Genetic Hitchhiking and Population Bottlenecks Contribute to Prostate Cancer Disparities in Men of African Descent

Joseph Lachance1, Ali J. Berens1, Matthew E.B. Hansen2, Andrew K. Teng1, Sarah A. Tishkoff2, and Timothy R. Rebbeck3

Abstract

Prostate cancer incidence and mortality rates in African and African American men are greatly elevated compared with other ethnicities. This disparity is likely explained by a combination of social, environmental, and genetic factors. A large number of susceptibility loci have been reported by genome-wide association studies (GWAS), but the contribution of these loci to prostate cancer disparities is unclear. Here, we investigated the population structure of 68 previously reported GWAS loci and calculated genetic disparity contribution statistics to identify SNPs that contribute the most to differences in prostate cancer risk across populations. By integrating GWAS results with allele frequency data, we generated genetic risk scores for 45 African and 19 non-African populations. Tests of natural selection were used to assess why some SNPs have large allele frequency differences across populations. We report that genetic predictions of prostate cancer risks are highest for West African men and lowest for East Asian men. These differences may be explained by the out-of-Africa bottleneck and natural selection. A small number of loci appear to drive elevated prostate cancer risks in men of African descent, including rs9623117, rs6983267, rs10896449, rs10993994, and rs817826. Although most prostate cancer–associated loci are evolving neutrally, there are multiple instances where alleles have hitchhiked to high frequencies with linked adaptive alleles. For example, a protective allele at 2q37 appears to have risen to high frequency in Europe due to selection acting on pigmentation. Our results suggest that evolutionary history contributes to the high rates of prostate cancer in African and African American men.

Significance: A small number of genetic variants cause an elevated risk of prostate cancer in men of West African descent.

Introduction

Men of African descent suffer disproportionately from prostate cancer compared with men of other ethnicities (1, 2). The International Agency for Research on Cancer GLOBOCAN program estimates that prostate cancer has the highest incidence of any tumor site in African-American, Caribbean, and African men (3). Prostate cancer is also the leading cause of cancer-specific mortality in African and Caribbean men, and it is second only to lung cancer as the leading cause of cancer-related deaths in African-American men (3). In the United States, rates of prostate cancer are substantially higher for African-American men than those from other groups, including a 69% higher incidence rate and a 134% higher death rate relative to European-American men (4). This health disparity is likely to be explained by the complex multifactorial effects of social, environmental, and genetic factors.

Prostate cancer has the highest familial risks of any major cancer (5, 6). Heritability estimates of prostate cancer also range from 42% to 58% (7, 8), indicating that there is a strong genetic component to prostate cancer etiology. To date, over 100 prostate cancer susceptibility loci have been identified via genome-wide association studies (GWAS), the majority of which have been identified in studies of men with European or Asian ancestry (9). Admixture mapping has identified a large genomic region at 8q24 that plays an important role in the elevated prostate cancer risk of African-American men (10), and novel prostate cancer associations have been detected in studies of African and African-American men (11, 12). There also is evidence that rare variants contribute to a substantial fraction of prostate cancer risk within each population (13). A previous study that included seven African populations from the Human Genome Diversity Project (HGDP) reported that men of African descent have elevated risks of prostate cancer (14). However, HGDP populations capture only a small subset of the genetic diversity that exists in sub-Saharan Africa (15), and it is unknown how much the genetic risks of prostate cancer vary across the continent. In addition, many additional prostate cancer susceptibility loci have been identified since this report (9, 14).

Allele frequency differences at disease susceptibility loci may contribute to differences in prostate cancer risk across populations (16, 17). Large allele frequency differences can quickly arise via founder effects and population bottlenecks (18). The great reshuffling of allele frequencies that occurred during the out-of-Africa
bottleneck may have contributed to health disparities and differences in prostate cancer risk across populations (19, 20). Another possibility is that allele frequencies at prostate cancer loci changed during the forced migration of African slaves to the new world (21). However, the middle passage did not involve a severe population bottleneck (22). Local adaptation is an alternative, but not mutually exclusive, mechanism that can result in large allele frequency differences between populations (23). Selective sweeps of adaptive alleles also lead to allele frequency changes at linked loci—a process known as genetic hitchhiking (24).

The late age of onset of prostate cancer implies that disease-causing alleles are likely to have negligible fitness effects, which is likely to limit the role of natural selection acting directly on prostate cancer susceptibility loci. Therefore, we hypothesize that genetic hitchhiking of prostate cancer risk alleles with closely linked adaptive alleles at other loci contributes to allele frequency differences and health disparities between populations.

In the present study, we combine GWAS results with allele frequencies from a broad panel of African and non-African populations to better understand the genetic architecture of prostate cancer risk and disparities. We address three questions: (i) Which prostate cancer susceptibility SNPs contribute the most prostate cancer risk and disparities. We address three questions: (i) Which prostate cancer susceptibility SNPs contribute the most prostate cancer risk and disparities. We address three questions: (ii) Which populations have the highest genetic risk of prostate cancer? (iii) Which prostate cancer susceptibility SNPs contribute the most prostate cancer risk and disparities. We address three questions: (iii) How much does natural selection contribute to prostate cancer disparities and increased risk in African men?

Materials and Methods

Population genetic data

Allele frequencies were obtained for 64 global populations. This dataset includes 648 individuals from 38 African populations collected by the laboratory of Dr. Sarah Tishkoff and 2,504 individuals from 26 populations from Phase III of the 1000 Genomes Project (25). Populations from the Tishkoff Laboratory dataset were genotyped using the Illumina1M-Duo array, and populations from Phase III of the 1000 Genomes Project were genotyped using low-coverage whole-genome sequencing. These data were combined to produce a set of 1,034,074 genotyped SNPs across 64 global populations (Supplementary Tables S1 and S2, and Supplementary Fig. S1). Genotypes at these SNPs were observed, rather than imputed.

GWAS-identified prostate cancer susceptibility loci

We obtained information on all known prostate cancer susceptibility loci based on GWAS archived in the NHGRI-EBI GWAS Catalog (9). We included one additional prostate cancer study that was not in the GWAS Catalog (26). These 26 prostate cancer GWAS yielded 169 unique SNPs that met a P value cutoff of $1.0 \times 10^{-5}$. This P value cutoff was chosen because of prior evidence that genetic risk scores (GRS) are more accurate if less stringent cutoffs are used (27, 28). After excluding eight X-linked SNPs, 161 unique autosomal SNPs remained. X-linked SNPs were excluded from our analysis because genome-wide CMS scores (see below) are only available for autosomal loci. We then filtered out SNPs that lacked odds ratio (OR) information and SNPs that were not genotyped on the Illumina1M-Duo array. When prostate cancer susceptibility loci were identified in multiple studies, we retained the OR from the prostate cancer study reporting the strongest association (i.e., smallest P value). When two or more candidate SNPs were found within 100 kb of each other we retained the SNP with the strongest prostate cancer association (i.e., smallest P value). With this approach, we winnowed the candidate prostate cancer–associated SNPs to 68 autosomal SNPs in linkage equilibrium. Sixty-three of these SNPs had P values below $5 \times 10^{-5}$. Ancestral versus derived states of risk alleles were inferred from dbSNP and 1000 Genomes Project data (25, 29). After controlling for strand flipping issues, risk allele frequencies were found for each SNP and population (Supplementary Table S2).

Genetic disparity contribution statistics

We developed the genetic disparity contribution (GDC) statistic to quantify the contribution of each SNP to differences in prostate cancer risk across populations. Here, we use an overbar notation to distinguish β coefficients that refer to diploid genotypes, as opposed to alleles. Assuming additive allelic effects, β coefficients are given by:

$$\bar{\beta}_{ijk} = \ln(1 + 2p_{ijk}(OR_i - 1)).$$

(A)

where $p_{ijk}$ is the frequency of the risk allele at the $i^{th}$ prostate cancer susceptibility locus in individual $j$ from population $k$, and OR refers to the OR of the $i^{th}$ risk allele. ORs used to calculate β coefficients were obtained from the NHGRI-EBI GWAS Catalog (9). We generated GDC statistics by calculating the mean difference in β coefficients for two different populations. The GDC of the $i^{th}$ SNP to differences in risk between African and non-African populations is given by:

$$GDC_{i,A-N} = \bar{\beta}_{iA} - \bar{\beta}_{iN},$$

(B)

where $\bar{\beta}_{iA}$ is the β coefficient associated with the $i^{th}$ SNP in African populations, $\bar{\beta}_{iN}$ is the β coefficient associated with the $i^{th}$ SNP in non-African populations. Beta coefficients in Equation B use the mean frequency of the risk allele at locus $i$ in African and non-African populations. GDC statistics were also used to quantify which SNPs contribute the most to hereditary differences in disease risks between West and East African populations.

Genetic risk score calculations

GRS were calculated to yield predictions of prostate cancer risk across populations. These scores take into account allele frequency and OR information for each SNP. Assuming that all 68 prostate cancer susceptibility loci are in linkage equilibrium, the GRS of individual $j$ from population $k$ is given by:

$$GRS_{i,k} = \sum_{i=1}^{68} \bar{\beta}_{ijk}.$$

(C)

where $\bar{\beta}_{ijk}$ is the β coefficient associated with the $i^{th}$ prostate cancer susceptibility locus in individual $j$ from population $k$. An individual who is homozygous for the protective allele at all 68 loci would have a GRS of zero (although the likelihood of this occurring is exceedingly small). For each prostate cancer–associated SNP, we assume that risk alleles have the same effect size in each population.

To compare prostate cancer risks in different populations and obtain GRS distributions for each population, we simulated genotypes for one million individuals per population. These simulations used population-specific allele frequency information at each prostate cancer susceptibility locus. Assuming...
We tested whether population ranks of prostate cancer risk are robust to weighting GRS calculations by OR. Unweighted GRS statistics were calculated using the more OR for each SNP (i.e., OR, 1.167, the mean OR of all 68 prostate cancer risk alleles). Weighted GRS statistics were calculated using ORs from the GWAS Catalog (i.e., using the reported OR for each prostate cancer SNP). Note that the median GRS of simulated ORs from the GWAS Catalog (i.e., using the reported OR for each prostate cancer risk allele) and the mean OR of all 68 prostate cancer risk alleles are robust to weighting GRS calculations by OR. We also tested whether population ranks of prostate cancer risk are the direct targets of positive selection, we also obtained the maximum CMS score per 200kb window for each continent. Note that direct targets of selection tend to have CMS scores that are among the highest in each 200kb window (33), whereas alleles that are subject to genetic hitchhiking can have moderate CMS scores. CMS tests have an ability to detect positive selection when selection coefficients exceed 0.5% (33).

We used Berg and Coop’s framework to test whether prostate cancer SNPs exhibit signatures of polygenic adaptation (34). This approach uses a weighted sum of allele frequencies to estimate genetic values for each population, and it assumes that allele frequency differences among populations should be uncorrelated with the sign and magnitude of prostate cancer effect sizes if SNPs are evolving neutrally. Berg and Coop’s approach uses the test statistic Qx to quantify the among population variance in estimated genetic values that is not explained by genetic drift, and Qx statistics can be viewed as a type of QST/FST comparison. We converted Qx statistics into P-values by comparing Qx for our set of 68 independent prostate cancer SNPs to an empirical null distribution of putatively neutral SNPs. To generate this null distribution, we matched SNPs by 2% minor allele frequency bin and distance to the nearest known gene.

### Results

#### Prostate cancer allele frequencies

Risk allele frequencies at many prostate cancer loci vary across the globe (Fig. 1). On a broad scale, allele frequencies differ for African and non-African populations. These patterns are due in part to population bottlenecks that followed the out-of-Africa migration. Allele frequency differences yield genetic risks of prostate cancer that are elevated for West African populations, intermediate for East African populations, and low for non-African populations. However, risk allele frequencies also show extensive heterogeneity within each continent. Some prostate cancer–associated disease SNPs have small risk allele frequency differences across populations while others have marked differences in risk allele frequencies across populations. Africa is not monomorphic with respect to the genetic risk of prostate cancer, and there are some non-African populations with higher predicted risks of prostate cancer than African populations.

#### Ranking SNPs by their contribution to prostate cancer disparities

The relative contribution of different SNPs to prostate cancer disparities depends on allele frequency differences between African and non-African populations and whether these SNPs have small or large ORs (Fig. 2). Thirty-eight out of 68 prostate cancer SNPs have higher risk allele frequencies in African populations, and we are unable to reject the null hypothesis that this difference is due to chance (P-value = 0.3961, exact binomial test). However, nine of the 13 most divergent SNPs have higher risk allele frequencies in African populations than in non-African populations. In addition, the five SNPs with the largest ORs have higher risk frequencies in Africa. These patterns suggest that prostate cancer health disparities between African and non-African populations may be driven by a relatively small number of SNPs. Each SNP has an ancestral allele that is shared with other primates, and a derived allele that is due to a recent mutation (35). Figure 2 reveals that risk allele frequencies tend to be higher in Africa when risk alleles are ancestral, and risk allele frequencies tend to be

---

**Ethnicity-specific estimates of prostate cancer risk**

Age-adjusted prostate cancer mortality and incidence rates were obtained from the United States Centers for Disease Control and Prevention and the National Cancer Institute (CDC and NCI, 2012 data; ref. 31). None of the 64 populations in our dataset came from American Indian or Alaska Native populations, so this category was removed from our analysis. We focused on a subset of populations from the 1000 Genomes Project that are representative of different ethnicities (African: ACB and ASW; Asian/Pacific Islander: CDX, CHB, CHS, KHV, and JPT; European: CEU, FIN, GBR, IB5, and TSI; Hispanic: CLM, MXL, PHL, and PUR). After converting prostate cancer incidence and mortality rates to a natural log scale, ethnicity-specific estimates of prostate cancer risk were plotted versus median GRS for each population. Constraining the slope of the line to equal one, linear least squares regression was used to fit trend lines.

#### Tests of positive selection

Composite of multiple signals (CMS) scores (32, 33) were used to detect selection in populations from Europe (CEU), Asia (CHB+JPT), and Africa (YRI). CMS scores integrate multiple signatures of natural selection, including extended haplotype homozygosity, high values of 

\[
F_{ST}
\]

and elevated frequencies of derived alleles. Genome-wide CMS scores were downloaded as a UCSC Genome Browser track (https://www.broadinstitute.org/cms/results). Approximately two-thirds (611,675) of the over 1 million genotyped SNPs in our dataset had CMS scores. For the remaining SNPs, we selected the closest neighboring SNP with a CMS score as a representative value. The 95th percentile of this genome-wide set of CMS scores was used as a cutoff to identify outlier SNPs that are potential targets of positive selection. Because prostate cancer SNPs need not be the direct targets of positive selection, we also obtained the maximum CMS score per 200kb window for each continent. Note that direct targets of selection tend to have CMS scores that are among the highest in each 200kb window (33), whereas alleles that are subject to genetic hitchhiking can have moderate CMS scores. CMS tests have an ability to detect positive selection when selection coefficients exceed 0.5% (33).

We used Berg and Coop’s framework to test whether prostate cancer SNPs exhibit signatures of polygenic adaptation (34). This approach uses a weighted sum of allele frequencies to estimate genetic values for each population, and it assumes that allele frequency differences among populations should be uncorrelated with the sign and magnitude of prostate cancer effect sizes if SNPs are evolving neutrally. Berg and Coop’s approach uses the test statistic Qx to quantify the among population variance in estimated genetic values that is not explained by genetic drift, and Qx statistics can be viewed as a type of QST/FST comparison. We converted Qx statistics into P-values by comparing Qx for our set of 68 independent prostate cancer SNPs to an empirical null distribution of putatively neutral SNPs. To generate this null distribution, we matched SNPs by 2% minor allele frequency bin and distance to the nearest known gene.

### Results

#### Prostate cancer allele frequencies

Risk allele frequencies at many prostate cancer loci vary across the globe (Fig. 1). On a broad scale, allele frequencies differ for African and non-African populations. These patterns are due in part to population bottlenecks that followed the out-of-Africa migration. Allele frequency differences yield genetic risks of prostate cancer that are elevated for West African populations, intermediate for East African populations, and low for non-African populations. However, risk allele frequencies also show extensive heterogeneity within each continent. Some prostate cancer–associated disease SNPs have small risk allele frequency differences across populations while others have marked differences in risk allele frequencies across populations. Africa is not monomorphic with respect to the genetic risk of prostate cancer, and there are some non-African populations with higher predicted risks of prostate cancer than African populations.

#### Ranking SNPs by their contribution to prostate cancer disparities

The relative contribution of different SNPs to prostate cancer disparities depends on allele frequency differences between African and non-African populations and whether these SNPs have small or large ORs (Fig. 2). Thirty-eight out of 68 prostate cancer SNPs have higher risk allele frequencies in African populations, and we are unable to reject the null hypothesis that this difference is due to chance (P-value = 0.3961, exact binomial test). However, nine of the 13 most divergent SNPs have higher risk allele frequencies in African populations than in non-African populations. In addition, the five SNPs with the largest ORs have higher risk frequencies in Africa. These patterns suggest that prostate cancer health disparities between African and non-African populations may be driven by a relatively small number of SNPs. Each SNP has an ancestral allele that is shared with other primates, and a derived allele that is due to a recent mutation (35). Figure 2 reveals that risk allele frequencies tend to be higher in Africa when risk alleles are ancestral, and risk allele frequencies tend to be
higher in non-African populations when risk alleles are derived \((P = 0.0005,\) Fisher exact test).

A small number of loci appear to have a disproportionately large contribution to elevated prostate cancer risks in African men (Fig. 3A). Some of these SNPs have risk allele frequencies that are over 40\% higher in Africa than outside of Africa (e.g., rs7584330 near \(MLPH\) and rs9623117 near \(TNRC6B\)) whereas others have alleles of large effect (e.g., rs817826 near \(RAD23B\) and \(KLF4\), which has an OR of 1.41). We developed a novel statistic, the GDC, to rank order SNPs by how much they contribute to increased prostate cancer risk in Africa. GDC scores are a function of ORs and the relative frequencies of risk alleles in African and non-African populations. SNPs with positive GDC scores have risk allele frequencies that are higher in African populations than non-African populations. SNPs that contribute the most to increased risk in African populations include rs9623117 at 22q13.1, rs6983267 at 8q24.1, rs10896449 at 11q13.3, rs10993994 at 10q11.23, and rs817826 at 9q31.2. Note that rs10993994 is near \(MSMB\), which encodes PSP94, a major protein secreted by the prostate (36). Conversely, some SNPs increase the risk of prostate cancer in non-African populations, including rs636291 at 1p36.22 and rs2252004 at 10q26.21. Most SNPs have minimal allele frequency differences between continents, and these SNPs are unlikely to make a major contribution to health disparities. Continental differences in prostate cancer risk disappear if the six SNPs that contribute the most to elevated prostate cancer risk in African men are ignored (i.e., the sum of 62 remaining GDC scores is close to zero: 0.0099). We also examined whether the SNPs that contribute most to prostate cancer risk differences between African and non-African populations also contribute to differences in risk between West Africa and East Africa. Overall, GDC statistics between continents were positively...
correlated with GDC statistics within Africa ($r^2 = 0.217$, Supplementary Fig. S3). Many of the same SNPs contribute to health disparities on multiple spatial scales.

**Genetic risk prediction in different populations**

Using computer simulations, we calculated GRSs for one million simulated individuals from each population and ranked populations by median GRS (Fig. 3B). Here, we define an individual’s GRS as equal to the sum of single locus $\beta$ coefficients. Populations with higher median GRS scores have a higher predicted risk of prostate cancer.

Overall, populations from West Africa have the highest predicted risk, followed by East African populations, and then non-African populations. Seven of the eight populations with the highest predicted risk are located in West Africa and the nine populations with the lowest predicted risk are non-African. Focusing on extreme populations, the Fulani from Cameroon have the highest median GRS and Japanese from Tokyo have the lowest median GRS. East African populations with the highest predicted risk include the Pare and Luhya of Kenya (which share Bantu ancestry with West African populations; ref. 37), as well as the Hadza of Tanzania. The Yaaku from Kenya have the lowest predicted risk of prostate cancer of any East African population.

Among non-African populations, predicted risk tends to be greater for South Asian populations, intermediate for European populations, and lower for East Asian populations. The GRS of different populations show substantial overlap. For example, although African-Americans (ASW) tend to have a higher predicted risk than Northern-Europeans (CEU), there is a 24.9% chance that an ASW genome will have a lower GRS than a CEU genome. Note that GRS only measure genetic contributions to disease risk, and it is likely that other factors, including environmental exposures, also contribute to prostate cancer risk.

**Figure 2.** Allele frequency differences and ORs of prostate cancer susceptibility loci. Differences in risk allele frequency between pooled African and non-African populations are plotted versus ORs from published GWAS data (positive y-axis values indicate elevated risk in African populations). SNPs where the ancestral allele increases risk are labeled blue and SNPs where the derived allele increases risk are labeled gray.
predictions follow the same trend as public health data: African-American populations have the highest GRS and death rates, European and Hispanic populations have intermediate GRS and death rates, and Asian populations have the lowest GRS and death rates. Furthermore, GRSs capture the magnitude of ethnic differences in clinical disease risks (Fig. 5). There is a strong positive relationship between GRS and prostate cancer mortality (residual standard error $= 0.1816$, 15 df, Fig. 5A). Differences in GRS within each ethnicity appear to be due to admixture and different amounts of African ancestry components. There is also a strong positive relationship between GRS and age-standardized estimates of prostate cancer incidence (residual SE = 0.1862, 15 df, Fig. 5B). Taken together, our results indicate that GRS successfully quantify ethnic differences in prostate cancer incidence and mortality rates.

Tests of positive selection
Because local adaptation can result in large allele frequency differences between populations, we tested whether prostate cancer hits were under selection in CEU, YRI, or JPT+CHB...
populations using the Composite of Multiple Signals (CMS) test (32, 33). CMS scores combine several tests of positive selection into a single metric, with higher CMS scores indicating SNPs that are more likely to be adaptive. To distinguish between prostate cancer loci that are direct targets of selection as opposed to linked hitchhiking loci, we found the CMS score nearest to each GWAS SNP and the largest CMS score within a 200kb window centered around each GWAS SNP (Fig. 6A). SNPs that are direct targets of

![Image of ADMIXTURE plot](image_url)

**Figure 4.** Prostate cancer risks vary for different genetic ancestries. A, ADMIXTURE plot ($k = 12$) of 3,152 genotyped individuals. Each color corresponds to a different genetic ancestry component. Individuals are grouped by study population, and study populations are ranked by median GRS. B–M, GRS are plotted against the proportion of each individual’s genome that has a particular ancestry component. Data were fit to a linear model, and correlation coefficients indicate the extent to which each ancestry component is associated with GRS.
cancer SNPs with the highest East Asian CMS scores include rs103294 at 19q13.42 and rs1465618 at 2p21, and prostate cancer SNPs with the highest West African CMS scores include rs9623117 at 22q13.1 and rs11672691 at 19q13.2. We note that there are multiple prostate cancer SNPs that have low CMS scores, but are found in genomic regions with high CMS scores. This is consistent with genetic hitchhiking of disease SNPs. Further support for the genetic hitchhiking hypothesis comes from the fact that there are no instances of a prostate cancer SNP having the highest CMS score in its 200kb window.

We also tested whether prostate cancer SNPs exhibit signatures of polygenic adaptation using the approach of Berg and Coop (34). For our collective set of 68 unlinked prostate cancer SNPs, we find that the cumulative variance in prostate cancer risk among African populations is not significantly higher than expected by genetic drift alone ($P$ value $= 0.429$, $Q_X = 51.3$). $Q_X$ statistics measure the extent to which trait-associated SNPs vary more across population than one would expect from genetic drift. Although some prostate cancer SNPs are outliers with large allele frequency differences between populations, the overall set of prostate cancer SNPs resembles the rest of the genome. Berg and Coop’s approach can also be used to determine if groups of populations have genetic values that deviate from neutral expectations (34). Focusing on different geographic regions of Africa, we find that East African populations have lower prostate cancer risks than expected by genetic drift alone ($P$ value $= 0.019$, $Z$-score $= – 2.339$) and that West African populations have slightly higher prostate cancer risks than expected by drift alone ($P$ value $= 0.087$, $Z$-score $= 1.711$). This lends additional support to our inference that the genetic risk of prostate cancer in East Africans differs from that of West Africans.

**Evaluation of underlying assumptions**

There are some caveats to our approach, including the use of published ORs and the fact that most reported prostate cancer susceptibility loci have been identified in non-African populations (39). The GWAS Catalog provides reported ORs for study populations in which associations were detected, but these alleles may confer different ORs in other (yet unstudied) populations. To address the potential for OR heterogeneity across populations, we calculated GRS two different ways. Weights were calculated using ORs reported in the GWAS Catalog, and unweighted GRS calculations assumed that each disease SNP confers the same OR magnitude of association on prostate cancer risk in all populations. Comparing these two approaches, we find that the rank order of populations from different geographic regions is broadly conserved regardless of the OR used, and that GRS are highly correlated for both approaches ($R^2 = 0.967$, Supplementary Fig. S4). We also tested whether GRS predictions are robust to inclusion of specific prostate cancer SNPs via bootstrapping. West African populations consistently had higher predicted risk than East African populations (97.5% bootstrap support), and East African populations consistently had higher predicted risk than non-African (99.5% bootstrap support). The rank order of prostate cancer risk for individual populations depends upon which SNPs are included in GRS calculations. Although Fulani (CAFU) had the highest median GRS when all 68 prostate cancer SNPs were included, only 3% of the bootstrap runs resulted in Fulani having the highest GRS. By contrast, 25% of the bootstrap runs resulted in Lemande (CALM) having the

---

**Figure 5.**

Strong concordance between ethnicity-specific estimates of prostate cancer risk and GRS predictions. Age-adjusted prostate cancer death rates and incidence rates from the United States were compared with the median GRS of each population. Ethnicity-specific data from CDC and NCI are grouped by ethnicity. The solid line in each panel is the best fit to data after constraining the slope to equal one. **A,** Prostate cancer–related death rates (natural log scale) versus median GRS. **B,** Prostate cancer incidence rates (natural log scale) versus median GRS.
Finally, we note that the set of known disease loci is incomplete. In particular, European and Asian GWAS are likely to miss prostate cancer susceptibility loci that have intermediate frequency alleles in African populations and rare alleles in non-African populations. To the extent that additional prostate cancer SNPs will be found, disease risks may be underestimated for men of African descent in our current analysis. Nevertheless, we note that there is a striking concordance between our GRS predictions and clinical estimates of disease risk (Fig. 5).

Discussion

Like most of the genome, the majority of prostate cancer susceptibility loci evolve neutrally with moderate allele frequency differences between populations. However, a small number of SNPs make a disproportionally large contribution to population-level differences in the genetic risks of prostate cancer. Characteristics of these loci include larger effect sizes as estimated in GWAS studies and large allele frequency differences between African and non-African populations (Fig. 3). A relatively small number of...
SNPs have large GDC statistics, which in part explains how genetic health disparities can arise for a polygenic disease like prostate cancer. We note that SNPs with the highest GDC statistics need not be same SNPs that contribute the most to heritability within a single population. Both neutral and selective evolutionary mechanisms appear to have contributed to disparities in the genetic risk of prostate cancer. These mechanisms include founder effects due to the out-of-Africa migration and genetic hitchhiking of disease susceptibility alleles with locally adaptive alleles.

Here, we highlight three genomic regions that contribute to prostate cancer health disparities: 2q37, 22q13, and 8q24. rs7584330 is a prostate cancer SNP that is located at 2q37. For the populations studied in this paper, the mean frequency of the risk allele at rs7584330 is 47% higher in African populations compared with non-African populations. European CMS scores indicate that there is evidence of positive selection at 2q37 (Supplementary Fig. S5). This region contains the melanophilin (MLPH) and prolactin releasing hormone (PRLH) genes. MLPH mutations have been associated with dilated skin and hair pigmentation (40), and signatures of positive selection at 2q37 are consistent with previous studies that suggest variation at pigmentation genes is adaptive in non-African populations (41, 42). The PRLH gene encodes prolactin-releasing peptide, and levels of prolactin are known to affect male fertility (43). These data suggest that a haplotype that protects against prostate cancer has hitchhiked to high frequencies in Europe as an incidental byproduct of linkage and selection for lighter pigmentation. We also observed risk allele frequencies differences and signatures of positive selection at 2q37 that are consistent with previous studies that suggest variation at pigmentation genes is adaptive in non-African populations (41, 42). The PRLH gene encodes prolactin-releasing peptide, and levels of prolactin are known to affect male fertility (43). These data suggest that a haplotype that protects against prostate cancer has hitchhiked to high frequencies in Europe as an incidental byproduct of linkage and selection for lighter pigmentation.

For the populations studied in this paper, the mean frequency of the risk allele at rs9623117 is 61% higher in African populations compared with non-African populations. Elevated CMS scores at 22q13 suggest that there may have been recent positive selection in West Africa and East Asia (Supplementary Fig. S6). rs9623117 is close to the gene TNRC6B, the products of which play a role in RNA-mediated gene silencing (44). However, it is unclear what trait is responsible for the signatures of positive selection near rs9623117. For the populations studied in this paper, the mean frequency of the risk allele at rs9623117 is 61% higher in African populations compared with non-African populations. Elevated CMS scores at 22q13 suggest that there may have been recent positive selection in West Africa and East Asia (Supplementary Fig. S6). rs9623117 is close to the gene TNRC6B, the products of which play a role in RNA-mediated gene silencing (44). However, it is unclear what trait is responsible for the signatures of positive selection near rs9623117.

The 8q24 locus has been repeatedly implicated in prostate cancer and other cancers. In this region alone, there are at least three independent prostate cancer susceptibility loci and 21 additional prostate cancer SNPs that have been excluded from our analysis of 68 SNPs due to linkage disequilibrium. Two of the independent 8q24 prostate cancer loci have much higher risk allele frequencies in African populations (rs6938267 and rs4242382) whereas a third locus has only a small contribution to health disparities, as indicated by a GDC statistic that is close to zero (rs10163343). Some of the excluded prostate cancer SNPs in 8q24 have large effect sizes (e.g. rs116041037, rs188140481, and rs4242384 have ORs that exceed 1.8). The 8q24 region contains MYC, an oncogene that encodes a transcription factor that is associated with many human cancers (45). However, despite the importance of 8q24 to prostate cancer risk and disparities, CMS scores do not reveal evidence of recent positive selection in this genomic region (Supplementary Fig. S7).

We report that the predicted genetic risks of prostate cancer vary geographically, with African populations showing the highest risk. This is consistent with existing evidence that African ancestry is enriched in African-American prostate cancer cases relative to African-American controls (10). Within Africa, GRS are highest for men of West African descent. However, there is substantial heterogeneity in GRS within and between populations. This heterogeneity is particularly relevant to the health of African-Americans, as individuals can trace their ancestry back to many different African and non-African source populations (15, 46). We also note that ASW individuals represent only a subset of the diversity found in African-American populations. Finally, we note that Fulani who have settled in West African towns (CAFU) have higher GRS statistics than Mbororo Fulani (CAMF) who are nomadic pastoralists. Differences in risk among Fulani individuals may be due to differential admixture with West African populations that have Bantu ancestry (15).

Although GRS are lower for East African populations than West African populations, GLOBOCAN estimates of prostate cancer incidence and mortality rates are similar for East African and West African countries (3). This pattern may be explained by environmental risk factors that have a larger contribution to prostate cancer in East African populations, or because GLOBOCAL estimates of prostate cancer incidence and morality are based on limited data. Furthermore, we note that having a high GRS does not guarantee that an individual will develop prostate cancer. Genotype-by-environment interactions and socioeconomic factors also contribute to prostate cancer risks and health disparities (47).

Going forward, there is a need to undertake GWAS and linkage studies in African populations to identify loci that may not be detectable in non-Africans, translate raw GRSs into estimates of absolute risk (48), and identify the causal alleles in genomic regions that are associated with prostate cancer (49, 50). Future studies will increase our understanding of health disparities and the genetics of prostate cancer as additional susceptibility loci are found in Africa.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

Conception and design: J. Lachance, S.A. Tishkoff, T.R. Rebbeck

Development of methodology: J. Lachance, T.R. Rebbeck

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S.A. Tishkoff, T.R. Rebbeck

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J. Lachance, A.J. Berens, M.E.B. Hansen, A.K. Teng, T.R. Rebbeck

Writing, review, and/or revision of the manuscript: J. Lachance, S.A. Tishkoff, T.R. Rebbeck

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): T.R. Rebbeck

Study supervision: J. Lachance, S.A. Tishkoff, T.R. Rebbeck

Acknowledgments

J. Lachance and T.R. Rebbeck are supported in part by NIH/NCI grant U01CA184374. M.E.B. Hansen was funded by a CEET training grant (T32ES019851). This work was also supported by NIH grant DP1ES022577 to S.A. Tishkoff. J. Lachance and T.R. Rebbeck are members of the MADCaP network (Men of African Descent, Carcinoma of the Prostate, https://www.madcapnetwork.org/). We thank C. Hanson, M. Harlemon, M.S. Kim, and U. Martinez-Marigorta for helpful discussions during the writing of the article.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 26, 2017; revised November 2, 2017; accepted February 5, 2018; published first February 8, 2018.
Lachance et al.

References


Genetic Hitchhiking and Population Bottlenecks Contribute to Prostate Cancer Disparities in Men of African Descent


Cancer Res 2018;78:2432-2443. Published OnlineFirst February 8, 2018.

Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-17-1550

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2018/02/08/0008-5472.CAN-17-1550.DC1

Cited articles
This article cites 46 articles, 10 of which you can access for free at:
http://cancerres.aacrjournals.org/content/78/9/2432.full#ref-list-1

Citing articles
This article has been cited by 9 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/78/9/2432.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/78/9/2432.
Click on “Request Permissions” which will take you to the Copyright Clearance Center's (CCC) Rightslink site.