

ARTICLE

Received 24 Jun 2014 | Accepted 10 Mar 2015 | Published 5 May 2015

DOI: 10.1038/ncomms7889

Two susceptibility loci identified for prostate cancer aggressiveness

Sonja I. Berndt^{1,*}, Zhaoming Wang^{1,2,*}, Meredith Yeager^{1,2}, Michael C. Alavanja¹, Demetrius Albanes¹, Laufey Amundadottir¹, Gerald Andriole³, Laura Beane Freeman¹, Daniele Campa⁴, Geraldine Cancel-Tassin⁵, Federico Canzian⁶, Jean-Nicolas Cornu¹, Olivier Cussenot⁵, W. Ryan Diver⁷, Susan M. Gapstur⁷, Henrik Grönberg⁸, Christopher A. Haiman⁹, Brian Henderson⁹, Amy Hutchinson², David J. Hunter¹⁰, Timothy J. Key¹¹, Suzanne Kolb¹², Stella Koutros¹, Peter Kraft¹⁰, Loic Le Marchand¹³, Sara Lindström¹⁰, Mitchell J. Machiela¹, Elaine A. Ostrander¹⁴, Elio Riboli¹⁵, Fred Schumacher⁹, Afshan Siddiq¹⁶, Janet L. Stanford^{12,17}, Victoria L. Stevens⁷, Ruth C. Travis¹¹, Konstantinos K. Tsilidis¹⁸, Jarmo Virtamo¹⁹, Stephanie Weinstein¹, Fredrik Wilkum⁸, Jianfeng Xu²⁰, S. Lilly Zheng²⁰, Kai Yu¹, William Wheeler²¹, Han Zhang¹, African Ancestry Prostate Cancer GWAS Consortium†, Joshua Sampson¹, Amanda Black¹, Kevin Jacobs¹, Robert N. Hoover¹, Margaret Tucker¹ & Stephen J. Chanock¹

Most men diagnosed with prostate cancer will experience indolent disease; hence, discovering genetic variants that distinguish aggressive from nonaggressive prostate cancer is of critical clinical importance for disease prevention and treatment. In a multistage, case-only genome-wide association study of 12,518 prostate cancer cases, we identify two loci associated with Gleason score, a pathological measure of disease aggressiveness: rs35148638 at 5q14.3 (*RASA1*, $P = 6.49 \times 10^{-9}$) and rs78943174 at 3q26.31 (*NAALADL2*, $P = 4.18 \times 10^{-8}$). In a stratified case-control analysis, the SNP at 5q14.3 appears specific for aggressive prostate cancer ($P = 8.85 \times 10^{-5}$) with no association for nonaggressive prostate cancer compared with controls ($P = 0.57$). The proximity of these loci to genes involved in vascular disease suggests potential biological mechanisms worthy of further investigation.

¹ Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892, USA. ² Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Leidos Biomedical Research Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland 21701, USA. ³ Division of Urologic Surgery, Washington University School of Medicine, St Louis, Missouri 63110, USA. ⁴ Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. ⁵ CeRePP, Assistance Publique-Hôpitaux de Paris, UPMC University Paris 6, Paris, France. ⁶ Genomic Epidemiology Group, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. ⁷ Epidemiology Research Program, American Cancer Society, Atlanta, Georgia 30303, USA. ⁸ Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm 17177, Sweden. ⁹ Department of Preventative Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California 90033, USA. ¹⁰ Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ¹¹ Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford OX3 7BN, UK. ¹² Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. ¹³ Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, USA. ¹⁴ National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹⁵ Department of Epidemiology and Biostatistics, School of Public Health, Imperial College, London SW7 2AZ, UK. ¹⁶ Department of Genomics of Common Disease, School of Public Health, Imperial College London, London SW7 2AZ, UK. ¹⁷ Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington 98195, USA. ¹⁸ Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece. ¹⁹ Department of Chronic Disease Prevention, National Institute for Health and Welfare, FI-00271 Helsinki, Finland. ²⁰ Center for Cancer Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, USA. ²¹ Information Management Services Inc., Rockville, Maryland 20852, USA. † List of members and affiliations appears at the end of the paper. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to S.I.B. (email: berndts@mail.nih.gov).

Prostate cancer is the most common cancer diagnosed in men and second leading cause of cancer death among men in the United States¹; however, little is known about why the disease progresses in some men but not in others. Determining which cancers are likely to progress and cause death is of critical clinical importance. Prostate cancer aggressiveness is thought to be partially determined by genetic factors, as studies have shown an increased risk of death from prostate cancer among offspring with a family history of fatal disease^{2,3}. The definitions of aggressive prostate cancer differ between studies; however, one important and widely used descriptor is tumour grade at diagnosis, as measured by the Gleason score, which ranks pathological changes, namely, tumour differentiation, and has been associated with disease progression and survival⁴. Linkage studies using the Gleason score as a measure of aggressiveness have implicated several chromosomal regions, including 1p, 5q, 6q, 7q and 19q; however, no specific genetic mutations have been conclusively identified^{5–8}. Although previous genetic association studies have identified or suggested markers for aggressive prostate cancer^{9–11}, these single-nucleotide polymorphisms (SNPs) have either also been associated with nonaggressive disease, making them nonspecific, or have not been convincingly replicated¹².

Genome-wide association studies (GWAS) have successfully identified roughly 100 loci associated with prostate cancer risk^{11–26}; however, most loci have minor allele frequencies (MAFs) >10% and, so far, none conclusively differentiate aggressive from nonaggressive disease. To discover additional loci associated with risk and to identify loci specific for aggressive prostate cancer, here we conduct a multistage GWAS for prostate cancer among men of European ancestry using the Illumina HumanOmni2.5 Beadchip, which provides greater coverage of uncommon SNPs for individuals of European ancestry than microarrays used in previous GWAS of prostate cancer. We identify two novel loci associated with the Gleason score, a pathological measure of prostate cancer aggressiveness, that are located near genes involved in vascular development and maintenance.

