–2.05)	6.66E-04	0.170	0.445	-0.029	0.039	--	--	--
rs115628473*	7	116402123	<i>MET</i>	A/C	0.05	1.63 (1.24–2.14)	4.11E-04	0.170	0.259	-0.044	0.039	--	--	--
rs377420134*	7	116413125	<i>MET</i>	GA/G	0.05	1.56 (1.21–2.02)	7.10E-04	0.170	0.231	-0.045	0.038	--	--	--
rs916941*	7	116417210	<i>MET</i>	T/C	0.05	1.57 (1.20–2.04)	8.10E-04	0.184	0.445	-0.029	0.039	--	--	--
rs138238598*	7	116421879	<i>MET</i>	G/A	0.05	1.65 (1.26–2.17)	2.94E-04	0.170	0.384	-0.036	0.041	--	--	--
rs6966012*	7	116422214	<i>MET</i>	C/T	0.05	1.59 (1.22–2.07)	5.54E-04	0.170	0.502	-0.026	0.039	--	--	--
rs115240747*	7	116423320	<i>MET</i>	A/G	0.05	1.59 (1.23–2.06)	4.57E-04	0.170	0.231	-0.045	0.038	--	--	--
rs114707545*	7	116425961	<i>MET</i>	C/A	0.06	1.62 (1.26–2.08)	1.47E-04	0.170	0.410	-0.031	0.037	--	--	--
rs115293079*	7	116426224	<i>MET</i>	C/T	0.06	1.62 (1.26–2.08)	1.47E-04	0.170	0.410	-0.031	0.037	--	--	--
rs149188493*	7	116436394	<i>MET</i>	C/A	0.05	1.60 (1.23–2.09)	4.70E-04	0.170	0.597	-0.021	0.040	--	--	--
rs8187942*	9	75538072	<i>ALDH1A1</i>	C/G	0.06	0.60 (0.44–0.80)	4.90E-04	0.170	0.294	-0.051	0.049	--	--	--
rs722921	9	75544299	<i>ALDH1A1</i>	T/A	0.18	0.73 (0.61–0.88)	6.84E-04	0.170	0.848	0.007	0.037	0.679	0.013	0.032



SNP	CHR	Position	Gene	Allele <sup>d</sup>	EAF <sup>b</sup>	OR (95%CI) <sup>c</sup>	P <sup>c</sup>	FDR <sup>d</sup>	eQTL in Africans <sup>e</sup>			eQTL in Europeans <sup>e</sup>		
									P	Beta	SE	P	Beta	SE
rs55723	12	49370547	WNT1	A/G	0.48	1.26 (1.10–1.43)	5.80E-04	0.170	0.863	0.001	0.008	0.038	-0.093	0.045
<b>8 Top SNPs in Non-Hispanic Whites</b>														
rs2072454	7	55214348	EGFR	T/C	0.48	1.23 (1.10–1.37)	2.66E-04	0.164	0.361	-0.005	0.006	0.031	0.036	0.017
rs2270247	7	55214647	EGFR	T/G	0.48	1.23 (1.10–1.37)	2.47E-04	0.164	0.367	-0.005	0.006	0.031	0.036	0.017
rs730437	7	55215018	EGFR	C/A	0.48	1.23 (1.10–1.37)	2.47E-04	0.164	0.367	-0.005	0.006	0.031	0.036	0.017
rs2075109	7	55218903	EGFR	C/T	0.48	1.23 (1.11–1.37)	1.74E-04	0.164	0.675	-0.001	0.004	0.031	0.036	0.017
rs2075110	7	55219159	EGFR	T/C	0.48	1.23 (1.11–1.38)	1.69E-04	0.164	0.675	-0.014	0.033	0.031	0.036	0.017
rs6944695	7	55219290	EGFR	C/T	0.48	1.23 (1.10–1.37)	2.21E-04	0.164	0.910	-0.012	0.107	0.031	0.036	0.017
rs2075111	7	55219307	EGFR	G/C	0.48	1.23 (1.10–1.37)	2.21E-04	0.164	0.313	-0.012	0.012	0.031	0.036	0.017
rs12718946	7	55221447	EGFR	G/C	0.48	1.23 (1.10–1.37)	1.86E-04	0.164	0.675	-0.012	0.028	0.031	0.036	0.017

Abbreviations: CHR, chromosome; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval; FDR, false discovery rate; eQTL, expression quantitative trait loci; SE, standard error

\* 13 SNPs in *MET* and one SNP in *ALDH1A1* were found only in populations of African descent.

<sup>a</sup> Referring to "reference allele/effect allele".

<sup>b</sup> EAF in controls.

<sup>c</sup> Meta-analysis of the two studies in the same racial group. Logistic regression analysis was adjusted for age and principal components in each study.

