Cancer Research – Research Article

Whole Genome Sequencing Reveals Elevated Tumor Mutational Burden and Initiating Driver Mutations in African Men with Treatment-Naïve, High-Risk Prostate Cancer

Weerachai Jaratlerdsiri[1], Eva K. F. Chan[1,2], Tingting Gong[1,3], Desiree C. Petersen[1,2], Anton M.F. Kalsbeek[1,2], Philip A. Venter[4], Phillip D. Stricker[5], M.S. Riana Bornman[6], and Vanessa M. Hayes[1,2,3,4,6]

[1]Laboratory for Human Comparative and Prostate Cancer Genomics, Genomics and Epigenetics Division, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia. [2]St Vincent's Clinical School, University of New South Wales, Randwick, NSW, Australia. [3]Central Clinical School, University of Sydney, Camperdown, NSW, Australia. [4]Department of Medical Sciences, University of Limpopo, Turfloop Campus, Limpopo, South Africa. [5]Department of Urology, St. Vincent's Hospital Sydney, Darlinghurst, NSW 2010, Australia. [6]School of Health Systems and Public Health, University of Pretoria, Pretoria, South Africa.

Corresponding Author: Vanessa M. Hayes, Laboratory for Human Comparative and Prostate Cancer Genomics, Genomics and Epigenetics Division, Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, NSW 2010, Australia; e-mail: v.hayes@garvan.org.au

Conflicts of interest statement: The authors declare no potential conflicts of interest.

Keywords: prostate cancer, genomes, Africa, tumor mutational burden, high-risk disease

Running title: Genomic hallmark of high-risk prostate cancer in African men

Abstract

African American men are more likely than any other racial group to die from prostate cancer. The contribution of acquired genomic variation to this racial disparity is largely unknown, as genomic from Africa is lacking. Here we performed the first tumor-normal paired deep whole-genome sequencing for Africa. A direct study-matched comparison between African- and European-derived, treatment-naïve, high-risk prostate tumors for 15 cases allowed for further comparative analyses of existing data. Excluding a single hyper-mutated tumor with 55 mutations per megabase, we observed a 1.8-fold increase in small somatic variants in African- versus European-derived tumors (*P*-value = 1.02e-04), rising to 4-fold when compared with published tumor-matched data. Furthermore, we observed an increase in oncogenic driver mutations in African tumors (*P*-value = 2.92e-03); roughly 30% of impacted genes were novel to prostate cancer, and 79% of recurrent driver mutations appeared early in tumorigenesis. Although complex genomic rearrangements were less frequent in African tumors, we describe a uniquely hyper-duplicated tumor impacting 149 transposable elements. Comparable to African Americans, *ERG* fusions and *PIK3CA* mutations were absent and *PTEN* loss less frequent. *CCND1* and *MYC* were frequently gained, with somatic copy number changes more likely to occur late in tumorigenesis. In addition to traditional prostate cancer gene pathways, genes regulating calcium ion-ATPase signal transduction were disrupted in African tumors. Although preliminary, our results suggest that further validation and investigation into the potential implications for elevated tumor mutational burden and tumor-initiating mutations in clinically unfavorable prostate cancer can improve patient outcomes in Africa.

Significance: The first whole genome sequencing study for high-risk prostate cancer in African men allows a simultaneous comparison of ethnic differences relative to European populations and of the influences of the environment relative to African American men.

Introduction

Prostate cancer (PCa) is the most commonly diagnosed male malignancy and second most common cause of cancer-related death in men in the developed world (1). However, racial disparity in incidence and mortality rates has been observed within the United States. African American men are 1.6 and 2.9 times more likely to be diagnosed, and 2.4 and 5.0 times more likely to die from PCa than Americans of European or Asian ancestries, respectively (2). Although socioeconomic or lifestyle factors have been suggested, after adjusting for these differences African ancestry remains a significant risk factor (3). Together with the risk associated with a family history of PCa (4) and a large heritable contribution (estimated 58%; ref. 5), a genetic basis for the greater African-ancestral disease burden cannot be ignored. Besides inherited risk, disparities in PCa outcomes suggest the influence of somatic drivers. Along with an increase in mortality rates, African American men are generally younger at diagnosis, present with higher tumor grade and volume at surgery, and are at a greater potential for metastasis (4). Within Africa, a greater disease burden has been suggested (1, 6). The Southern African Prostate Cancer Study (SAPCS) has shown that South African men present with a 2-fold increase in aggressive disease compared with age-adjusted African Americans (7).

Molecular profiling has identified PCa as a highly complex disease with substantial somatic genomic heterogeneity. While whole exome sequencing of as part of The Cancer Genome Atlas (TCGA) has identified seven major PCa subtypes, including three genes associated with recurrent mutations and four fusion events (8), correlating these subtypes

with clinical outcomes has been less obvious. Studies focused on whole genome sequencing (WGS) have highlighted the significance of interrogating the entire prostate genome, reporting an abundance of acquired complex genomic rearrangements or non-coding alterations over single nucleotide driver mutations (9, 10). Focused on men of European ancestry, more recent interrogation of African American derived primary prostate tumors include whole genome (n=24; ref. 11) and whole exome sequencing (AAPC, n=102; ref. 12). Significant differences in somatic mutational patterns observed between Americans of African versus European ancestries and reported in those studies include lack of *TMPRRSS2-ERG* gene fusions, *PTEN* loss, and mutated *PIK3CA* while suggesting novel African-enriched PCa genes, such as *CDC27-OAT* gene fusions and recurrent mutations in *ERF*. No such study has been performed within the context of Africa.

This study, to our knowledge, is the first to genome profile PCa in men from Africa, in particular, Southern Africa with the greatest within individual inherited heterogeneity (13). Focusing on a high-risk disease, it was critical that we could provide a direct and unbiased non-African comparison, factoring in sample source and pathology, sequencing technology, depth of coverage and computational pipelines. As such we generated comparable data for nine European ancestral Australians while providing further comparisons with publically available data for American men of European and African ancestries.

Materials and Methods

Patient description, genetic ancestries and ethics

African patients were recruited and consented in 2013, at time of diagnosis (biopsy core tissue), from the University of Pretoria's Steve Biko Academic Hospital, South Africa, as part of the SAPCS (7). Inclusion criteria included a histopathological Gleason score ≥8

PCa (high-risk disease) and self-identified African ancestry, classified as Southern Bantu (n=4) or South African Coloured (n=2; Table 1). African heritage was confirmed using 103,670 autosomal ancestry informative genetic markers and ancestral fractions defined as African-Bantu (Yoruba), African-KhoeSan (Ju/'hoansi), European (CEPH-Utah) and Asian (Han Chinese), as previously described (13, 14). The study was reviewed and approved by the University of Pretoria Human Research Ethics Committee (HREC #43/2010), including US Federal-wide assurance (FWA00002567 and IRB00002235 IORG0001762). Isolated DNA was shipped under the Republic of South Africa Department of Health Export Permit, in accordance with the National Health Act 2003 (J1/2/4/2 No1/12), to the Garvan Institute of Medical Research in Australia, written informed consent was obtained and the study conducted in accordance with ethical guidelines as outlined by the Declaration of Helsinki. WGS was performed in accordance with site-specific approval granted by St Vincent's Hospital HREC in Australia (SVH 15/227). The European comparative cohort included Australian patients selected based on an treatment-naive surgically removed tumor with a Gleason score ≥8, as previously described (Supplementary Table S1; ref. 15).

Whole genome sequencing and quality control

DNA was derived from fresh frozen prostate tumor tissue, derived either from biopsy core tissue at diagnosis (SAPCS) or surgical tissue at surgery ($APCRC_{EUR}$), as well as matched blood, and underwent 2×150 cycle paired-end sequencing on an Illumina HiSeq X Ten instrument (Genome.One), averaging over 80X and 45X coverage, respectively. Reads were adapter-trimmed using Illumina's Bcl2fastq Conversion software (http://www.illumina.com/) and filtered using cutadapt v1.9 (16) to remove low-quality bases (<Q15), short reads (<70bp), and missing read pairs. Remaining reads were aligned to the hg38 reference using bwa-mem v0.7.12 (17). The GATK pipeline v3.6 was used for identifying duplication, performing small (<50 bp) insertion/deletion (indel) realignment,

base quality score recalibration and co-realignment of the tumor-blood pair (18). Tumor duplication rates ranged from 8.62% to 15.25%, while for blood these ranged from 9.70% to 22.31%. Alignment statistics of the reads were calculated using QualiMap v2.1.3 (19). Stromal cell contamination was calculated using Sequenza (20). Tumor purity calculations (range 41–78%) are presented in Table 1. Quality control included interrogation of 21 somatic mitochondrial DNA mutations previously reported (Supplementary Table S2; ref. 21) and consistent size distribution of inserts from different sequencing lanes that were prepared from the same WGS library (Supplementary Table S3). Sequencing data for six SAPCS samples have been deposited in the NCBI Sequence Read Archive (accession SRP119289) with BioProject number PRJNA412953 and the following BioSample accessions: SAMN07732894 – SAMN07732905 and SAMN08985113 – SAMN08985130.

Variant calling and annotation

In this study, small variants include single nucleotide variants (SNVs) and insertions or deletions ≤50 bases (indels), while structural variants (SVs) are greater than 50 bases. Read alignment of the blood underwent FreeBayes (22) and GATK HaplotypeCaller analysis (18) to generate a consensus set of SNVs and indels. Germline variants were further assessed for known PCa risk alleles (23) and rare variants in highly penetrant genes including *BRCA1, BRCA2, HOXB13, ATM,* and *CHEK2*. All the risk variants observed were manually inspected using Integrative Genomics Viewer (IGV; ref. 24).

High-confidence somatic variants were called for each tumor-blood pair using multiple callers. Specifically, somatic SNVs and indels were called using MuTect (25), VarScan v2.4.0 (26), and Strelka v2.7.1 (27), generating a consensus callset (two or more callers). Differences between the MuTect and consensus callsets were insignificant (35,004 vs 37,975 mutations; t = -0.46; *P*-value=0.66; Supplementary Table S4). The MuTect

results of small somatic mutations were counted and divided by genome size 3,088 Mbp for tumor mutational burden. Germline SNVs and indels were further excluded, and those in known cancer-associated genes were reviewed using IGV. High confidence somatic SVs were called when detected by two or more SV callers using a previously described five-tooled MetaSV analysis (28). Inter-chromosomal translocations were identified based on identical SVs (bedtools window –w 1000) called by Lumpy v0.2.13 (29) and Manta v1.1.0 (30). Genomic rearrangements were excluded if both breakpoints were within low complexity regions, with all remaining genomic rearrangements undergoing manual inspection. Additional *de novo* genome assembly of tumor reads was performed to identify inserted sequences of somatic duplications using ABySS v1.9.0, as previously mentioned (28), and subsequently searched for repeat elements against the Repbase database (http://repeatmasker.org). Chromoplexic and chromothriptic rearrangements were tested using ChainFinder version 1.0.1 (ref. 9) and ShatterProof (31), respectively. Significant focal somatic copy number alterations (SCNAs) were identified through a comparison of segmented copy number profiles between tumor and matched blood (Supplementary Table S5), using GISTIC v2.0.22 (32), with the confidence level and q-value equal to 0.90 and 0.10, respectively. The segmentation profiles were generated using CNVkit (33) and their unsupervised hierarchical clustering, together with copy number data from the Memorial Sloan Kettering Cancer Center cohort of low- to high risk PCa (MSKCC; n=157 primary tumors), were computed as previously described (34). Percentage of genome alteration (PGA) was calculated based on the cumulative number of base pairs altered for each gain or loss per patient divided by a reference genome size.

Functional annotation and deleterious alterations of germline variants were performed using ANNOVAR (35). Functional impact, affected genes, transcripts, and other relevant features of small somatic mutations and SVs were annotated using Oncotator

v1.8.0.0 (36) and GEMINI v0.18.3 (37), respectively. Highly mutated noncoding regions were identified using LARVA v1.0, with β-binomial distribution and replication timing correction (38). Additional annotation of somatic SNVs to 30 COSMIC (Catalogue Of Somatic Mutations In Cancer) mutational signatures (39) was performed using SomaticSignatures in R package (40). Missense mutations were identified as potential oncogenic drivers using either TransFIC (http://bg.upf.edu/transfic/home) or CanDrA (41) with PCa-specific databases.

Comparative cohort analysis

Besides our comparative WGS cohort ($APCRC_{EUR}$), we scanned the literature and databases for publically available somatic mutation data for the American population with the inclusion of associated racial and pathological classification. We identified 52 men with high-risk primary PCa (Gleason score ≥8; Supplementary Table S6; as at 1 September 2017) including: $BACA_{EUR}$ WGS (n=10), and TCGA exome sequencing ($TCGA_{EUR}$; n=26 and $TCGA_{AA}$; n=3). An additional 13 African American men were excluded because their tumor DNA was extracted from formalin-fixed paraffin embedded tissues (12). Published results of the 39 patients using the MuTect calling and Oncotator annotation pipeline were compared to our genomes analyzed with the same pipeline. Only SNVs (missense, nonsense, splice site and silent mutations), indels (frameshift and in-frame mutations) and SVs (>900 bp and ≥4 reads based on read-pair evidence) available between the studies were counted and $\log_{10}$-transformed for a subsequent significance test using two-tailed t-test in R. Deleterious impact and oncogenic potential of missense (nonsynonymous) mutations were assessed using PolyPhen-2 (42) and CanDrA, respectively.

Clone phylogenetic analysis

Estimation of clonality and clonal segregation of somatic mutations was computed using PhyloWGS v1.0-rc1 (43). The somatic SNVs and indels from each patient described above were analyzed for their segregation into each clone or subclone, following the program's instruction with copy number estimation (TITAN program). The clonality was run with 10,000 Markov chain Monte Carlo samples. Matching of the copy number alterations and somatic SVs observed was assessed through the visual inspection of SV frequencies and $\log_2$ copy number ratio.

Results

Germline variation and prostate cancer risk alleles

Representing the earliest diverged contemporary human population (13), our African participants all share an ancient KhoeSan (range 2.5–41.7%) along with a predominant Bantu heritage (range 42.1–97.5%) and therefore an exasperated inherited genomic variation, averaging 5.5 million small variants (4,432,761 SNVs and 1,131,271 indels; Table 1), 1.3-fold greater than observed for the $APCRC_{EUR}$. We found the largely European-derived PCa risk alleles (7) to be fixed (6/100) or common in SAPCS (78/100 allele frequency >0.1; 48/100 ≥0.5; Supplementary Table S7) and unlikely to be risk-predicting for this population, as previously suggested (14). Missense mutations in high penetrance genes were found to be common, compared with an internal dataset of nine Southern African genomes (Supplementary Table S7), and/or previously predicted as non-oncogenic (Supplementary Table S8).

Spectrum of small somatic variation and associated racial disparities

The total number, type, and distribution of somatic mutations observed within the SAPCS are summarized in Fig. 1A. We observed a total of 227,289 putative small somatic

mutations (Supplementary Table S9). Notably, 74.6% (169,542) were from a single hyper-mutated tumor (UP2113), with a 1.8- to 3-fold within study increase in transition mutations (Supplementary Fig. S1) and a 'small' tumor mutational burden (TMB) of 55 mutations per Mbp (2,180 protein-coding). Excluding the hyper-mutated tumor, our study suggests a roughly 4-fold increase in the burden of small somatic variants than previously reported for high-risk PCa (8, 9), with a range of 3.0 to 4.7 per Mbp. Direct comparison with our high-risk $APCRC_{EUR}$ tumors confirmed a 1.8-fold increased TMB (t=-5.679; *P*-value=1.02e-04), rising to a 7.7-fold increase when considering published TMB for low to intermediate-risk PCa (10). Importantly, our comparison with $APCRC_{EUR}$ confirms the racial mutational disparity is not a technical or computational artifact. Additional racial differences (SAPCS versus $BACA_{EUR}$, $TCGA_{EUR}$ and $APCRC_{EUR}$) observed include: (*i*) the total number of protein coding variants (median 103 vs 42, 32 and 45; t=-3.314, -4.739, -2.819; *P*-values=0.005, 5.49e-05 and 1.45e-02; Fig. 1B); (*ii*) the number of predicted 'probably' or 'possibly' damaging nonsynonymous mutations using PolyPhen-2 (median 33 vs 15, 10 and 12; t=-2.698, -4.159, -2.884; *P*-values=0.017, 2.46e-04 and 1.28e-02; Fig. 1C); and (*iii*) the total number of predicted oncogenic driver mutations (median 17 vs 4, 4 and 6; t=-3.831, -4.957, -3.653; *P*-values=1.84 e-03, 3.12e-05, and 2.92e-03; Fig. 1D). Depending on the computational prediction tool applied, CanDrA or TransFIC, eight to 28 oncogenic driver mutations were observed per tumor, respectively, with two to eight overlapping driver mutations predicted between the tools. The single exception included UP2113 with 237 and 613 predicted oncogenic driver mutations, respectively, including 113 tool overlaps (Supplementary Table S10). Roughly 30% of genes mutated in the tumors were novel to PCa (COSMIC Genome Screens v82 as at 1 September 2017). We observed no impact of hyper-mutated tumor exclusion on the statistical significance (t=-3.175 to -6.055; *P*-value=7.31e-03 to 1.37e-06) and no significant differences in the number and impact of

small somatic mutations between our study and the $TCGA_{AA}$ tumors (t=-0.774 to -0.561; all *P*-values>0.46). Taken together, we suggest that the TMB, defined by small somatic variants, is significantly increased in African men with high-risk PCa compared with men of European ancestry, irrespective of country of origin.

Spectrum of structural somatic variation and associated racial disparities

Compared to small variants, greater variability was observed in the total number (11 to 312 per tumor) and type of SV acquired (Table 1 and Supplementary Table S11). Comparative analysis with $BACA_{EUR}$ required us to restrict SVs to those >900 bp and evident in >3 reads based on read-pair algorithms only. Large deletions were most common, ranging from two to 61 events per tumor, compared with $BACA_{EUR}$ (5 to 182) and $APCRC_{EUR}$ (5 to 47; Supplementary Fig. S2A). Oncogenic impact based on GEMINI high impact and MalaCards human cancer annotation ranged from zero to nine. The *TMPRSS2–ERG* fusion resulting from 3 Mbp chromosome 21 deletions was distinctly absent in the SAPCS, yet found in almost half of European-derived primary prostate tumors (8, 9), including 3/9 $APCRC_{EUR}$. Our data suggest that Europeans are more likely to acquire megabase size deletions, with a 3.5-fold increase in $BACA_{EUR}$ (t=2.255; *P*-value=0.041) and 1.12-fold increase in $APCRC_{EUR}$ using conservative calling methods (t=0.904; *P*-value=0.383; Supplementary Fig. S2B). Unique to this study, we identified a hyper-duplicated tumor (UP2133) with 234 tandem duplications. The remaining SAPCS tumors presented with two to 16 tandem duplications (median 6), significantly increased compared with $BACA_{EUR}$ (1 to 7; median 2; t=-3.699; P-value=0.003), yet only marginally increased compared with $APCRC_{EUR}$ (0 to 15; median 4; *P*-value=0.326). Sequence scaffold interrogation through the *de novo* assembly of UP2133 verified 35.9% (84/234) of these events. Despite lack of known genes in the duplicated regions, their sequence assembly

resembled 149 transposable elements (57 SINEs, 48 LINEs, 25 LTR elements and 19 DNA elements; Supplementary Table S12). Additional distinguishing features of UP2133 included: (*i*) a single chromoplexic chained fusion event involving four inter-chromosomal translocations impacting 67 genes (Supplementary Table S13) and (*ii*) elevated levels of SCNAs (11,593 gains and 13,848 losses). Compared with European studies, we found the acquisition of larger genomic rearrangements, including chromoplexy and chromothripsis (5/9 and 2/9 $APCRC_{EUR}$, respectively; Supplementary Fig. S3A and S3B) to be less frequent in our African derived tumors, consistent with lack of *ERG*-positivity.

Mutational clusters, recurrence and associated racial disparities

In an attempt to further identify genomic hallmarks or drivers of racial disparities in lethal PCa, we performed mutational cluster and recurrence analyses. Our tumors fell into three main COSMIC (Catalogue of Somatic Mutations in Cancer) clusters (Fig. 2A). Specifically, hyper-mutated UP2113 showed a predominance of somatic mutational Signature 6, while UP2003 clustered with tumors corresponding to the largest contribution of Signature 3, with a notable under representation of Signature 1 common to all cancer subtypes. Signatures 3 and 6 have been associated with failure of DNA mismatch repair responses, which may explain the observed hyper-mutation and upper level for large deletions observed for UP2003. The remaining tumors, including hyper-duplicated UP2133, grouped with the most prominent mutational patterns attributed to unknown causes, Signatures 5, 8 and unknown (46–76%). Of 18,080 genes acquiring small mutations (Supplementary Table S9), 46% (8,333) were mutated in more than one SAPCS tumor and 2% (355) in all tumors. Two tumors acquired mutations within TCGA taxonomic genes, specifically *SPOP* and *FOXA1* (UP2003 and UP2113, respectively; Table 1). Out of 137 reported genes significantly mutated or driving PCa in approximately 1,000 tumors (44,