Results

Case-control association results for prostate cancer. A total of 4,600 cases and 2,941 controls of European ancestry from the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial were genotyped using the Illumina HumanOmni2.5 Beadchip and passed rigorous quality-control criteria (see Methods). Baseline characteristics of the cases and controls are shown in Supplementary Table 1. On the basis of a linear regression model, men with higher Gleason scores were more likely to be diagnosed at an older age ($P < 0.001$). Of the SNPs genotyped, 1,531,807 passed quality-control filters with a minimum call rate of 94%. Genotypes were analysed using regression models, assuming a log-additive genetic model and adjusting for age and significant eigenvectors. A quantile–quantile (Q–Q) plot of the P values for prostate cancer risk on the basis of logistic regression models showed enrichment of small P values compared with the null distribution, even after removing SNPs within 500 kb of the previously published loci (Supplementary Fig. 1, $\lambda = 1.007$). Fifty-six of the previously published loci were nominally associated with risk ($P < 0.05$) in stage 1 (Supplementary Table 2), and two previously published loci at chromosomes 8q24 and 17q12 reached genome-wide significance in stage 1 ($P < 5 \times 10^{-8}$, Supplementary Fig. 2). Rare variant analysis using SKAT²⁷ for SNPs with MAF < 2% revealed five gene regions with $P < 5 \times 10^{-8}$; however, all appeared to be artefacts driven by poorly clustered SNPs.

Sixteen promising SNPs with $P < 2 \times 10^{-5}$ on the basis of the logistic regression models were taken forward for Taqman replication in 5,139 cases and 5,591 controls from seven studies (see Methods); however, none replicated for prostate cancer overall (Supplementary Table 3). A more extensive replication was undertaken using a custom Illumina iSelect microarray comprising 51,207 SNPs selected for prostate cancer, 10,458 SNPs for other phenotypes (for example, obesity) and 1,435 candidate SNPs (see Methods). In stage 2, a total of 6,575 cases and 6,392 controls from five studies were genotyped with the custom iSelect and passed quality-control criteria (Supplementary Table 1). *In silico* data were also available for stage 3 for 1,204 nonoverlapping cases and 1,231 controls from a previous GWAS of advanced (defined as Gleason score ≥ 8 or stage C/D) prostate cancer¹², giving a total of 12,379 cases and 10,564 controls. As not all SNPs included on the iSelect were directly genotyped in stage 1 or stage 3, both scans were imputed using the 1000 Genomes Project release version 3 (ref. 28) and IMPUTE2 (ref. 29).

In a combined meta-analysis of the primary scan together with the custom SNP microarray replication and *in silico* look-up in a previous GWAS, 13 loci reached genome-wide significance ($P < 5 \times 10^{-8}$); however, each of them confirmed a previously reported locus^{13–20,23} (Supplementary Table 4). Although not reaching genome-wide significance, two new suggestive loci at chromosome 16q22.2 (*PKD1L3*, rs12597458, $P = 9.67 \times 10^{-8}$) and 6p22.3 (*CDKAL1*, rs12198220, $P = 2.13 \times 10^{-7}$) were also identified (Supplementary Table 5, Supplementary Fig. 3). Further studies are needed to confirm these findings.

Case-only results of disease aggressiveness. To evaluate disease aggressiveness, we modelled the Gleason score as a quantitative trait among the prostate cancer cases ($n = 4,545$) included in stage 1 in a case-only analysis using linear regression. We chose to model the Gleason score as a quantitative trait as opposed to a dichotomous outcome in order to maximize our statistical power to detect variants that differentiate aggressive from nonaggressive disease. In stage 1, the Q–Q plot of the association P values revealed a small number of SNPs with P values less than expected under the null distribution (Supplementary Fig. 4, $\lambda = 0.998$). We evaluated the SNPs previously reported to be associated with the risk of aggressive disease; however, none were significantly associated with the Gleason score among cases (Supplementary Table 6). As part of the custom SNP microarray replication, SNPs with a P value < 0.001 from the linear regression model of the Gleason score as a quantitative trait, filtered using $r^2 > 0.8$, were taken forward for the custom SNP microarray replication in 5,355 cases with the Gleason score from five studies (stage 2). One novel locus at chromosome 5q14.3 (rs35148638) reached genome-wide significance in the meta-analysis of stages 1 and 2. Five SNPs, including two moderately correlated SNPs at chromosome 5q14.3, with P values $< 2 \times 10^{-6}$, were taken forward for replication in 2,618 cases from the Cancer of the Prostate in Sweden study. In the meta-analysis of the Gleason score results for all three stages including a total of 12,518 cases, three SNPs reached genome-wide significance: rs35148638 at 5q14.3 ($P = 6.49 \times 10^{-9}$), rs62113212 at 19q13.33 ($P = 5.85 \times 10^{-9}$) and rs78943174 at 3q26.31 ($P = 4.18 \times 10^{-8}$; Table 1). The SNPs at chromosome 5q14.3 and 3q26.31 represent novel loci (Fig. 1), whereas the chromosome 19q13.33 locus has been previously identified to be associated with prostate cancer risk overall¹⁸.

Stratified case-control association results for novel loci. To evaluate the extent to which these three loci (identified from our case-only study of the Gleason score) could be associated with aggressive prostate cancer risk, we conducted a case-control

Table 1 | Loci associated with prostate cancer aggressiveness as measured by the Gleason score among cases only*.

SNP	Nearest gene(s)	Chr	Position	Risk allele [†]	Other allele	Stage	RAF [‡]	No. of subjects	Beta	SE	P-value
<i>Novel loci</i>											
rs35148638	RASA1, CCNH	5q14.3	86610989	C	A	Stage 1	0.25	4,545	0.085	0.024	3.09×10^{-4}
						Stage 2	0.255	5,353	0.125	0.022	1.86×10^{-6}
						Stage 3	0.262	2,618	-0.025	0.040	0.50
						Combined		12,516	0.088	0.015	6.49×10^{-9}
rs78943174	NAALADL2	3q26.31	175252736	C	T	Stage 1	0.985	4,545	0.361	0.085	2.44×10^{-5}
						Stage 2	0.987	5,336	0.217	0.085	0.03
						Stage 3	0.979	2,604	0.293	0.110	0.008
						Combined		12,485	0.289	0.053	4.18×10^{-8}
<i>Previously reported loci</i>											
rs62113212	KLK3	19q13.33	51360840	T	C	Stage 1	0.063	4,544	0.177	0.043	4.00×10^{-5}
						Stage 2	0.061	5,348	0.124	0.041	0.009
						Stage 3	0.107	2,616	0.119	0.040	0.004
						Combined		12,508	0.138	0.024	5.85×10^{-9}

*Results from case-only analysis of the Gleason score as a quantitative trait using linear regression.

[†]Risk allele is the allele associated with an increased risk of aggressive disease (that is, higher Gleason score).

[‡]Frequency of the risk allele.

analysis stratified by the overall Gleason score (that is, scores of 2–10), recognizing that our power to detect an association at genome-wide significance would be reduced. We did not have data on the individual components of Gleason 7 from most studies in order to subclassify them as 3 + 4 or 4 + 3, and so in order to clearly differentiate between aggressive and non-aggressive disease, we stratified our cases by those with Gleason scores ≤ 6 (nonaggressive) and those with Gleason scores ≥ 8 (aggressive). Although it did not reach genome-wide significance, rs35148638 at 5q14.3 showed an increased risk with aggressive prostate cancer ($P = 8.85 \times 10^{-5}$; Supplementary Table 7). There was no association for nonaggressive disease ($P = 0.57$) and the P value for heterogeneity between the two outcomes was modestly significant ($P = 2.89 \times 10^{-4}$). As rs78943174 at 3q26.31 was not common among controls with a MAF of 1–2%, we had limited power to detect an association with aggressive disease; however, we did observe a marginal positive association between rs78943174 and aggressive prostate cancer risk ($P = 0.07$) and the P value for heterogeneity between aggressive and nonaggressive prostate cancer risk was nominally significant ($P = 0.006$). Consistent with previous studies³⁰, the SNP at 19q13.33 was strongly associated with nonaggressive prostate cancer ($P = 3.51 \times 10^{-13}$) with a weak association in the opposite direction for aggressive prostate cancer ($P = 0.01$) and highly significant P value for heterogeneity ($P = 1.44 \times 10^{-10}$).

As African Americans have an elevated risk of prostate cancer, we evaluated the extent to which these three SNPs were associated with aggressive prostate cancer risk in African Americans using data from the African American Prostate Cancer GWAS Consortium (see Methods). In this smaller study, none of the SNPs were significantly associated with the risk of aggressive disease (Supplementary Table 8), and only rs62113212 at chromosome 19q13.33 was nominally associated with non-aggressive disease ($P = 0.04$). However, the direction of the effects for African Americans for the SNPs at 3q26.31 and 19q13.33 were consistent with what we observed among Europeans.

Further examination of novel loci. Examination of the three identified loci for the Gleason score using data from ENCODE revealed significant DNase enrichment in lymphoblastoid and embryonic myoblast cells and evidence for altered motifs (Supplementary Table 9). Rs62113212 at 19q13.33 is in strong linkage disequilibrium with a missense SNP (rs17632542, $r^2 = 1$).

No significant expression quantitative trait locus (eQTL) associations were observed using data from the Genotype-Tissue Expression (GTEx) Project³¹; however, for a proxy of rs35148638 at 5q14.3 (rs4421140, $r^2 = 0.82$), we did find nominally significant eqtl associations with *RASA1* and *CCNH* expression and meqtl associations with CpG sites in *RASA1* and *CCNH* in adipose tissue^{32,33}.

Discussion

Linkage studies of prostate cancer aggressiveness have reported suggestive evidence of linkage to chromosome 5q (refs 5–8) and specifically 5q14 in TMPRSS2-ERG fusion-positive families³⁴. The 5q14.3 SNP identified in this study (rs35148638), associated with disease aggressiveness, is intronic to the RAS p21 protein activator 1 (*RASA1*) gene, which suppresses RAS function, helps regulate cellular proliferation and differentiation³⁵, and controls blood vessel growth³⁶. Rare mutations in *RASA1* lead to capillary malformation-arteriovenous malformation and Parkes-Weber syndrome³⁷ as well as lymphatic abnormalities³⁸, providing an intriguing plausibility for the gene in aggressive prostate cancer. The SNP is also ~ 79 kb downstream of the cyclin H (*CCNH*) gene, which encodes a regulatory component of a cyclin-dependent (CDK)-activating kinase necessary for RNA polymerase II transcription, nucleotide excision repair and p53 phosphorylation³⁹. *CCNH* has been shown to be differentially expressed between androgen-sensitive and androgen-resistant prostate cancer cell lines^{40,41}, suggesting a role in prostate cancer progression.

The 3q26.31 SNP (rs78943174) is intronic to the N-acetylated alpha-linked acidic dipeptidase-like2 (*NAALADL2*) gene, which is part of the glutamate carboxypeptidase II family. This gene is also related to prostate-specific membrane antigen, a well-characterized diagnostic indicator and potential drug target of prostate cancer⁴². *NAALADL2* has been shown to promote a pro-migratory and pro-metastatic microenvironment, and higher tumour expression of *NAALADL2* is associated with higher Gleason score and poor survival following radical prostatectomy⁴³. Variants in *NAALADL2* have also been identified to be associated with Kawasaki disease⁴⁴, a paediatric, autoimmune vascular disease. The SNP is also ~ 117 kb telomeric of the microRNA, MIR4789, which is predicted to target several genes involved in the insulin resistance (for example, *IRS1*, *PIK3R1*) among others^{45,46}. A previous GWAS of prostate cancer

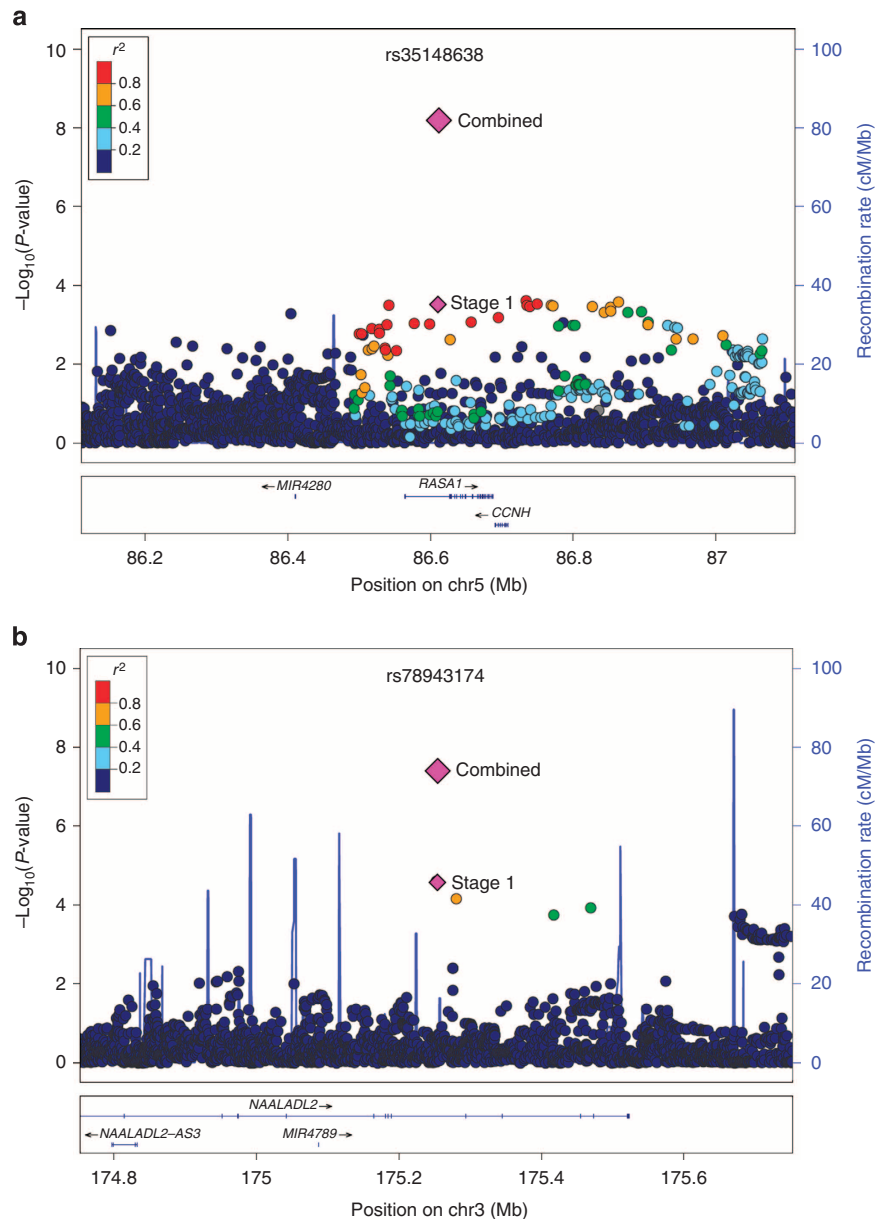


Figure 1 | Regional association plots of the two novel loci associated with the Gleason score as a quantitative trait among cases. (a) Chromosome 5q14.3 (rs35148638) and **(b)** chromosome 3q26.31 (rs78943174). Shown are the $-\log_{10}$ association P values from the linear regression model for the 4,545 cases in stage 1 (dots and lower purple diamond) and $-\log_{10}$ P values from the linear regression model for the 12,518 cases in the combined stage 1–3 analysis (upper diamond).

reported a suggestive association with a SNP at 3q26.31 (ref. 47); however, this SNP is not in linkage disequilibrium with the SNP identified in our study ($r^2 = 0.003$).

The chromosome 19q13.33 (*KLK3*) locus has been previously associated with prostate cancer risk overall¹⁸. Although several studies have suggested that the risk may differ by disease aggressiveness^{30,48–50}, this study shows for the first time a genome-wide significant difference between aggressive and nonaggressive disease as measured by the Gleason score. *KLK3* encodes the prostate-specific antigen (PSA) protein. The C allele of rs62113212 has been shown to be associated with higher PSA levels⁴⁸, suggesting that the association observed with the SNP is related to early prostate cancer detection.

Although one of our goals was to identify uncommon variants for prostate cancer, we did not identify any new independent SNPs with a MAF < 10%. We did, however, identify a suggestive

locus at chromosome 6p22.3 (rs12198220), which is 98 kb downstream of *CDKALI*. A pooled linkage study of prostate cancer previously reported suggestive evidence of linkage to this region⁵¹. Interestingly, SNPs at this locus have been associated with the risk of type 2 diabetes, adding to the list of susceptibility regions shared between prostate cancer and type 2 diabetes⁵². We also discovered a new suggestive locus at 16q22.2, which is in strong linkage disequilibrium with a missense variant (rs3213422, $r^2 = 0.74$) in dihydro-orotate dehydrogenase (quinone) gene (*DHODH*), which encodes an enzyme necessary for the biosynthesis of pyrimidines and cell proliferation. Further studies are needed to confirm these suggestive loci.

In this study, we used the Gleason score to differentiate between nonaggressive and aggressive prostate cancer. Gleason score is a powerful prognostic factor and predictor of disease behaviour; however, substantial changes in Gleason scoring have

changed since it was first proposed over 40 years ago, resulting in shifts towards higher scores⁵³. In addition, differences in scoring between pathologists remain⁵⁴. Whether these changes in Gleason scoring ultimately result in better outcome prediction and classification of disease from an aetiological standpoint remains to be seen. Unlike for breast cancer where classification by receptor status has resulted in significant advances in the aetiological understanding of the disease, clearly defining aggressive prostate cancer remains difficult. Regardless, the Gleason score is an important component of prostate cancer risk assessment and is the most commonly used tool for assessing prostate cancer aggressiveness.

In conclusion, we identified two new loci associated with prostate cancer aggressiveness as measured by the Gleason score in a case-only study of prostate cancer. Although additional studies are needed to confirm these findings and reveal the underlying biological mechanism, the proximity of these SNPs to genes involved in vascular disease, cell migration and metastasis makes them intriguing loci for further study.

Methods

Stage 1: discovery population and genotyping. A new GWAS was conducted in prostate cancer cases and controls of European ancestry from the PLCO Cancer Screening Trial. PLCO is a randomized trial for the early detection of prostate, lung, colorectal and ovarian cancers⁵⁵. In brief, 76,693 men were enrolled in the trial from 10 centres in the United States from 1993 to 2001 and randomized to receive annual screening with PSA for 6 years and digital rectal examination for 4 years or referred to their physician for routine care. Men with positive screening results were referred to their primary physician for further evaluation. All prostate cancer cases detected during screening or reported during the trial were pathologically confirmed, and information on stage and grade was abstracted from medical records. Blood or buccal cells were collected from participants in the trial⁵⁶. The study was approved by the institutional review board at each centre and National Cancer Institute (NCI); all study participants provided informed written consent.

A total of 4,838 prostate cancer cases and 3,053 controls of European ancestry, matched on age and year of randomization, were selected for stage 1. The sample size was chosen on the basis of statistical power estimates for detecting a modest association in a multistage GWAS. Including quality-control duplicates, 8,222 samples were genotyped on the Illumina HumanOmni2.5 Beadchip. Extensive quality-control metrics were employed to ensure that only high-quality genotype data were analysed using the GLU software package. Samples with a missing rate > 6% ($n = 323$) or heterozygosity < 16% or > 21% ($n = 7$) were excluded, and 221 samples were removed because of technical issues. Gender discordance on the basis of chromosome X heterozygosity was evaluated; however, no subjects were removed. One unexpected duplicate (> 99.9% concordance) and 28 full sibling pairs on the basis of an identity-by-descent threshold of 0.70 were detected and one subject from each pair was removed ($n = 29$). Ancestry was estimated using a set of population informative markers⁵⁷ and the GLU struct.admix module, which is similar to the method proposed in (ref. 58). Five subjects (three cases and two controls) were determined to have < 80% European ancestry and were removed from analysis (Supplementary Fig. 5). Principal components analysis was performed to evaluate population substructure in greater detail (Supplementary Fig. 6), and two significant eigenvectors ($P < 0.05$) were included in the analytic model. Expected duplicates yielded 99.9% concordance. SNPs without genotype calls, a completion rate < 94%, Hardy-Weinberg proportion test P value < 1×10^{-8} or MAF < 1% were excluded, leaving 1,531,807 SNPs for analysis. After quality-control exclusions, 4,600 cases and 2,840 controls remained. An additional 101 male controls from PLCO Cancer Screening Trial, genotyped previously on the HumanOmni2.5 (ref. 59) were also included, resulting in 4,600 cases and 2,941 controls for the primary prostate cancer analysis (Supplementary Table 1 and Supplementary Fig. 7). Of those cases, 4,545 men had information on the Gleason score available. Regression models were fit adjusting for significant principal components and age. Sixteen different models were fitted for prostate cancer-related outcomes, including overall prostate cancer risk and Gleason score.

Stage 2: follow-up studies and genotyping. Replication was conducted using a set of independent prostate cancer cases and controls from five studies: Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study⁶⁰ (ATBC, $n = 1,092$ cases/1,099 controls), Cancer Prevention Study II (ref. 61; CPSII, $n = 2,770$ cases/2,669 controls), Health Professional Follow-up Study⁶² ($n = 963$ cases/1,047 controls), French Prostate Cancer Case-Control Study ($n = 1,494$ cases/1,546 controls) and PLCO⁵⁵ ($n = 990$ cases/922 controls). Including quality-control duplicates, 14,592 samples were genotyped using a custom Illumina iSelect microarray comprising 51,207 SNPs selected for prostate cancer, 10,458 SNPs for other phenotypes (for

example, smoking and obesity) and 1,435 candidate SNPs. The SNPs for prostate cancer were filtered using $r^2 < 0.7$ and selected on the basis of the most significant results from 16 prostate cancer models, with the primary models being an overall prostate cancer risk model assuming a log-additive effect for each SNP and case-only Gleason score model, where Gleason was modelled as a quantitative linear trait among cases and each SNP was assumed to have an additive effect. SNPs with P values < 0.05 and < 0.001 from each model, respectively, were advanced for possible replication.

Similar to stage 1, samples genotyped in stage 2 underwent rigorous quality-control procedures. Samples with missing rate > 10% ($n = 1,158$) or mean heterozygosity < 20% or > 26% ($n = 4$) were excluded. In addition, 21 subjects without phenotype information were removed. Fifteen unexpected duplicates with concordance rates > 99.9% were observed, and 25 first-degree relative pairs were detected assuming an identity-by-descent threshold of 0.7. For each unexpected duplicate and relative pair, one subject was removed. Using the GLU struct.admix module, ancestry was estimated on the basis of genotyped SNPs with a MAF > 10% and HapMap data as the fixed reference population. Sixty-six subjects with < 80% European ancestry were removed from the analysis. Principal components analysis was conducted using a set of SNPs selected for traits unrelated to prostate cancer (for example, smoking and alcohol intake). After quality-control exclusions, a total of 6,575 cases and 6,392 controls remained for the primary analysis (Supplementary Table 1), including 5,355 cases with the Gleason score. SNPs with a MAF < 1% or completion rate < 90% were excluded from the analysis, leaving 55,497 SNPs for analysis (Supplementary Fig. 7). Regression models were fit adjusting for age, significant principal components and study.

In addition to the custom SNP microarray replication, 16 promising SNPs ($P < 2 \times 10^{-5}$) were taken forward for fast-track replication in the five studies listed above ($n = 2,495$ cases/2,532 controls) as well as three additional studies: Agricultural Health Study⁶³ ($n = 579$ cases/1,172 controls), Fred Hutchinson Cancer Research Center ($n = 1,315$ cases/1,152 controls) and the Multiethnic Cohort⁶⁴ ($n = 750$ cases/735 controls). In total, 5,139 cases and 5,591 controls, all of European ancestry, were genotyped (Supplementary Fig. 7). The SNPs were genotyped using individual TaqMan assays (Applied Biosystems Inc) and quality-control duplicates yielded > 99.9% concordance.

Stage 3a: *in silico* replication of prostate cancer findings. For replication of the overall prostate cancer results, nonoverlapping *in silico* GWAS data were available from 1,204 cases and 1,231 controls of European ancestry from four studies from a previous GWAS of advanced prostate cancer¹²: European Prospective Investigation into Cancer and Nutrition (EPIC; 431 cases/426 controls)⁶⁵, Multiethnic Cohort (244 cases/259 controls)⁶⁴, Physicians Health Study (PHS; 298 cases/255 controls) and American Cancer Society Cancer Prevention Study II (CPSII; 231 cases/291 controls not included in stage 2)⁶¹ (Supplementary Fig. 7). Subjects were genotyped using the Illumina HumanHap610K and extensive quality-control filters applied as described previously. All data were imputed using IMPUTE2 (ref. 29) and 1000 Genomes Project release version 3 (ref. 28) as the reference panel, and data analysed using SNPTEST assuming a log-additive genetic model and adjusting for age, study and significant principal components. Only SNPs with an information score > 0.3 were included in the meta-analysis.

Stage 3b: additional replication for Gleason score findings. For further replication of the results for the Gleason score, we genotyped five of the most significant SNPs ($P < 2 \times 10^{-6}$) in the Cancer of the Prostate in Sweden, a population-based case-control study of 2,618 cases and 1,728 controls using Sequenom (Supplementary Fig. 7). Regression models were fit adjusting for age.

Meta-analysis. Data from all three stages were meta-analysed using the fixed effects inverse variance method based on the beta estimates and s.e.'s from each stage.

Further follow-up analyses. To evaluate the associations observed in our study of men of European ancestry with those observed in African Americans, we obtained association results for three genome-wide significant hits from the African American Prostate Cancer GWAS Consortium²². Although it was not possible to evaluate the Gleason score as a quantitative trait among cases in this consortium, we were able to obtain stratified association results for cases with Gleason ≤ 6 versus controls and cases with Gleason ≥ 8 versus controls.

Using 1000 Genomes Project data, we identified SNPs with $r^2 > 0.8$ with the lead SNPs identified to be associated with Gleason score and evaluated whether they were nonsynonymous coding variants. We utilized HaploReg⁶⁶ to assess noncoding functional markers in the regions containing our lead SNPs and related proxy SNPs ($r^2 > 0.8$; Supplementary Table 9). We explored possible *cis* eQTL associations with our lead SNPs and related proxy SNPs ($r^2 > 0.8$) in adipose tissue, lymphoblastoid cell lines and skin using data from the MuTHER resource³² and all available tissues, including whole blood, in the GTEx³¹. We also examined possible methylation eQTL associations in adipose tissue using the MuTHER resource³³.

References

- Siegel, R., Ma, J., Zou, Z. & Jemal, A. Cancer statistics, 2014. *CA Cancer J. Clin.* **64**, 9–29 (2014).
- Lindstrom, L. S. *et al.* Familial concordance in cancer survival: a Swedish population-based study. *Lancet Oncol.* **8**, 1001–1006 (2007).
- Hemminki, K. Familial risk and familial survival in prostate cancer. *World J. Urol.* **30**, 143–148 (2012).
- Gleason, D. F. Histologic grading of prostate cancer: a perspective. *Hum. Pathol.* **23**, 273–279 (1992).
- Witte, J. S. *et al.* Genomewide scan for prostate cancer-aggressiveness loci. *Am. J. Hum. Genet.* **67**, 92–99 (2000).
- Witte, J. S. *et al.* Genome-wide scan of brothers: replication and fine mapping of prostate cancer susceptibility and aggressiveness loci. *Prostate* **57**, 298–308 (2003).
- Slager, S. L. *et al.* Genome-wide linkage scan for prostate cancer aggressiveness loci using families from the University of Michigan Prostate Cancer Genetics Project. *Prostate* **66**, 173–179 (2006).
- Schaid, D. J. *et al.* Genome-wide linkage scan of prostate cancer Gleason score and confirmation of chromosome 19q. *Hum. Genet.* **121**, 729–735 (2007).
- Sun, J. *et al.* Sequence variants at 22q13 are associated with prostate cancer risk. *Cancer Res.* **69**, 10–15 (2009).
- Xu, J. *et al.* Inherited genetic variant predisposes to aggressive but not indolent prostate cancer. *Proc. Natl Acad. Sci. USA* **107**, 2136–2140 (2010).
- Al Olama, A. A. *et al.* A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease. *Hum. Mol. Genet.* **22**, 408–415 (2013).
- Schumacher, F. R. *et al.* Genome-wide association study identifies new prostate cancer susceptibility loci. *Hum. Mol. Genet.* **20**, 3867–3875 (2011).
- Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **39**, 645–649 (2007).
- Gudmundsson, J. *et al.* Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet.* **39**, 631–637 (2007).
- Gudmundsson, J. *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat. Genet.* **39**, 977–983 (2007).
- Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* **40**, 310–315 (2008).
- Gudmundsson, J. *et al.* Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat. Genet.* **40**, 281–283 (2008).
- Eeles, R. A. *et al.* Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.* **40**, 316–321 (2008).
- Gudmundsson, J. *et al.* Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat. Genet.* **41**, 1122–1126 (2009).
- Eeles, R. A. *et al.* Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat. Genet.* **41**, 1116–1121 (2009).
- Takata, R. *et al.* Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat. Genet.* **42**, 751–754 (2010).
- Haiman, C. A. *et al.* Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. *Nat. Genet.* **43**, 570–573 (2011).
- Kote-Jarai, Z. *et al.* Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat. Genet.* **43**, 785–791 (2011).
- Xu, J. *et al.* Genome-wide association study in Chinese men identifies two new prostate cancer risk loci at 9q31.2 and 19q13.4. *Nat. Genet.* **44**, 1231–1235 (2012).
- Eeles, R. A. *et al.* Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat. Genet.* **45**, 385–391 (2013).
- Al Olama, A. A. *et al.* A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat. Genet.* **46**, 1103–1109 (2014).
- Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89**, 82–93 (2011).
- Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
- Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
- Lindstrom, S. *et al.* Characterizing associations and SNP-environment interactions for GWAS-identified prostate cancer risk markers—results from BPC3. *PLoS ONE* **6**, e17142 (2011).
- GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
- Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).
- Grundberg, E. *et al.* Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am. J. Hum. Genet.* **93**, 876–890 (2013).
- Luedeke, M. *et al.* Predisposition for TMPRSS2-ERG fusion in prostate cancer by variants in DNA repair genes. *Cancer Epidemiol. Biomarkers Prev.* **18**, 3030–3035 (2009).
- McCormick, F. ras GTPase activating protein: signal transmitter and signal terminator. *Cell* **56**, 5–8 (1989).
- Henkemeyer, M. *et al.* Vascular system defects and neuronal apoptosis in mice lacking ras GTPase-activating protein. *Nature* **377**, 695–701 (1995).
- Eerola, I. *et al.* Capillary malformation-arteriovenous malformation, a new clinical and genetic disorder caused by RASA1 mutations. *Am. J. Hum. Genet.* **73**, 1240–1249 (2003).
- Burrows, P. E. *et al.* Lymphatic abnormalities are associated with RASA1 gene mutations in mouse and man. *Proc. Natl Acad. Sci. USA* **110**, 8621–8626 (2013).
- Lolli, G. & Johnson, L. N. CAK-cyclin-dependent activating kinase: a key kinase in cell cycle control and a target for drugs? *Cell Cycle* **4**, 572–577 (2005).
- Karan, D., Kelly, D. L., Rizzino, A., Lin, M. F. & Batra, S. K. Expression profile of differentially-regulated genes during progression of androgen-independent growth in human prostate cancer cells. *Carcinogenesis* **23**, 967–975 (2002).
- Romanuik, T. L. *et al.* LNCaP Atlas: gene expression associated with *in vivo* progression to castration-recurrent prostate cancer. *BMC Med. Genomics* **3**, 43 (2010).
- Ristau, B. T., O’Keefe, D. S. & Bacich, D. J. The prostate-specific membrane antigen: lessons and current clinical implications from 20 years of research. *Urol. Oncol.* **32**, 272–279 (2013).
- Whitaker, H. C. *et al.* N-acetyl-L-aspartyl-L-glutamate peptidase-like 2 is overexpressed in cancer and promotes a pro-migratory and pro-metastatic phenotype. *Oncogene* **33**, 5274–5287 (2013).
- Burgner, D. *et al.* A genome-wide association study identifies novel and functionally related susceptibility Loci for Kawasaki disease. *PLoS Genet.* **5**, e1000319 (2009).
- Wang, X. miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA* **14**, 1012–1017 (2008).
- Wang, X. & El Naqa, I. M. Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics* **24**, 325–332 (2008).
- FitzGerald, L. M. *et al.* Genome-wide association study identifies a genetic variant associated with risk for more aggressive prostate cancer. *Cancer Epidemiol. Biomarkers Prev.* **20**, 1196–1203 (2011).
- Parikh, H. *et al.* Fine mapping the KLK3 locus on chromosome 19q13.33 associated with prostate cancer susceptibility and PSA levels. *Hum. Genet.* **129**, 675–685 (2011).
- Kote-Jarai, Z. *et al.* Identification of a novel prostate cancer susceptibility variant in the KLK3 gene transcript. *Hum. Genet.* **129**, 687–694 (2011).
- Bensen, J. T. *et al.* Genetic polymorphism and prostate cancer aggressiveness: a case-only study of 1,536 GWAS and candidate SNPs in African-Americans and European-Americans. *Prostate* **73**, 11–22 (2013).
- Schaid, D. J. *et al.* Pooled genome linkage scan of aggressive prostate cancer: results from the International Consortium for Prostate Cancer Genetics. *Hum. Genet.* **120**, 471–485 (2006).
- Machiela, M. J. *et al.* Association of type 2 diabetes susceptibility variants with advanced prostate cancer risk in the Breast and Prostate Cancer Cohort Consortium. *Am. J. Epidemiol.* **176**, 1121–1129 (2012).
- Gilliland, F. D. *et al.* Trends in Gleason score for prostate cancer diagnosed between 1983 and 1993. *J. Urol.* **165**, 846–850 (2001).
- Allsbrook, Jr. W. C. *et al.* Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists. *Hum. Pathol.* **32**, 74–80 (2001).
- Andriole, G. L. *et al.* Mortality results from a randomized prostate-cancer screening trial. *N. Engl. J. Med.* **360**, 1310–1319 (2009).
- Hayes, R. B. *et al.* Etiologic and early marker studies in the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Control Clin. Trials* **21**, 349S–355S (2000).
- Yu, K. *et al.* Population substructure and control selection in genome-wide association studies. *PLoS ONE* **3**, e2551 (2008).
- Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
- Wang, Z. *et al.* Improved imputation of common and uncommon SNPs with a new reference set. *Nat. Genet.* **44**, 6–7 (2012).
- The ATBC Cancer Prevention Study Group. The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. *Ann. Epidemiol.* **4**, 1–10 (1994).
- Calle, E. E. *et al.* The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer* **94**, 2490–2501 (2002).
- Giovannucci, E. *et al.* A prospective study of dietary fat and risk of prostate cancer. *J. Natl Cancer Inst.* **85**, 1571–1579 (1993).
- Alavanja, M. C. *et al.* The Agricultural Health Study. *Environ. Health Perspect.* **104**, 362–369 (1996).
- Kolonel, L. N. *et al.* A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am. J. Epidemiol.* **151**, 346–357 (2000).

65. Riboli, E. *et al.* European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr.* **5**, 1113–1124 (2002).
66. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).

Acknowledgements

This work was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH (Z01-CP010119, ZIA-CP010152-11) and in part by the National Institute of Environmental Health Sciences (Z01-ES049030). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services nor does mention of trade names, commercial products or organization indicate endorsement by the US Government. We thank Drs Christine Berg and Philip Prorok, Division of Cancer Prevention, NCI, the screening center investigators and staff of the PLCO Cancer Screening Trial, Mr Thomas Riley and staff at Information Management Services Inc. and Ms Barbara O'Brien and staff at Westat Inc. for their contributions to the PLCO. Finally, we are grateful to the study participants for donating their time and making this study possible.

Author contributions

S.I.B., Z.W., M.Y., L.A., J.S., A.B., K.J., R.N.H., M.T. and S.J.C. organized and designed the study. Z.W., A.H., M.Y. and S.J.C. conducted/supervised the genotyping of the samples.

S.I.B., Z.W., M.Y., M.M., J.S., H.Z., W.W., K.Y., K.J., M.T. and S.J.C. contributed to the design and execution of the statistical and/or bioinformatic analysis. S.I.B. and S.J.C. drafted the manuscript. S.I.B., M.C.A., D.A., G.A., L.B.F., D.C., G.C.-T., F.C., J.-N.C., O.C., W.R.D., S.M.G., H.G., C.A.H., B.H., D.J.H., T.J.K., S.K., S.K., P.K., L.L.M., S.L., E.A.O., E.R., F.S., A.S., J.S., V.L.S., R.C.T., K.K.T., J.V., S.W., F.W., J.X., S.L.Z., A.B. and R.N.H. conducted the epidemiologic studies and contributed samples to the GWAS and/or follow-up genotyping. All authors contributed to the writing of the manuscript.

Additional information

Accession number: The GWAS data from stage 1 is available on dbGaP under accession number phs000882.v1.p1.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Berndt, S. I. *et al.* Two Susceptibility Loci Identified for Prostate Cancer Aggressiveness. *Nat. Commun.* 6:6889 doi: 10.1038/ncomms7889 (2015).

Sue A. Ingles⁹, Rick A. Kittles²², Sara S. Strom²³, Benjamin A. Rybicki²⁴, Barbara Nemesure²⁵, William B. Isaacs²⁶, Wei Zheng²⁷, Curtis A. Pettaway²⁸, Edward D. Yeboah^{29,30}, Yao Tettey^{29,30}, Richard B. Biritwum^{29,30}, Andrew A. Adjei^{29,30}, Evelyn Tay^{29,30}, Ann Truelove³¹, Shelley Niwa³¹, Anand P. Chokkalingam³², Esther M. John^{33,34}, Adam B. Murphy³⁵, Lisa B. Signorello^{10,36}, John Carpten³⁷, M Cristina Leske²⁵, Suh-Yuh Wu²⁵, Anslem J. M. Hennis^{25,38}, Christine Neslund-Dudas²⁴, Ann W. Hsing^{33,34}, Lisa Chu^{33,34}, Phyllis J. Goodman³⁹, Eric A. Klein⁴⁰, John S. Witte^{41,42}, Graham Casey⁹, Sam Kaggwa⁴³, Michael B. Cook¹, Daniel O. Stram⁹ & William J. Blot^{27,36}

²²University of Arizona College of Medicine and University of Arizona Cancer Center, Tucson, AZ, USA. ²³Department of Epidemiology, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA. ²⁴Department of Public Health Sciences, Henry Ford Hospital, Detroit, MI, USA. ²⁵Department of Preventive Medicine, Stony Brook University, Stony Brook, NY, USA. ²⁶James Buchanan Brady Urological Institute, Johns Hopkins Hospital and Medical Institution, Baltimore, MD, USA. ²⁷Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, USA. ²⁸Department of Urology, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA. ²⁹Korle Bu Teaching Hospital, Accra, Ghana. ³⁰University of Ghana Medical School, Accra, Ghana. ³¹Westat, Rockville, MD, USA. ³²School of Public Health, University of California, Berkeley, CA, USA. ³³Cancer Prevention Institute of California, Fremont, CA, USA. ³⁴Stanford University School of Medicine and Stanford Cancer Institute, Palo Alto, CA, USA. ³⁵Department of Urology, Northwestern University, Chicago, IL, USA. ³⁶International Epidemiology Institute, Rockville, MD, USA. ³⁷The Translational Genomics Research Institute, Phoenix, AZ, USA. ³⁸Chronic Disease Research Centre and Faculty of Medical Sciences, University of the West Indies, Bridgetown, Barbados. ³⁹SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. ⁴⁰Glickman Urological & Kidney Institute, Cleveland Clinic, Cleveland, OH, USA. ⁴¹Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA. ⁴²Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA. ⁴³Department of Surgery, Makerere University College of Health Sciences, Kampala, Uganda