<sup>d</sup> FDR was calculated in each racial group.

<sup>e</sup> eQTL were analyzed based on datasets from the HapMap3 project with 107 Europeans and 326 Africans.

**Table 2**

The 14 SNPs exhibiting differences between two racial groups

SNP	Gene	Location	MEC & Ghana, African			PLCO & BPC3, non-Hispanic white			Heterogeneity <sup>c</sup>		Meta-regression <sup>d</sup>		Interaction <sup>e</sup>
			Allele <sup>a</sup>	EAF	OR (95%CI) <sup>b</sup>	P <sup>b</sup>	Allele <sup>a</sup>	EAF	OR (95%CI) <sup>b</sup>	P <sup>b</sup>	Q	I <sup>2</sup>	
<b>6 Top SNPs identified in African Descendants</b>													
rs56197129	<i>TP63</i>	Intron	C/A	0.17	1.35 (1.15–1.60)	3.16E-04	0.110	0.148	52.3	0.148	0.148	0.007	
rs6795002	<i>TP63</i>	Intron	T/A	0.18	1.35 (1.15–1.59)	2.51E-04	0.114	0.142	53.6	0.142	0.142	0.006	
rs55851920	<i>TP63</i>	Intron	T/C	0.17	1.36 (1.15–1.60)	2.77E-04	0.097	0.152	51.2	0.152	0.152	0.011	
rs7616437	<i>TP63</i>	Intron	A/G	0.18	1.37 (1.16–1.60)	1.44E-04	0.461	0.106	61.7	0.106	0.080	0.004	
rs729221 <sup>f</sup>	<i>ALDH1A1</i>	Intron	T/A	0.18	0.73 (0.61–0.88)	0.001	0.991	0.019	81.7	0.019	0.033	0.032	
rs855723	<i>WNT1</i>	5'-Upstream	A/G	0.48	1.26 (1.10–1.43)	0.001	0.589	0.067	70.1	0.067	0.067	0.040	
<b>8 Top SNPs identified in Non-Hispanic Whites</b>													
rs2072454	<i>EGFR</i>	Exon-Synonymous	T/C	0.51	0.89 (0.72–1.10)	0.285	2.66E-04	0.009	85.6	0.012	0.012	0.001	
rs2270247	<i>EGFR</i>	Intron	T/G	0.50	0.89 (0.73–1.10)	0.293	2.47E-04	0.008	85.6	0.012	0.012	0.001	
rs730437	<i>EGFR</i>	Intron	C/A	0.51	0.89 (0.73–1.10)	0.291	2.47E-04	0.008	85.7	0.012	0.012	0.001	
rs2075109	<i>EGFR</i>	Intron	C/T	0.45	0.98 (0.86–1.11)	0.755	1.74E-04	0.007	86.0	0.016	0.016	0.024	
rs2075110	<i>EGFR</i>	Intron	T/C	0.37	0.96 (0.84–1.09)	0.506	1.69E-04	0.004	88.3	0.008	0.008	0.013	
rs6944695	<i>EGFR</i>	Intron	C/T	0.50	0.93 (0.82–1.06)	0.281	2.21E-04	0.001	90.4	0.002	0.002	0.004	
rs2075111	<i>EGFR</i>	Intron	G/C	0.50	0.93 (0.82–1.06)	0.264	2.21E-04	0.001	90.6	0.002	0.002	0.003	
rs12718946	<i>EGFR</i>	Intron	G/C	0.42	0.91 (0.80–1.04)	0.160	1.86E-04	0.001	91.8	0.001	0.001	0.003	

Abbreviations: MEC, Multiethnic/Minority Cohort Study of Diet and Cancer; PLCO, Prostate, Lung, Colorectal and Ovarian; BPC3, Breast and Prostate Cancer Cohort Consortium; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval

<sup>a</sup>Referring to "reference allele/effect allele".

<sup>b</sup>Meta-analysis of the two studies in each population. Logistic regression analysis was adjusted for age and principal components in each study.

<sup>c</sup>Heterogeneity between results from the two racial groups were defined as Q-test  $P = 0.100$  or  $I^2 = 50.0\%$ .

<sup>d</sup>Meta-regression analysis of results from the four studies (MEC, Ghana, PLCO and BPC3) as independent datasets included the racial group as a moderator. Results of the four studies were shown in Supporting Information Table 6.

<sup>e</sup>The multiplicative interaction of each SNP and racial group in logistic regression analysis was adjusted for age and racial group in a pooled dataset from the four studies with 2,753 cases and 3,666 controls.

All four studies used A as the effect allele for rs722921 in analyses of Q-test,  $I^2$ , meta-regression and multiplicative interaction.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript