

Association of self-identified race and genetic ancestry with the immunogenomic landscape of primary prostate cancer

Thiago Vidotto,¹ Eddie L. Imada,² Farzana Faisal,¹ Sanjana Murali,¹ Adrianna A. Mendes,¹ Harsimar Kaur,¹ Siqun Zheng,³ Jianfeng Xu,³ Edward M. Schaeffer,⁴ William B. Isaacs,⁵ Karen S. Sfanos,¹ Luigi Marchionni,^{2,6} and Tamara L. Lotan^{1,5,6}

¹Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. ²Department of Pathology, Weill-Cornell School of Medicine, New York, New York, USA. ³Program for Personalized Cancer Care, NorthShore University Health System, Evanston, Illinois, USA. ⁴Department of Urology, Northwestern University School of Medicine, Chicago, Illinois, USA. ⁵Department of Urology and ⁶Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

The genomic and immune landscapes of prostate cancer differ by self-identified race. However, few studies have examined the genome-wide copy number landscape and immune content of matched cohorts with genetic ancestry data and clinical outcomes. Here, we assessed prostate cancer somatic copy number alterations (sCNA) and tumor immune content of a grade-matched, surgically treated cohort of 145 self-identified Black (BL) and 145 self-identified White (WH) patients with genetic ancestry estimation. A generalized linear model adjusted with age, preoperative prostate-specific antigen (PSA), and Gleason Grade Group and filtered for germline copy number variations (gCNV) identified 143 loci where copy number varied significantly by percent African ancestry, clustering on chromosomes 6p, 10q, 11p, 12p, and 17p. Multivariable Cox regression models adjusted for age, preoperative PSA levels, and Gleason Grade Group revealed that chromosome 8q gains (including *MYC*) were significantly associated with biochemical recurrence and metastasis, independent of genetic ancestry. Finally, Treg density in BL and WH patients was significantly correlated with percent genome altered, and these findings were validated in the TCGA cohort. Taken together, our findings identify specific sCNA linked to genetic ancestry and outcome in primary prostate cancer and demonstrate that Treg infiltration varies by global sCNA burden in primary disease.

Authorship note: LM and TLL contributed equally to this work.

Conflict of interest: TLL has received research support from Roche/Ventana, DeepBio, and Myriad Genetics for other studies.

Copyright: © 2023, Vidotto et al. This is an open access article published under the terms of the Creative Commons Attribution 4.0 International License.

Submitted: June 15, 2022

Accepted: December 27, 2022

Published: December 29, 2022

Reference information: *JCI Insight*. 2023;8(3):e162409.
<https://doi.org/10.1172/jci.insight.162409>.

Introduction

In the United States, self-identified Black (BL) men are more than twice as likely to die from prostate cancer as self-identified White (WH) men (1, 2). While much of this difference is likely attributable to access to care, fewer prostate tumors from BL men have undergone immunogenomic profiling. Intriguingly, published data in relatively small cohorts suggest significant differences in both the genomic (3–16) and immune tumor microenvironment (TME) (4, 9, 17–21) landscapes of prostate tumors from BL compared with WH men, and some of these may be relevant to understanding disparities in outcome between these groups. For example, *ERG* rearrangements (11, 22–24), *PTEN* deletions (7, 11, 23, 24), and *TP53* alterations (14–16, 25) have been identified by our group and others as reproducibly less common among prostate tumors from BL compared with WH men. In parallel, other studies have identified inflammatory gene signatures that are enriched among prostate cancers arising in BL men (4, 9, 17–20, 26).

Though somatic copy number alterations (sCNA) prevail over mutations as the most common genomic alteration in prostate cancer (27, 28) and may be linked to changes in the immune TME (25, 29–32), few prior studies have systematically compared the copy number and immune landscapes of primary prostate cancers using diverse and grade-matched cohorts. Here, we leverage methylation profiling to interrogate the genome-wide copy number landscape and immune cell content of 145 primary prostate tumors from BL men and compare these with a group of tumors from WH men, all surgically treated and with long-term clinical follow-up. BL and WH patients in our cohort were frequency matched by tumor Gleason Grade

Group (GG), the widely-used system to pathologically grade primary prostate tumors. To mitigate the effects of genetic heterogeneity among self-identified racial groups, we identified sCNA associated with percent African genetic ancestry. Finally, we are the first to our knowledge to integrate genome-wide copy number profiling with immune tumor content estimation in prostate tumors from diverse cohorts.

Results

In this retrospective study, we performed Infinium MethylationEPIC profiling on prostate tumors from a previously reported (14, 24, 29) cohort of 290 patients with surgically treated primary prostate cancer and clinical follow-up (Supplemental Figure 1; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.162409DS1>). The cohort included tumors from 145 self-identified BL and 145 self-identified WH men, matched for GG at radical prostatectomy, along with 15 normal prostate tissue samples from each group. Clinical and pathologic information for the cohort is summarized in Table 1. Median follow-up for the entire cohort was 6 years, and it was 6 years for BL and 7 years for WH patients. As previously reported for this cohort (33), self-identified race was highly correlated with percent African (Yoruba) ancestry (YRI), with rare exceptions (Supplemental Figure 2).

EPIC array-derived sCNA are significantly associated with those obtained from targeted sequencing and genomic alterations detected by IHC. We leveraged the *conumee* package (34) to assess the genome-wide sCNA landscape of tumors from methylation array data. We validated our approach by comparing copy number calls obtained using *conumee* with other methods (e.g., targeted sequencing and Oncoscan SNP arrays) using correlations and correspondence at the top (CAT) plots (35, 36). This latter method allows for assessment of the agreement for genes with large \log_2 fold change, and these genes ultimately are those affected by genomic amplifications or losses. For this analysis, we analyzed 7 cases by EPIC arrays, Oncoscan SNP arrays, and targeted sequencing, which revealed a consistent agreement between copy number calls beyond chance expectations, especially for measurements obtained by EPIC and Oncoscan arrays (Supplemental Figure 3). In our patient cohort, we further tested whether the continuous gene-level sCNA data obtained via *conumee* correlated with previously published sCNA data for 101 genes assessed by targeted sequencing in a subset of our cohort composed of 134 BL patients (14). The correlation between gene-level sCNA data from sequencing and *conumee* was significant and positive for 73 of 101 (72%) genes on the targeted panel (Supplemental Figure 4 and Supplemental Table 1). To further validate the *conumee* sCNA estimations, we examined the continuous *conumee* sCNA data for *PTEN*, *ERG*, and *TP53* stratified by previously published *PTEN*, *ERG*, and p53 IHC status (24, 25) using genetically validated protocols for this cohort. As expected, *PTEN* loss, *ERG* overexpression, and p53 nuclear accumulation detected by IHC were significantly associated with decreased *PTEN*, *ERG*, and *TP53* copy number ($P < 0.0001$) (Supplemental Figure 5).

Percent genome altered (PGA) obtained independently through EPIC array and targeted sequencing are correlated. Next, we performed a segment-level sCNA analysis and derived global sCNA burden estimates expressed as PGA from the EPIC array copy number data. The methylation array-based PGA data were significantly correlated with the previously published PGA estimates obtained from targeted sequencing in a subset of 134 BL patients (Spearman's $\rho = 0.44$, $P < 0.0001$) (Supplemental Figure 6A). As a final check, we analyzed the associations between PGA and *PTEN*, *ERG*, and *TP53* alterations detected by IHC. Nuclear accumulation of p53 protein is a previously genetically validated surrogate marker of underlying *TP53* missense mutation (37, 38). PGA levels were significantly increased in samples with p53 nuclear accumulation (median = 7.09) compared with those without (median = 3.81) ($P = 0.02$) (Supplemental Figure 6, B–D). We did not observe significant differences in PGA between tumors with intact *PTEN* (median = 4.53) versus *PTEN* loss (median = 3.41) ($P = 0.74$). Similarly, tumors with *ERG* expression indicative of underlying *ERG* fusions did not have increased PGA levels (median = 3.18) compared with *ERG*⁻ samples (median = 4.65) ($P = 0.37$).

PGA is not associated with ancestry but is linked to an increased risk of metastasis. Having validated the sCNA estimates via methylation array, we next tested associations between PGA and clinical pathologic variables, including self-identified race and genetic ancestry. The segment-level copy number landscape of primary prostate tumors in BL and WH patients highlighted patterns of losses (6q, 8p, 13q, 16q) and gains (3q, chr 7, 8q) that have previously been reported in prostate cancer (9, 27, 28) (Figure 1, A and B). As expected, PGA was significantly associated with GG in the overall cohort by logistic regression analysis as well as separately for each self-identified race group (Supplemental Table 2). Median PGA was 4.13 and 3.82 for

Table 1. Clinical, pathological, and genetic features of the investigated cohort

Clinical feature	Patients self-identified as Black	Patients self-identified as White	P value
%CEU (median)	0.18	0.97	<0.0001
%YRI (median)	0.78	0	<0.0001
Age (median, range)	58 (36–73)	62 (37–73)	<0.001
PSA (median, range)	7.5 (0.3–43.5)	5.8 (0.1–39.8)	<0.001
Biochemical recurrence (count, %)	57 (39)	52 (35)	0.6
Time to BCR (years; median, range)	2.5 (1–20)	4.0 (1–19)	0.04
Metastasis (count, %)	13 (8)	17 (11)	0.5
Time to metastasis (years; median, range)	5 (1–20)	6 (1–20)	0.03
Follow-up time (years; median, range)	6 (1–20)	7 (1–22)	0.03
Clinical stage (count, %)			0.7
0	104 (71)	100 (68)	
1	23 (15)	23 (15)	
2	17 (11)	22 (15)	
Pathological stage (count, %)			0.8
T2	65 (44)	63 (43)	
T3a	46 (31)	50 (34)	
T3b	22 (15)	20 (13)	
N1	10 (6)	12 (8)	
Gleason group at RP (count, %)			1
GG1	20 (13)	18 (12)	
GG2	20 (13)	20 (13)	
GG3	67 (46)	70 (48)	
GG4	22 (15)	21 (14)	
GG5	16 (11)	16 (11)	

P values shown for continuous variables were obtained from Mann-Whitney *U* tests. Categorical variables were compared using χ^2 tests. BCR, biochemical recurrence; RP, radical prostatectomy; GG, tumor Gleason Grade Group; CEU, European ancestry; YRI, African (Yoruba) ancestry.

BL and WH patients, respectively ($P = 0.28$). Though median PGA was lower among BL compared with WH patients among all GGs, this finding did not reach statistical significance (Figure 1, C and D, and Supplemental Table 2). Similarly, there was no significant correlation between percent African ancestry and PGA in the overall cohort, nor among the self-identified BL or WH patients (Figure 1E).

Median PGA for patients with metastasis was 8.69, while patients without metastasis had a median PGA of 3.49 ($P = 0.001$). PGA was significantly associated with time to metastasis on univariable Cox regression analysis of the combined cohort ($P = 0.0009$) and among self-identified BL ($P = 0.001$) and WH patients ($P = 0.04$) (Supplemental Table 3). In a multivariable Cox regression model adjusted for age, prostate-specific antigen (PSA), and GG, PGA remained significantly associated with metastasis within the entire cohort ($P = 0.005$) and among self-identified BL patients ($P = 0.002$) but not among self-identified WH patients ($P = 0.11$). Similar results were seen when PGA was assessed by quartile in these models (Supplemental Table 3).

Self-identified race is associated with somatic and germline copy number changes in prostate cancer. Though self-identified race and genetic ancestry were not significantly associated with PGA, some segment-level differences were apparent in the prostate copy number landscape when compared by self-identified race, including a higher frequency of chr 6 and 10q deletions as well as chr 3 and 7 amplifications, among WH patients (Figure 1, A and B). Thus, we next sought to model gene-level sCNA associated with self-identified race and/or genetic ancestry. Using limma models adjusted for preoperative PSA and GG, we first identified 407 FANTOM-CAT-annotated loci where copy number varied significantly by self-identified race (FDR < 0.05) (Supplemental Figure 7A and Supplemental Table 4). Among common cancer driver genes frequently altered at the copy number level in prostate cancer, only sCNA involving *PTEN* varied significantly by race, with relative increased copy number among the BL compared with WH samples. These data are consistent with a number of previous reports confirming that *PTEN* gene deletions are less common in prostate tumors arising in BL patients (7, 11, 23, 24), and they validated our models.

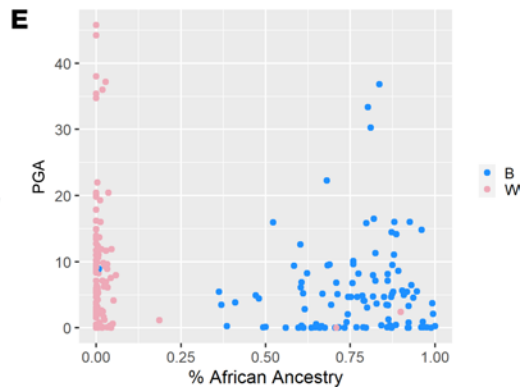
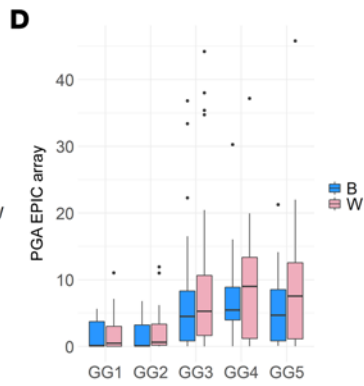
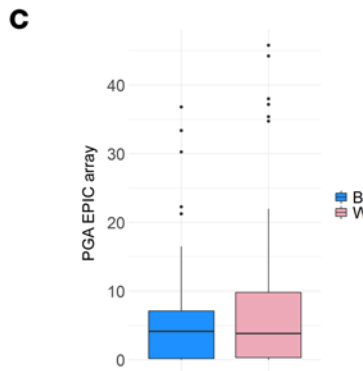
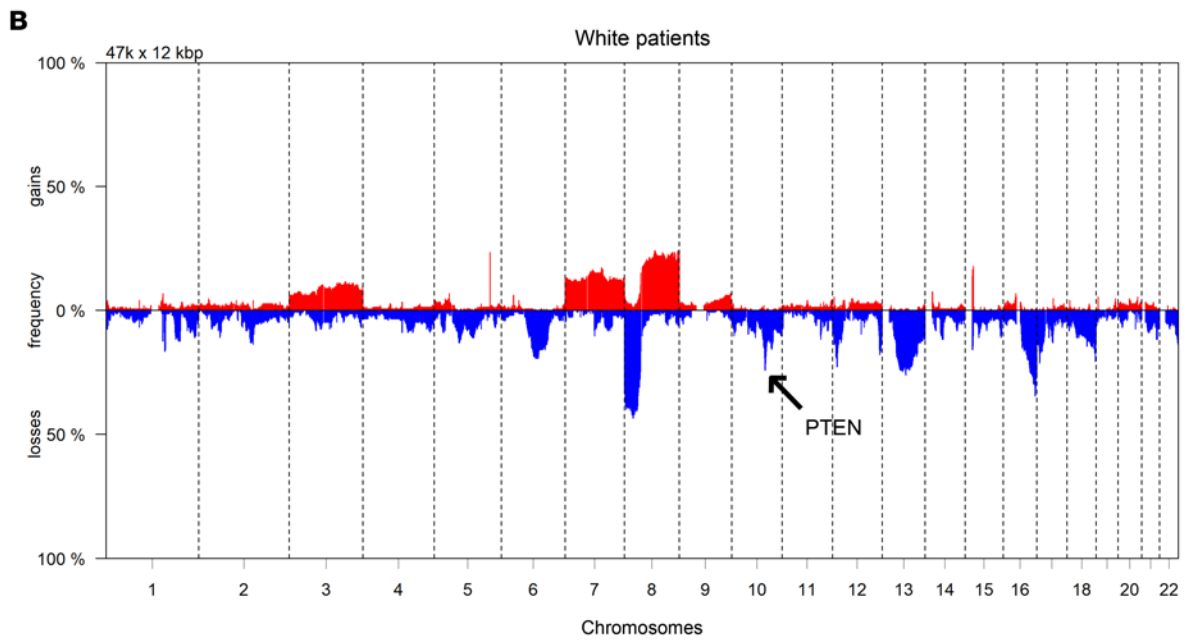
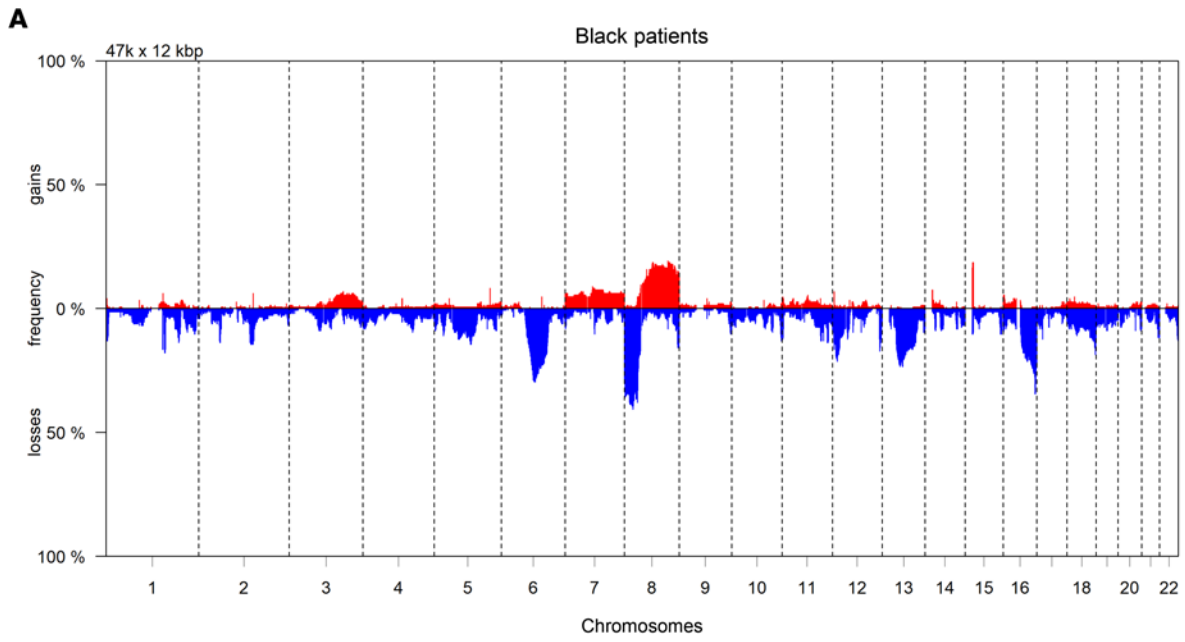


Figure 1. Segment-level somatic copy number alteration (sCNA) landscape of primary prostate tumors for patients who self-identified as Black (BL) or White (WH). (A) Landscape of copy number loss and gain for tumors arising in BL patients ($n = 145$). (B) Landscape of copy number loss and gain for tumors arising in WH patients ($n = 145$). The arrow indicates an increased frequency of PTEN loss at 10q23.31 among tumors from WH patients. (C) EPIC array-derived percent genome altered (PGA) was not associated with self-reported ancestry ($n = 249$). (D) EPIC array-derived percent genome altered (PGA) increases by Gleason Grade Groups, but this relationship was not statistically significant. (E) Percent African ancestry (YRI) did not correlate with PGA ($n = 249$).

However, among the CNAs varying by self-identified race, we noted several loci with previously described ancestry-associated germline copy number variations (gCNV). This was particularly apparent for a cluster of genes lying in a 90 Kb repeat unit on 17q12, including *CCL3L1* (the most potent ligand for CCR5, with 96% homology to CCL3), *CCL4*, and *TBC1D3*, all of which showed significantly increased relative copy number in prostate tumors from self-identified BL patients (Supplemental Figure 7A). This region is well documented to be associated with gCNV, with established higher germline copy numbers — as well as increased gene expression — for *CCL3L1* among African-ancestry individuals compared with European-ancestry individuals (39, 40).

These observations validated that our models were correctly identifying genomic regions with CNVs by race. However, these findings also suggest that the tumor sCNA comparisons could be confounded by underlying gCNV, despite normalization of our sCNA data using a pooled sample of BL and WH normal prostate tissues. In order to systematically identify gCNV associated with self-identified race using the methylation array platform, we leveraged a publicly available normal colon tissue data set, consisting of 80 self-identified BL and 40 self-identified WH subjects with InfiniumEPIC methylation profiling (41). Limma models in the colon data set revealed 1,668 loci (FDR < 0.001) (Supplemental Table 5) with significant gCNV between the 2 self-identified race groups, of which 119 loci overlapped with the sCNA loci discovered in our prostate tumor comparison by self-identified race, comprising 29% (119 of 407) of the identified prostate sCNA (Limma, FDR < 0.05) and prominently featuring the area on 17q12 (Supplemental Figure 7, B and C).

Increased percent African ancestry is linked to sCNA in prostate cancer. Next, we sought to model sCNA associated with genetic ancestry. Generalized linear models (GLMs) adjusted for age, GG, and PSA were performed across the entire FANTOM-CAT-annotated genome to identify how much of the CNAs were explained by percent African ancestry. This comparison yielded 237 loci whose copy number varied significantly by percent African ancestry (FDR < 0.05) (Supplemental Table 6), with 86% (204 of 237) overlapping with the 407 loci found with our limma model conducted with self-identified race (Figure 2A). We then accounted for gCNV present in the GLM results by filtering out loci identified through the colon analysis, which included 94 copy number variable loci likely due to gCNV (Supplemental Table 7) and a total of 143 presumed sCNA that varied by percent African ancestry (Figure 2B and Table 2).

As with the modeling by self-identified race described above, a cluster of genes identified by the GLM on 10q23 (including *PTEN* and *BMPRIA*) showed a higher relative copy number in patients with higher percent African ancestry (Figure 2, B–D, and Supplemental Figure 8). Similarly, a region on 17p near (but not including) *TP53* showed relative higher copy number in patients with increased African ancestry. This is consistent with reports of less common *TP53* mutations in prostate tumors from self-identified BL patients (15), since most *TP53* mutations are accompanied by shallow deletions/loss of heterozygosity (LOH) of the other allele in prostate cancer (28). However, we also identified a number of potentially novel regions with sCNA variable by ancestry, such as a cluster on 11p13 (*WT1*, *WT1-AS*, and *CD59*), with relatively decreased copy number in patients with a higher percentage of African ancestry, and we identified a region centered around 6p24 (including *HIVEP1*) and 12p11-12 (including *SOX5* and *DNM1L*) with increased relative copy number in patients with increased African ancestry (Supplemental Figure 9).

Chromosome 8q gains are significantly associated with biochemical recurrence and metastasis independent of clinical variables and percent African ancestry. Next, we assessed which sCNA were associated with the outcomes of biochemical recurrence and metastasis using age, GG, and PSA Cox regression models adjusted with percent African ancestry. We utilized continuous copy number data across the entire FANTOM-CAT-annotated genome for this analysis. Multivariable Cox regression revealed 162 loci (FDR < 0.1) that were significantly associated with biochemical recurrence independently of percent African ancestry and clinical pathologic variables, including gains in 8q (near but not including *MYC*), losses in 8p, losses in 10q24 (directly adjacent to *PTEN*, including *ATAD1*, which is commonly codeleted with *PTEN*), 12p13 (near *CDKN1B*), and 17p12 (Supplemental Figure 10 and Supplemental Table 8). Cox models for metastasis yielded 175 significant loci (FDR < 0.1). Strikingly, only chromosome 8q gains, including *MYC*, were significantly associated with metastasis independently of percent African ancestry (Figure 3, A and B, and

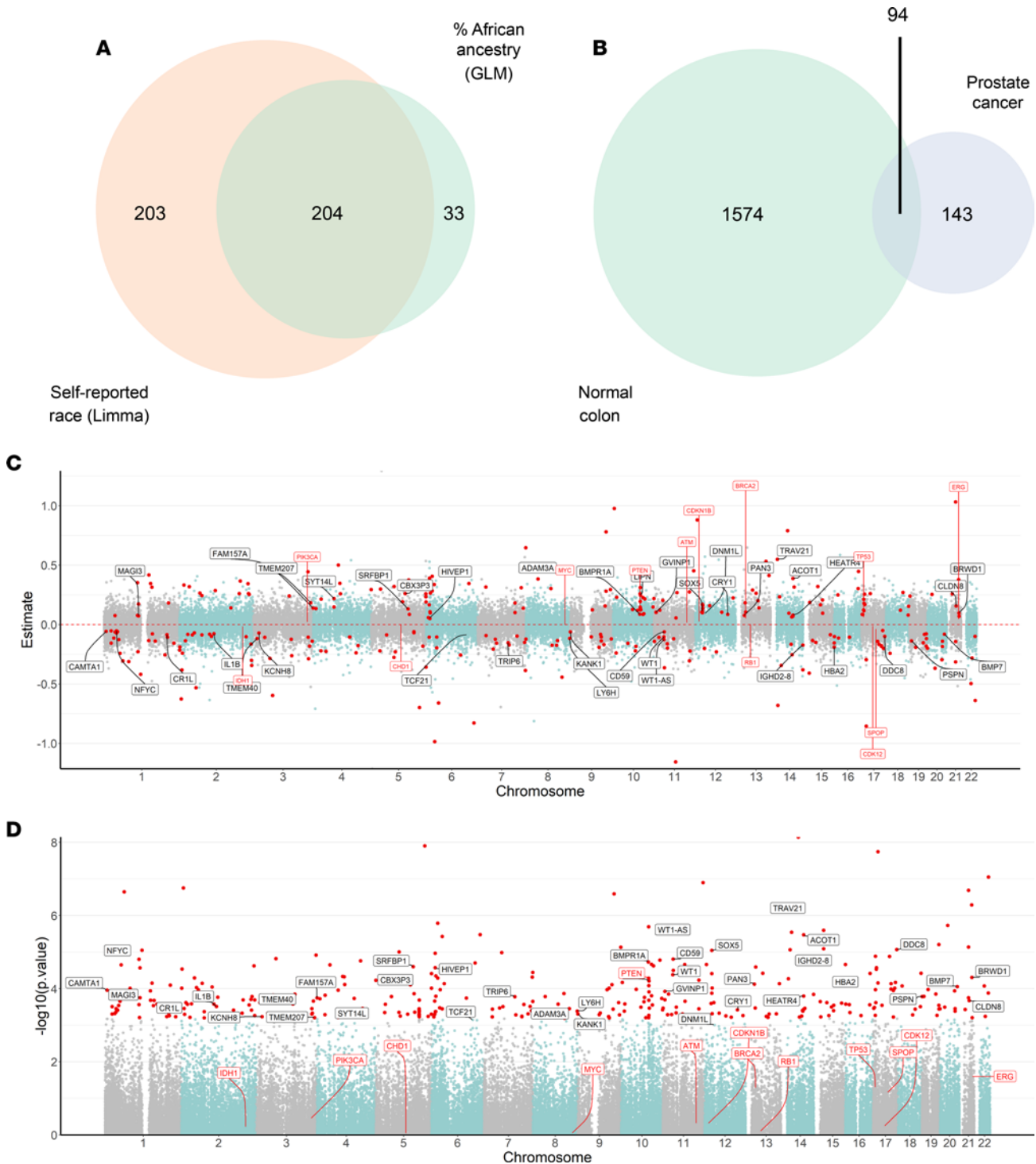


Figure 2. Gene-level sCNA landscape of primary prostate tumors by genetic ancestry. Generalized linear models (GLM) were used to compare the genomic copy number landscape of prostate tumors ($n = 290$) based on percent African ancestry adjusted with Gleason Grade Group (GG1 and GG2 versus GG3, GG4, and GG5), age, and preoperative PSA levels after removing genes likely to represent gCNV based on analysis of benign colon tissue. **(A)** Overlap between loci with sCNA varying significantly by self-reported race or by percent African ancestry ($FDR < 0.05$). **(B)** Overlap between loci with copy number alteration varying by self-reported race in normal colon tissue ($FDR < 0.001$) versus prostate cancer by percent African ancestry. From the 237 genes significantly associated with percent African ancestry in the prostate cohort, 94 overlapped with regions found in the comparison between self-identified race in colon normal tissue (Supplemental Table 7). **(C)** Estimates from GLM for each locus shown by chromosome. Each dot represents an individual locus. Red dots show loci with sCNA that vary significantly by genetic ancestry in the GLM using FDR-adjusted $P < 0.1$, with 94 overlapping loci with colon limma models excluded from the figure, as they represent presumptive gCNV. **(D)** GLM P values on the y axis are distributed by chromosome in the x axis. For ease of visualization, labels are shown for a subset of coding loci with adjusted $P < 0.01$, with the complete list available in Supplemental Table 6. Previously described prostate cancer driver genes are indicated with red labels for reference (regardless of significance in model).

Table 2. Coding genes with sCNA that vary significantly by percent African ancestry in the GLM model

Gene name	Chromosome band	Estimate	Adjusted P value
NFYC	1p34.2	-0.053	0.008
MAGI3	1p13.2	0.075	0.046
CAMTA1	1p36.31	-0.058	0.035
C1orf204	1q23.2	-0.080	0.046
STON1-GTF2A1L	2p16.3	-0.069	0.040
C2orf91	2p21	-0.086	0.022
KIAA1211L	2q11.2	-0.081	0.036
ISL1	5q11.1	-0.086	0.047
SRFBP1	5q23.1	0.132	0.015
HIVEP1	6p24.1	0.055	0.011
TRIP6	7q22.1	-0.154	0.047
BMPR1A	10q23.2	0.090	0.008
PTEN	10q23.31	0.133	0.018
LCOR	10q24.1	0.086	0.046
LIPN	10q23.31	0.250	0.023
CD59	11p13	-0.061	0.010
WT1	11p13	-0.102	0.023
FCHSD2	11q13.4	0.057	0.048
SOX5	12p12.1	0.116	0.006
PAN3	13q12.2	0.077	0.031
N4BP2L2	13q13.1	0.107	0.012
ACOT1	14q24.3	0.379	0.006
TRAV21	14q11.2	0.558	0.001
IGHD2-8	14q32.33	-0.175	0.009
IGHD1-7	14q32.33	-0.384	0.040
HBA2	16p13.3	-0.188	0.028
DDC8	17q25.3	-0.100	0.007
PSPN	19p13.3	-0.138	0.046
BMP7	20q13.31	-0.082	0.032
BRWD1	21q22.2	0.096	0.025

Estimates and adjusted P values are shown for each gene, along with the chromosomal location. Genes overlapping with benign colon analysis (presumed to be gCNV) are not shown. See Supplemental Table 6 for complete list.

Supplemental Table 9). We confirmed these results via Kaplan-Meier analysis, where *MYC* gains were significantly linked to metastasis for the entire cohort (Figure 3C) and for self-identified BL (Figure 3D) and WH patients (Figure 3E). Overall, *MYC* gains were found in 19% (54 of 282) of samples in our cohort. By self-identified race, 16% (24 of 118) WH and 21% (30 of 110) BL patients had this genomic alteration.

PGA is linked to increased immunosuppressive TME mediated by Tregs. Since BL and WH patients with prostate cancer have previously been shown to have distinct immune TME profiles (4, 9, 17–21, 26) and since these profiles are associated with tumor-specific genomic alterations (25, 29–31), we next assessed whether PGA was associated with immune cell infiltration and if this association was modified by self-identified race. We leveraged methylation profiling to estimate prostate tumor immune cell content via methylCIBERSORT, as we previously reported for B cells for this cohort (21). Treg immune cell content estimated from methylCIBERSORT was positively and significantly correlated with PGA in the combined cohort and separately for BL and WH patients (Figure 4, A–C). Consistent with this, we found Treg density quantified by a previously published FOXP3 IHC assay (29) was significantly correlated with PGA among BL men, though this correlation was not significant in the overall cohort or among WH men (Figure 4, D–F). We further validated these findings using previously published data from the TCGA cohort (42), where methylCIBERSORT-based estimates of Treg content were positively and significantly correlated with PGA estimates among BL and WH men and overall (Figure 4, G–I). Consistent with the recently reported positive prognostic associations with plasma cell content, and the increased density of these cells in prostate tumors from BL patients (21), B cell content by IHC for CD79a was significantly negatively correlated with

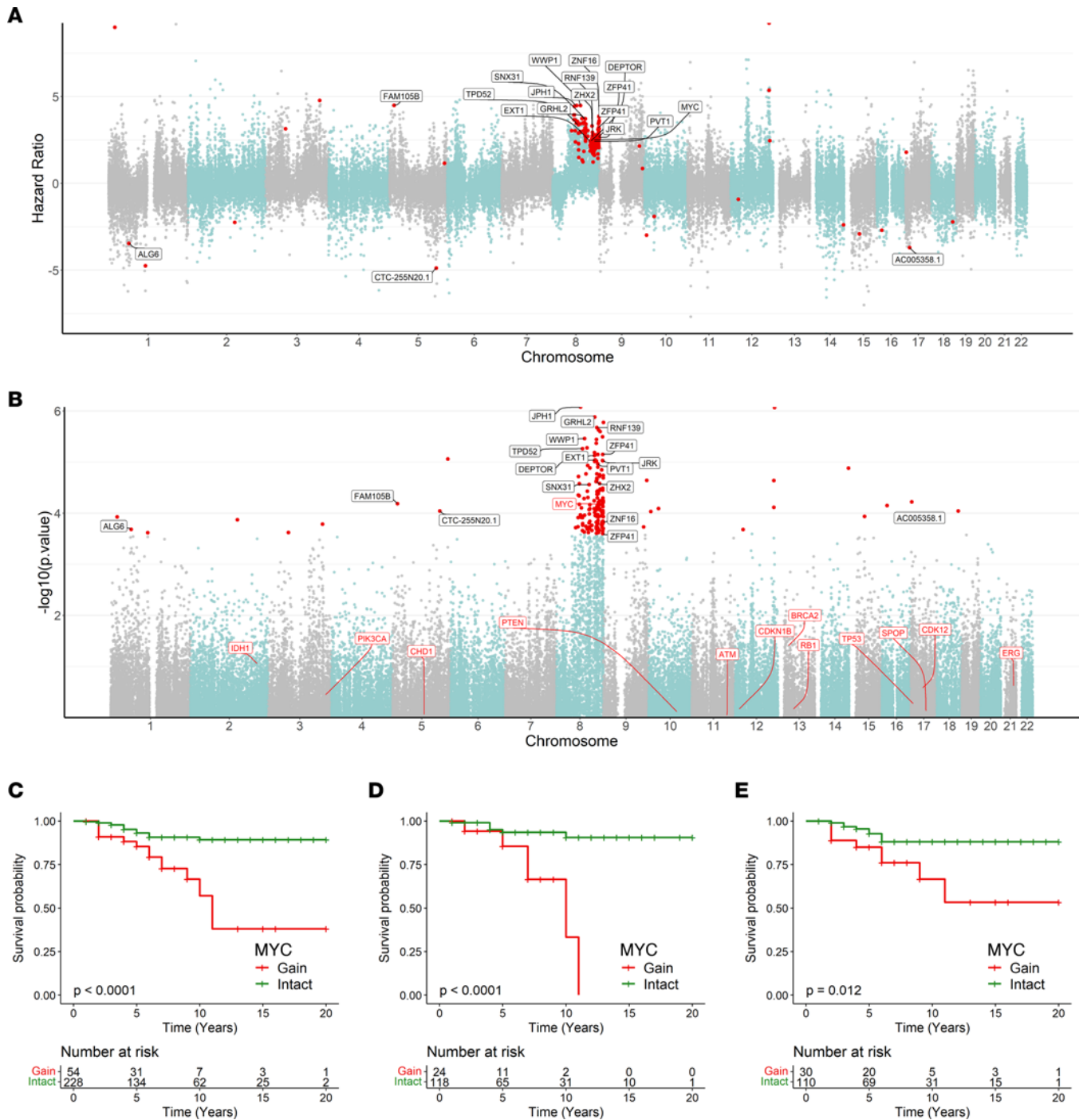


Figure 3. Chromosome 8q gains are significantly associated with risk of prostate cancer metastasis independent of ancestry. Cox regression model adjusted with percent African ancestry, age, preoperative PSA, and tumor Gleason Grade Group (GG1 and GG2 versus GG3, GG4, and GG5) was applied to all samples from our cohort ($n = 290$) and showed significant associations between 8q gains and metastasis. **(A)** Hazard ratio shown by chromosome. Each dot represents an individual locus. Red dots indicate loci where sCNA are significantly associated with risk of metastasis using FDR-adjusted $P < 0.1$. **(B)** P values on the y axis are distributed by chromosome in the x axis. For visualization purposes, only a subset of coding genes are labeled; however, the complete list is available in Supplemental Table 9. Previously described prostate cancer driver genes are indicated with red labels for reference (regardless of significance in model). **(C–E)** *MYC* gains (>2 copies) are associated with prostate cancer metastasis in Kaplan-Meier analysis. Log-rank test was employed to identify the association between *MYC* gains and metastasis. **(C)** All patients ($n = 282$). **(D)** BL patients ($n = 142$). **(E)** WH patients ($n = 140$).

PGA in the overall Johns Hopkins cohort (as well as BL and WH subsets) (Supplemental Figure 11). This finding was also replicated among BL patients in the TCGA cohort using methylCIBERSORT, but significant correlations were not found in other comparisons. Immune content estimates for other cellular subsets, including cytotoxic T cells and macrophages, were not consistently significantly correlated with PGA in either cohort (Supplemental Figures 12 and 13).

Discussion

In prostate cancer, gene deletions, amplifications, and rearrangements prevail over point mutations as the most common genomic alterations, and the total burden of sCNA, expressed as PGA, is the genomic variable most reproducibly associated with GG and clinical outcome in primary disease (43, 44). Though mortality from prostate cancer is more than twice as high among patients who self-identify as BL compared with those who self-identify as WH (1, 2), it is remarkable that few prior studies have systematically compared the copy number and immune landscapes of primary prostate cancers using racially diverse and GG-matched cohorts. In a review of prior studies, the vast majority have included only smaller cohorts (~20 patients in each group), and most larger studies, such as those using TCGA data, have contended with comparing groups that were disparate based on clinical pathologic variables (3–16) (Supplemental Table 10). Compounding these issues, only a few prior studies in unmatched cohorts have integrated genetic ancestry information into these analyses to mitigate against the heterogeneity introduced by social constructs of race and ethnicity (9, 11). Finally, no prior studies, to our knowledge, incorporated long-term clinical follow-up for metastasis and compared sCNA landscape of BL and WH patients with prostate cancer.

Given these limitations, it is unsurprising that few reproducible associations of self-identified race and sCNA have been previously reported. Only 1 major sCNA difference between BL and WH patients has held up across multiple studies using orthogonal technologies: the lower frequency of *PTEN* gene deletions among BL patients. First reported using IHC assays (23, 24), this finding has now been replicated in a number of sequencing studies and is among the few sCNA that have been tested with regard to percent African ancestry (7, 9). The lower rate of *PTEN* deletion among BL patients may be, in large part, due to the established lower frequency of *ERG* gene rearrangement in this population (11, 22–24), since *PTEN* deletion is highly enriched among tumors with *ERG* rearrangement (45, 46) and *ERG* fusions precede *PTEN* loss in the majority of cases (47, 48). In fact, when patients are stratified by dual *PTEN* and *ERG* status, *PTEN* losses are equally as common among *ERG*⁺ and *ERG*⁻ cases, respectively, in BL and WH patients (14) (data not shown). It is curious that the frequency of a genomic alteration with an established connection to poor outcomes, such as *PTEN* deletion, is actually lower among BL compared with WH men, despite the increased mortality in this population. Given that *PTEN* deletion has been associated with increased genomic instability in prostate cancer (49), this finding raises the question of whether PGA might be lower among prostate tumors from BL compared with WH men, and if not, whether there are other genomic alterations driving adverse outcomes that may be more common among BL men.

To begin to answer these questions, in the current study, we combined in silico and tissue analyses for a grade-matched cohort of nearly 300 primary tumors from self-identified BL and WH men with accompanying genetic ancestry data, to model the association of sCNA with self-identified race and clinical outcomes. Though sCNA are typically queried using SNP arrays or array comparative genomic hybridization (aCGH), methylation array data can be used with similar sensitivity (50, 51). We leveraged InfiniumEPIC data via the *conumee* platform to infer copy number across more than 70,000 coding and noncoding loci using the extended FANTOM-CAT genome annotations. Because we are not aware of previous studies taking a similar approach for prostate cancer, we extensively compared our gene-level sCNA estimates to panel-based sequencing and Oncoscan data for a subset of BL patients, as well as to genetically validated and previously published *PTEN* and p53 IHC assays in the entire cohort (25, 29). In addition, we compared methylation-based PGA to previously published PGA data based on targeted sequencing of 101 genes in a subset of BL patients (14), though this comparison was somewhat limited in that it only included a subset of patients and a panel of 101 genes.

As a final clinical validation of our methodology, we demonstrated that the methylation-based PGA estimates were associated with GG (with higher grade tumors showing increased PGA) as well as with risk of metastasis in models adjusted for clinical pathologic variables. Notably, we did not see significant differences in PGA by either self-identified race or percent African ancestry. Though tumors from BL patients

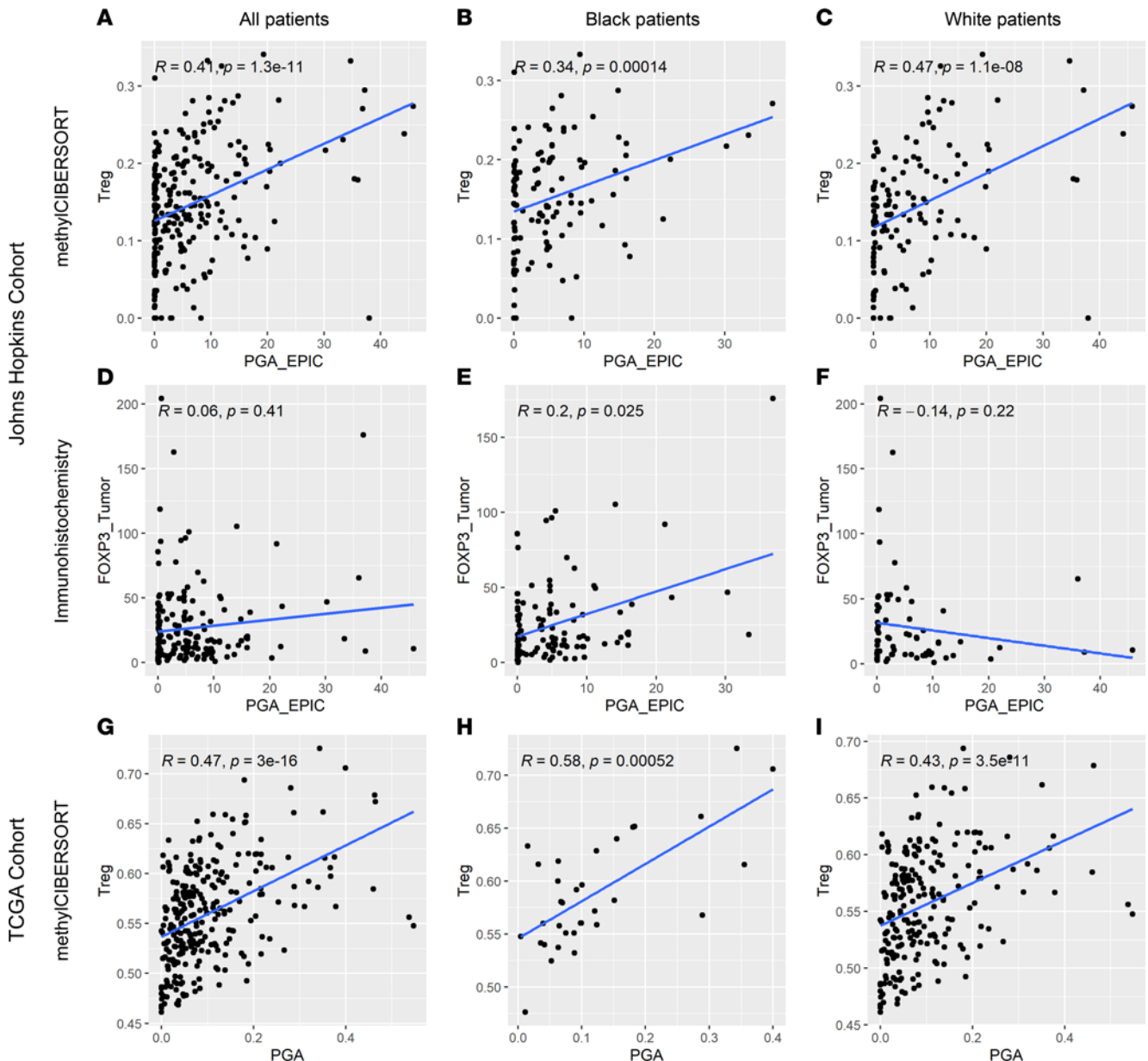


Figure 4. Tregs are significantly associated with PGA independently of self-identified race. (A–C) Tumor Treg content estimated from methyICIBERSORT was significantly associated with PGA for the entire Johns Hopkins cohort ($n = 249$) (A), for BL patients ($n = 120$) (B), and for WH patients ($n = 134$) (C), independently. (D and F) FOXP3⁺ cell density quantified by previously reported IHC (29) was not significantly associated with PGA in the entire Johns Hopkins cohort (D), nor for WH patients (F). (E) However, this association was significant for BL patients. (G–I) Validation using previously published methyICIBERSORT data (42) from the TCGA cohort showed a significant correlation between Treg content and PGA for the entire cohort ($n = 273$) (G) and separately in BL ($n = 33$) (H) and WH patients ($n = 222$) (I).

had lower absolute median PGA across GGs, these differences were not statistically significant. This suggests that, while frequencies of specific alterations such as *PTEN* deletion may vary by ancestry, PGA remains similar for primary tumors between BL and WH men and may not contribute to observed differences in clinical outcomes based on our single-institution cohort composed of African-American patients. Additional studies with more diverse populations are required to validate our findings.

We next set out to identify gene-level sCNA differences between BL and WH patients. Using linear models to compare the 2 groups while adjusting for clinical pathologic variables, we identified 407 genomic loci where copy number differed by self-identified race. Validating this methodology, *PTEN* was identified as one of the only known prostate cancer drivers on this list, consistent with previous studies (7, 11, 23, 24). Given that we had normalized our tumor copy number data using a reference cohort of benign tissues from

BL and WH men, we were surprised to also find among the 407 loci several other genes where germline copy number has been reported to vary by genetic ancestry. Many of the genes with the highest effect sizes in our model mapped to segmental duplications, including a region on chromosome 17q12 containing *CCL3L1*, as well as *CCL3* and *CCLA* (52). These genes also appeared in models using percent African ancestry as the output, which were generally highly overlapping with models using self-identified race. *CCL3L1* is the paralogue of *CCL3*, encoding the most potent ligand for CCR5, which also serves as a coreceptor for HIV (39). Because of this association, gCNV of the *CCL3L1* locus have been investigated in depth, especially with respect to gene expression of *CCL3L1* in T cells. Both genomic copy number and transcript levels for *CCL3L1* are significantly higher in patients of African compared with European ancestry, with median germline copy numbers of 4 (range 1–11) and 2 (range 0–9), respectively, and a linear relationship between genomic copy number and T cell gene expression (39). Consistent with this, we found that *CCL3L1* was overexpressed in BL prostate cancer tumors from the TCGA cohort compared with WH patients (Supplemental Figure 14). Nearby, *CCL4L1* and *TBC1D3* also exhibit gCNV, with higher copy numbers in African patients (40). While *CCL3* and *CCLA* are reported to be germline copy number invariant by quantitative PCR (qPCR) assays (53), they often appear as copy number variable in array-based genome-wide studies (including this study) (40), likely due to the high level of homology with their respective isoforms, and this high level of homology makes them difficult to distinguish in many assays due to nonspecific hybridization.

While the finding of higher copy numbers of *CCL3L1*, *CCLA*, and *TBC1D3* validated our models comparing patients by self-identified race and genetic ancestry, these data suggest that our sCNA list also inadvertently contained gCNV. To more systematically exclude contributions from gCNV in our tumor CNA comparison, we leveraged a public EPIC methylation data set derived from benign colon tissue from both BL and WH patients. Comparing copy number stratified by self-identified race using limma models in this data set, we comprehensively identified more than 1,600 loci that are likely to be affected by gCNV, which comprised nearly 30% (119 of 407) of the loci identified in our prostate tumor comparison by self-identified race and 40% (94 of 237) of the loci identified in our prostate tumor comparison by genetic ancestry. To our knowledge, we are the first group to leverage methylation profiling of benign tissues to discern gCNV by self-identified race.

In retrospect, the confounding effect of gCNV is apparent in several previous prostate cancer studies that have compared BL and WH patients. For example, a prior aCGH study comparing 20 BL and 21 WH prostate tumors found a region within 17q12 to be an area of relative gains among BL patients, though a germline contribution was not acknowledged (4). Strikingly, *CCL3* and *CCLA* also comprise 2 of the immune genes with reproducibly increased expression in BL compared with WH prostate tumors, a finding replicated in multiple independent array-based studies (20, 26, 54) but conspicuously absent from RNA-Seq-based studies such as TCGA (Supplemental Figure 14). Given the high level of homology of these isoforms to *CCL3L1* and *CCL4L1*, where copy number and prostate gene expression are established to be higher in individuals of African descent (Supplemental Figure 14), it seems likely that these findings are interrelated and a consequence of previously unrecognized germline copy number differences. Future studies will be required to determine whether *CCL3L1* and *CCL4L1* expression levels may be misclassified as *CCL3* or *CCLA*, respectively, in published gene expression data sets.

After excluding gCNV, we identified close to 150 loci where sCNA varies by percent African ancestry, a list that highly overlapped with comparisons by self-identified race. In addition to the region around *PTEN*, which showed higher relative copy number in patients with higher African ancestry, similar findings were noted for a region on 17p near *TP53*, consistent with reports of more infrequent *TP53* alterations in prostate tumors from self-identified BL patients (14–16, 25). Notably, we also identified a number of potentially novel regions with CNVs by ancestry, including clusters on 11p13, 6p24.1, and 12p12.1 (Table 2). These regions have not been identified as differing by ancestry in previous studies (Supplemental Table 10). However, 11p13 has previously been reported to contain a putative metastasis suppressor for prostate cancer (55), and 12p12 has been reported to be affected by deletions in predominantly WH cohorts (56). Future studies may delve into the functional and clinical effects of deletions in these regions that appear to be less common among BL patients. Of interest, the 8q region (including *MYC*), which was most closely associated with risk of metastatic disease in our cohorts, did not show significant CNV by genetic ancestry in our analyses, despite prior reports suggesting more frequent amplification in BL men in clinically unmatched cohorts (9, 16). Due to the underrepresentation of BL patients in most genomic studies of prostate cancer, additional large, matched cohorts will be useful to confirm our findings.

Previous studies have demonstrated distinct patterns of immune cell infiltration in the TME from self-identified BL and WH patients with prostate cancer (4, 9, 17–21). In parallel, *in silico* investigations have demonstrated that increased levels of aneuploidy — a measure of chromosomal instability — are linked to lower immune response in several tumor types (30, 57), though prostate cancer (a classically immunologically “cold tumor”) did not show a strong association in these studies. Herein, we leveraged tumor immune content estimates via methylCIBERSORT and assessed whether they were correlated with PGA in a diverse cohort. Strikingly, we found that PGA, a measure of the percent of the genome affected by genomic losses or gains, was positively correlated with increased Treg content in tumor tissue, an observation that held true for both self-identified races and one that was validated using previously published IHC for FOXP3⁺ cell densities in BL patients (29) and methylCIBERSORT data from the TCGA cohort (42). This finding is reminiscent of a recent observation from our group that Treg content by methylCIBERSORT in the TCGA cohort is increased in tumors with higher homologous recombination deficiency (HRD) score (42), another measure of genomic instability that integrates whole-genome tumor LOH, telomeric allelic imbalance, and large-scale state transition scores (58). Notably, we found the association of Treg content with PGA was significantly stronger than the association of FOXP3⁺ cell density with PGA. This might be due to the nearly 10-fold larger volume of tissue sampled for DNA isolation compared with tissue microarray preparation, and this nearly 10-fold larger volume of tissue may better capture intratumoral heterogeneity in immune cell content.

Consistent with our previous observation of improved prognosis with increased B cell content in primary prostate cancer (21), we found that B cell densities by IHC were inversely correlated with PGA, which is itself an adverse prognostic marker. This finding held up in both self-identified race groups in the Johns Hopkins cohort but was only replicated among BL patients in the TCGA cohort via methylCIBERSORT, perhaps due to higher B cell content in these tumors, as we have previously described (21). Notably, however, we were unable to replicate this in the methylCIBERSORT data from the Hopkins cohort. Taken together, these findings suggest that immunosuppressive T cell content is positively correlated with measures of genome-wide genomic instability regardless of race in primary prostate cancer, while there is a weaker negative correlation for B cells and PGA. Future studies will be required to parse out the extent to which these findings are causal or confounded by associations with other variables.

Our study has several limitations. First, we were unable to obtain germline blood or saliva samples from the patients due to the long interval since treatment in this cohort. The lack of these data precluded germline normalization on a case-by-case basis, and we instead used a pooled subset of normal prostate tissues for normalization and a large race-matched cohort of normal colon tissue samples for filtering for gCNV. Though we do not have evidence of tissue-specific gCNV that might affect colon and not prostate, we cannot completely exclude this possibility. Second, though we conducted several checks to validate our EPIC-derived sCNA data — including comparison with SNP-based sCNA assessment, in some cases — and found good agreement, it is possible that sCNA assessment via methylation profiling systematically over- or undercalls sCNAs at some loci. Third, to account for potential losses and gains occurring randomly, we increased the density of our investigated loci by employing FANTOM-CAT-annotated genomes, which contain > 70,000 coding and noncoding loci. However, many regions with lower density of loci may remain undetected, and structural rearrangements, such as translocations, were not assessed. Lastly, this is a single-institution and single-region cohort study. Therefore, additional studies with patients from other institutions and other geographic regions are necessary to validate these findings.

Conclusions. Herein, we have leveraged methylation profiling to define the sCNA and immune landscape of primary prostate cancer in a diverse cohort with integrated genetic ancestry information and clinical follow-up. After extensive validation, we provide the first comprehensive mapping to our knowledge of genomic loci where loss or gain is associated with genetic ancestry, and we demonstrate that the global sCNA burden of prostate tumors does not vary by ancestry. We note that gCNV are potential confounders in sCNA analyses and may account for differences in immune gene expression previously reported by other groups for prostate tumors arising in BL versus WH patients. Finally, we identify an association between global burden of sCNA and Treg infiltrate that is independent of ancestry.

Methods

Samples. Our study cohort has been previously described (14, 24, 29) and comprised 290 patients with prostate cancer treated with radical prostatectomy at the Johns Hopkins Hospital or Bayview Medical Center (Baltimore, Maryland, USA), including 145 men self-identified as BL and 145 men self-identified

as WH. Radical prostatectomies were performed between 1995 and 2010. For GG 1 and 2, which were relatively more common, 50 BL men with clinical follow-up were randomly selected from the total number of consecutive cases during that time period. For other GG categories, all BL men meeting criteria were included. Within each GG, an equivalent number of prostatectomy samples from self-identified WH men were selected from the same time period via stratified random sampling to provide a grade-matched WH cohort for comparison. Of this final cohort of 391 cases, we selected 290 grade-matched samples with at least 500 ng of tumor DNA required for methylation analyses. The entire 290-sample cohort was used for further analyses. However, methods employing quality control thresholds and other cutoffs led to a reduced sample size for a few comparisons (Supplemental Figure 15).

Normal adjacent prostate tissue was obtained from a subset of 30 randomly selected patients from the above cohort with adequate benign tissue in the radical prostatectomy specimen, including 15 samples from BL and 15 samples from WH men (Supplemental Table 11). In the vast majority of cases (28 of 30), a completely normal radical prostatectomy quadrant section, without any tumor present, was used to sample benign prostate for normal DNA isolation. These blocks were typically several centimeters from tumor. In cases where most blocks were extensively involved by tumor, we sampled adjacent normal tissue from tumor-containing blocks, as long as it was > 3 mm from the tumor sample. Genetic ancestry estimation on this cohort was obtained through SNP array data (Global Screening Array v3, GSAv3), as previously published (33).

Infinium MethylationEPIC BeadChip arrays. Tumor regions with at least 50% tumor cellularity by pathologic examination were visually annotated for the dominant tumor nodule (highest GG), and 4–6 tumor punches (0.6 mm diameter) were obtained in these regions for DNA isolation. In a prior publication utilizing tumor DNA samples from BL patients in this cohort (14), the median variant allele frequency for SPOP mutations (23%) and FOXA1 (27%) suggested that our visual estimates of 50% tumor cellularity are reasonable; however, visual estimation may be inaccurate. DNA from formalin-fixed paraffin-embedded tumor regions was obtained as previously described (14). All tumors and benign samples were profiled via Infinium MethylationEPIC BeadChip (Illumina), which covers 850,000 methylation sites at single-nucleotide resolution. IDAT files containing red and green channels from the methylation output were combined for each sample using the *minfi* package in R (59). Quality control checks performed via *shinyMethyl* (60) were used to identify abnormally distributed probe quantification sites. Normalized data were obtained through the Subset quantile Within-Array Normalization (SWAN), which normalizes Infinium type I and II probes simultaneously. We confirmed that our experiments and analyses were in line with previous findings confirming the hypermethylation state of *GSTP1* and *PTGS1* in tumor compared with normal samples (data not shown). Raw and processed methylation data are available on the Gene Expression Omnibus (GEO) repository (GSE221219; <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE221219>).

sCNA assessment from EPIC array methylation assays using conumee. We used *Methylset* objects obtained from *preprocessRaw* functions in *minfi* to derive whole-genome copy number data using the *conumee* package in R (34), which has been used for CNA estimation in many tumor types (61–67). The 30 normal samples were used as the reference genome to calibrate the copy number calls for the tumor cohort. Segmented regions of copy number change derived from *conumee* were then used in further analyses. For gene-level copy number estimation, we leveraged a comprehensive genomic annotation using the permissive set of the FANTOM-CAT (68, 69), which is composed of 124,245 coding and noncoding gene annotations. These loci were then overlapped with those from EPIC array, which yielded 70,198 loci. This annotation increased the density of genes and other regions in the genome to improve our ability to determine copy number regions more precisely. To visualize and compute region-level frequency of genomic gains and losses, we used the *CGHcall* package (70) to obtain regions that were altered by self-identified race.

Identification of sCNA for self-identified race in prostate cancer. Aiming to identify copy number loci that varied according to self-identified race, we employed a GLM approach (*limma*) (71) using the continuous gene-level copy number data derived from our EPIC array methylation analysis. These models were adjusted with preoperative PSA, age, and GG (with GG1 and GG2 grouped as low-risk and GG3, GG4, and GG5 grouped as high-risk). *P* values were adjusted using the Benjamini-Hochberg method.

Assessment of gCNV by self-identified race. To filter out CNAs varying by race that might be due to underlying gCNV, we leveraged a large publicly available EPIC array methylation data set encompassing 128 normal colon samples, of which 88 were from BL and 40 were from WH healthy individuals without a history of any cancer diagnosis or a family history of colon cancer syndromes (GSE151732) (41). The colon data

were processed, normalized, and analyzed via the same bioinformatic pipeline used for our data, comparing cases by self-identified race in order to identify gCNV. Genomic regions/genes of copy number change between BL and WH individuals identified in the normal colon samples were deemed to be germline and removed from all subsequent analyses when overlapping those identified in prostate cancer.

Modeling sCNA associated with percent African ancestry. To complement our analyses and findings based on self-identified race, we used a GLM approach. To this end, we used the EPIC-derived continuous copy number data to model the percent African ancestry in our prostate cancer cohort, adjusting by GG, age, and preoperative PSA levels. *P* values were adjusted using the Benjamini-Hochberg method.

Modeling biochemical recurrence and metastasis. Next, we performed multivariable Cox proportional hazards regression to identify loci (genes and lncRNAs) associated with biochemical recurrence and metastasis in our cohort. These models were also adjusted for GG, PSA, age, and percent African ancestry. The Benjamini-Hochberg method was used to adjust for multiple comparisons. Gene-level copy number gains and losses obtained from the *CGHcall* package were used to validate our findings from Cox regression models.

Segment-level and PGA of sCNA calls. We next obtained PGA estimates for each sample from EPIC array copy number data obtained from *conumee* at the genomic segment level. We used the expectation-maximization (EM) algorithm to model the copy number distribution in the complete prostate cancer cohort as a mixture of 3 normal distributions — representing gained, lost, and intact copy number regions; we then assigned each segment for each case to 1 of such normal distributions. In order to calculate the PGA, we obtained the genomic coordinates for all segments, calculated the cumulative size of those deemed gained or lost separately for each sample, and then divided such sums by the whole genome size. In this analysis, 36 samples did not have any segmental sCNA that passed our cutoff points and were not used in downstream analyses. We validated the PGA data obtained from the EPIC array platform with sequencing-derived PGA data from a subset of 134 patients from our cohort. PGA was also used in logistic regression and Cox regression models adjusted by GG, PSA levels, and age.

Estimation of tumor immune cell content. To identify the association between PGA and immune cell content, we used *methylCIBERSORT* (72), a methylation-based deconvolution method to obtain cell density scores for CD14 cells (macrophages), CD19 cells (B cells), CD4 effector T cells, CD56 cells (NK cells), CD8 T cells, endothelial cells, eosinophils, fibroblasts, neutrophils, and Tregs, as previously published for B cells in this cohort (21) and in the TCGA cohort (42). We used previously published IHC data from our cohort to validate our *in silico* deconvolution method. Markers for Tregs (FOXP3), B cells (CD79), CD8 T cells, and macrophages (CD163) were employed as described previously (21, 29, 33).

Availability of data and materials. InfiniumEpic Methylation array data are pending deposition in GEO. Public-domain data set used (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE151732>).

Statistics. Fisher's exact and χ^2 tests were used to compare self-identified race groups for categorical data. Mann-Whitney *U* test was used for clinical continuous data. Linear regression models were conducted to determine sCNA status by genetic ancestry. Spearman's correlation was used to determine the relationship between immune cells and PGA. *P* value threshold was set to 0.05, and Benjamini-Hochberg FDR was set to 0.1 or less as described for each analysis.

Study approval. This study was approved by the Johns Hopkins IRB in accordance with the US Common Rule under a waiver of consent to enable research on excess archival biospecimens (IRB00208526).

Author contributions

TLL, TV, and LM conceived the study. FF, SM, HK, SZ, JX, KSS, ELI, AAM, WBI, and EMS contributed to data collection and analysis. TLL, TV, and LM drafted the manuscript. All authors critically reviewed the manuscript and agreed to submit for publication.

Acknowledgments

This research was supported in part by 2 Health Disparity Research Awards from the CDMRP-PCRP (W81XWH-17-1-0286 to KSS and TLL; W81XWH-19-1-0292 to TLL). Additional funding and resources were provided by the Schaufeld Program for Prostate Cancer in Black Men and the NCI Cancer Center Support grants 5P30CA006973-52 to TLL and R01 CA200859 to LM.

Address correspondence to: Tamara Lotan, 1550 Orleans Street, Baltimore, Maryland 21231, USA. Phone: 410.614.9196; Email: tlotan1@jhmi.edu.

1. American Cancer Society. Cancer Facts & Figures for African Americans. <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/cancer-facts-and-figures-for-african-americans/cancer-facts-and-figures-for-african-americans-2019-2021.pdf>. Accessed January 5, 2023.
2. Giaquinto AN, et al. Cancer statistics for African American/Black People 2022. *CA Cancer J Clin.* 2022;72(3):202–229.
3. Castro P, et al. Genomic profiling of prostate cancers from African American men. *Neoplasia.* 2009;11(3):305–312.
4. Rose AE, et al. Copy number and gene expression differences between African American and Caucasian American prostate cancer. *J Transl Med.* 2010;8:70.
5. Petrovics G, et al. A novel genomic alteration of LSAMP associates with aggressive prostate cancer in African American men. *EBioMedicine.* 2015;2(12):1957–1964.
6. Lindquist KJ, et al. Mutational landscape of aggressive prostate tumors in African American men. *Cancer Res.* 2016;76(7):1860–1868.
7. Huang FW, et al. Exome sequencing of African-American prostate cancer reveals loss-of-function ERF mutations. *Cancer Discov.* 2017;7(9):973–983.
8. Jaratlersiri W, et al. Whole-genome sequencing reveals elevated tumor mutational burden and initiating driver mutations in African men with treatment-naïve, high-risk prostate cancer. *Cancer Res.* 2018;78(24):6736–6746.
9. Yuan J, et al. Integrative comparison of the genomic and transcriptomic landscape between prostate cancer patients of predominantly African or European genetic ancestry. *PLoS Genet.* 2020;16(2):e1008641.
10. Yadav S, et al. Somatic mutations in the DNA repairome in prostate cancers in African Americans and Caucasians. *Oncogene.* 2020;39(21):4299–4311.
11. Koga Y, et al. Genomic profiling of prostate cancers from men with African and European ancestry. *Clin Cancer Res.* 2020;26(17):4651–4660.
12. Chen Y, et al. Breast and prostate cancers harbor common somatic copy number alterations that consistently differ by race and are associated with survival. *BMC Med Genomics.* 2020;13(1):116.
13. Liu W, et al. Distinct genomic alterations in prostate tumors derived from African American men. *Mol Cancer Res.* 2020;18(12):1815–1824.
14. Faisal FA, et al. CDKN1B deletions are associated with metastasis in African American men with clinically localized, surgically treated prostate cancer. *Clin Cancer Res.* 2020;26(11):2595–2602.
15. Mahal BA, et al. Racial differences in genomic profiling of prostate cancer. *N Engl J Med.* 2020;383(11):1083–1085.
16. Stopsack KH, et al. Differences in prostate cancer genomes by self-reported race: contributions of genetic ancestry, modifiable cancer risk factors, and clinical factors. *Clin Cancer Res.* 2022;28(2):318–326.
17. Wallace TA, et al. Tumor immunobiological differences in prostate cancer between African-American and European-American men. *Cancer Res.* 2008;68(3):927–936.
18. Powell IJ, et al. Genes associated with prostate cancer are differentially expressed in African American and European American men. *Cancer Epidemiol Biomarkers Prev.* 2013;22(5):891–897.
19. Hardiman G, et al. Systems analysis of the prostate transcriptome in African-American men compared with European-American men. *Pharmacogenomics.* 2016;17(10):1129–1143.
20. Awasthi S, et al. Comparative genomics reveals distinct immune-oncologic pathways in African American men with prostate cancer. *Clin Cancer Res.* 2021;27(1):320–329.
21. Weiner AB, et al. Plasma cells are enriched in localized prostate cancer in Black men and are associated with improved outcomes. *Nat Commun.* 2021;12(1):935.
22. Magi-Galluzzi C, et al. TMPRSS2-ERG gene fusion prevalence and class are significantly different in prostate cancer of Caucasian, African-American and Japanese patients. *Prostate.* 2011;71(5):489–497.
23. Khani F, et al. Evidence for molecular differences in prostate cancer between African American and Caucasian men. *Clin Cancer Res.* 2014;20(18):4925–4934.
24. Tosoian JJ, et al. Prevalence and prognostic significance of PTEN Loss in African-American and European-American men undergoing radical prostatectomy. *Eur Urol.* 2017;71(5):697–700.
25. Kaur HB, et al. TP53 missense mutation is associated with increased tumor-infiltrating T cells in primary prostate cancer. *Hum Pathol.* 2019;87:95–102.
26. Rahmatpanah F, et al. RNA expression differences in prostate tumors and tumor-adjacent stroma between Black and White Americans. *Oncotarget.* 2021;12(15):1457–1469.
27. Taylor BS, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell.* 2010;18(1):11–22.
28. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell.* 2015;163(4):1011–1025.
29. Kaur HB, et al. Association of tumor-infiltrating T-cell density with molecular subtype, racial ancestry and clinical outcomes in prostate cancer. *Mod Pathol.* 2018;31(10):1539–1552.
30. Davoli T, et al. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science.* 2017;355(6322):eaaf8399.
31. Vidotto T, et al. PTEN-deficient prostate cancer is associated with an immunosuppressive tumor microenvironment mediated by increased expression of IDO1 and infiltrating FoxP3+ T regulatory cells. *Prostate.* 2019;79(9):969–979.
32. Maynard JP, et al. P2X4 purinergic receptors offer a therapeutic target for aggressive prostate cancer. *J Pathol.* 2022;256(2):149–163.
33. Mendes AA, et al. Association of B7-H3 expression with racial ancestry, immune cell density, and androgen receptor activation in prostate cancer. *Cancer.* 2022;128(12):2269–2280.
34. Hovestadt V, Zapatka M. Conumee: enhanced copy-number variation analysis using Illumina DNA methylation arrays. <https://bioconductor.org/packages/release/bioc/html/conumee.html>. Accessed January 5, 2023.
35. Irizarry RA, et al. Multiple-laboratory comparison of microarray platforms. *Nat Methods.* 2005;2(5):345–350.
36. Ross AE, et al. Gene expression pathways of high grade localized prostate cancer. *Prostate.* 2011;71(14):1568–1577.
37. Guedes LB, et al. Analytic, preanalytic, and clinical validation of p53 IHC for detection of TP53 missense mutation in prostate cancer. *Clin Cancer Res.* 2017;23(16):4693–4703.

38. Maughan BL, et al. p53 status in the primary tumor predicts efficacy of subsequent abiraterone and enzalutamide in castration-resistant prostate cancer. *Prostate Cancer Prostatic Dis.* 2018;21(2):260–268.
39. Bhattacharya T, et al. CCL3L1 and HIV/AIDS susceptibility. *Nat Med.* 2009;15(10):1112–1115.
40. Sudmant PH, et al. Diversity of human copy number variation and multicopy genes. *Science.* 2010;330(6004):641–646.
41. Devall M, et al. Racial disparities in epigenetic aging of the right versus left colon. *J Natl Cancer Inst.* 2020;113(12):1779–1782.
42. Kaur HB, et al. Association between pathogenic germline mutations in BRCA2 and ATM and tumor-infiltrating lymphocytes in primary prostate cancer. *Cancer Immunol Immunother.* 2021;71(4):943–951.
43. Hieronymus H, et al. Copy number alteration burden predicts prostate cancer relapse. *Proc Natl Acad Sci U S A.* 2014;111(30):11139–11144.
44. Stopsack KH, et al. Aneuploidy drives lethal progression in prostate cancer. *Proc Natl Acad Sci U S A.* 2019;116(23):11390–11395.
45. Yoshimoto M, et al. Absence of TMPRSS2:ERG fusions and PTEN losses in prostate cancer is associated with a favorable outcome. *Mod Pathol.* 2008;21(12):1451–1460.
46. Carver BS, et al. Reciprocal feedback regulation of PI3K and androgen receptor signaling in PTEN-deficient prostate cancer. *Cancer Cell.* 2011;19(5):575–586.
47. Gumuskaya B, et al. Assessing the order of critical alterations in prostate cancer development and progression by IHC: further evidence that PTEN loss occurs subsequent to ERG gene fusion. *Prostate Cancer Prostatic Dis.* 2013;16(2):209–215.
48. Bismar TA, et al. PTEN genomic deletion is an early event associated with ERG gene rearrangements in prostate cancer. *BJU Int.* 2011;107(3):477–485.
49. Williams JL, et al. Recurrent copy number alterations in prostate cancer: an in silico meta-analysis of publicly available genomic data. *Cancer Genet.* 2014;207(10–12):474–488.
50. Capper D, et al. DNA methylation-based classification of central nervous system tumours. *Nature.* 2018;555(7697):469–474.
51. Sievers P, et al. Recurrent fusions in PLAGL1 define a distinct subset of pediatric-type supratentorial neuroepithelial tumors. *Acta Neuropathol.* 2021;142(5):827–839.
52. Modi WS. CCL3L1 and CCL4L1 chemokine genes are located in a segmental duplication at chromosome 17q12. *Genomics.* 2004;83(4):735–738.
53. Townson JR, et al. Gene copy number regulates the production of the human chemokine CCL3-L1. *Eur J Immunol.* 2002;32(10):3016–3026.
54. Rayford W, et al. Comparative analysis of 1152 African-American and European-American men with prostate cancer identifies distinct genomic and immunological differences. *Commun Biol.* 2021;4(1):670.
55. Gao AC, et al. CD44 is a metastasis suppressor gene for prostatic cancer located on human chromosome 11p13. *Cancer Res.* 1997;57(5):846–849.
56. Kibel AS, et al. Deletion mapping at 12p12-13 in metastatic prostate cancer. *Genes Chromosomes Cancer.* 1999;25(3):270–276.
57. Taylor AM, et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell.* 2018;33(4):676–689.
58. Timms KM, et al. Association of BRCA1/2 defects with genomic scores predictive of DNA damage repair deficiency among breast cancer subtypes. *Breast Cancer Res.* 2014;16(6):475.
59. Aryee MJ, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 2014;30(10):1363–1369.
60. Fortin JP, et al. shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R. *F1000Res.* 2014;3:175.
61. Kommos FKF, et al. Clinicopathologic and molecular analysis of embryonal rhabdomyosarcoma of the genitourinary tract: evidence for a distinct DICER1-associated subgroup. *Mod Pathol.* 2021;34(8):1558–1569.
62. Barrett JE, et al. The DNA methylome of cervical cells can predict the presence of ovarian cancer. *Nat Commun.* 2022;13(1):448.
63. Huang M, et al. Engineering genetic predisposition in human neuroepithelial stem cells recapitulates medulloblastoma tumorigenesis. *Cell Stem Cell.* 2019;25(3):433–446.
64. Banan R, et al. Infratentorial IDH-mutant astrocytoma is a distinct subtype. *Acta Neuropathol.* 2020;140(4):569–581.
65. Reinhardt A, et al. Tumors diagnosed as cerebellar glioblastoma comprise distinct molecular entities. *Acta Neuropathol Commun.* 2019;7(1):163.
66. Suwala AK, et al. Glioblastomas with primitive neuronal component harbor a distinct methylation and copy-number profile with inactivation of TP53, PTEN, and RB1. *Acta Neuropathol.* 2021;142(1):179–189.
67. Kilaru V, et al. Critical evaluation of copy number variant calling methods using DNA methylation. *Genet Epidemiol.* 2020;44(2):148–158.
68. Hon CC, et al. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature.* 2017;543(7644):199–204.
69. Imada EL, et al. Transcriptional landscape of PTEN loss in primary prostate cancer. *BMC Cancer.* 2021;21(1):856.
70. Van de Wiel MA, et al. CGHcall: calling aberrations for array CGH tumor profiles. *Bioinformatics.* 2007;23(7):892–894.
71. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
72. Chakravarty A, et al. Pan-cancer deconvolution of tumour composition using DNA methylation. *Nat Commun.* 2018;9(1):3220.