45), 32 were also identified in African patients with reported mutation frequencies ranging from 18.7% (*TP53*) to as low as 0.7% (*HRAS, KRAS, TAP1*, and *WASF3*; Fig. 2B). Among them, *CDH1* (0.9%), *CHD7* (1.5%) and *KMT2A* (1.7%) were recurrent in two tumors studied. To expand the recurrent gene set, machine learning of a curated PCa database training set identified an additional 14 genes within SAPCS tumors as recurrent nonsynonymous predicted oncogenic drivers (Supplementary Fig. S4), with *TTN, PABPC3*, *RAVER2*, and *RIF1* also mutated in African Americans (8, 12). Recurrence of small somatic mutations (single nucleotide or indel) impacting genes, such as *FOXA1*, *SPOP, FLG2*, *HSPG2*, *KMT2A*, and *MUC16* concurs with African American studies (8, 11, 12), while large deletions (over 50 bp in length) impacting *BRCA2*, *DEFA1B* and *MFF* were recurrent in our study. Noncoding regions significantly mutated (corrected *P*-values=1.81e-02–5.20e-08) included: previously reported transcription factor binding sites (TFBS) of *CTCF*, *FOXA1* or *MYC* (8, 9) and the promoter of *SMARCA5* (39) and recurrent mutations within the promoters of *PRPSAP2* and *MFF*, TFBS of *GABPA* or *FOXA1,* and DNase I hypersensitive site (DHS) of *MIR8065* or *RBFOX1*. Besides the traditional gene pathways disrupted in PCa, namely AR, PI3K, WNT, p53 and DNA repairing, we identified an abundance of genes implicated in Calcium ion-ATPase signal transduction (Supplementary Table S14). Our study implies that abnormal Calcium-mediated cell death (46) may represent a novel pathway of prostate tumorigenesis.

Recurrent SCNAs revealed 61 gains and 30 losses (Supplementary Table S15), the most significant including gains (q-value<1.82e-4) at 1p36.13 (*NBPF1*), 4p16.3 (*ZNF595* and *ZNF718*) and 9q13 (*LOC442421*) and a loss at 8p23.1 (q-value=4.16e-08; *DEFB130*, *DEFB134* and *DEFB135*). Comparing the magnitude of this copy number with those reported for low- to high-risk PCa (MSKCC; PGA range 0-41%, median 2%) suggested the same two major subtypes, minimal and substantial SCNAs, with all SAPCS tumors

clustered in the latter (Supplementary Fig. S5). Their percentage of genome alteration was high (range 18–34%) and their 276 reported DNA signature genes for PCa was highly altered (range 60–196 genes), but not in UP2113 (3% and 29/276; Supplementary Table S16). The subtype and PGA that impacts specific genes has provided a disadvantage of prognosis in 5-year biochemical relapse-free survival (34, 47) and metastasis within study SAPCS patients (48).

Tumor evolution

Clone phylogenetic analyses showed that most SNVs and indels clustered at the clonal, rather than subclonal stage of tumor evolution (Fig. 3A), with an affected cancer cell fraction (CCF) estimated to be 100% (t=4.009; *P*-value=9.09e-4; Fig. 3B) and a likelihood for tumor initiation (Tajima's *D*=-2.23*; P*-value=0.01; Supplementary Fig. S6). Among the 14 recurrent oncogenic drivers described above, 11 were mutated at the clonal stage. The noncoding mutations that were highly mutated and showed high mutant allele frequency were also found early in their corresponding phylogenies. A similar observation was seen with the *SPOP* missense mutation in UP2003, a previously described TCGA PCa molecular subtype (8). Although 8p23.1 losses within four of the six tumors were found to be clonal, most copy number alterations were at the subclonal stage with CCF ranging from 20% to 60% (t=-2.389; *P*-value=0.03; Fig 3C). Notably, recurrent gains at 1p36.13, 4p16.3, and 9q13 were subclonal among all the tumors analyzed.

Discussion

To our knowledge, this is the first study to use deep WGS to profile genomic variation in men with lethal primary PCa from Africa. The significance of this study lies not only in the extent of naturally occurring genome variation within the study population,

averaging a million more variants than Europeans, but importantly the significance of previously described aggressive disease presentation (7), allowing for comparable analyses for genomic signatures determining racial disparities observed within the United States. Future African-based studies would also advance our understanding of how the contribution of tumor initiating inherited variation, along with environmentally induced epigenetic changes, drives tumor evolution and associated heterogeneity between racial/ethnic groups (49).

Studies characterizing somatic mutations in primary PCa from African American men have focused on intermediate- to high-risk disease (Gleason grade ≥7) reporting a similar overall burden of small somatic variants (defined in this study as TMB) as observed for men of European ancestry (11, 12). Although our numbers are small, we observe an elevated TMB compared with European men, averaging a 7.7-fold increase for low- to intermediate-risk disease (10) and 4-fold for more aggressive disease presentation (8, 9, 50). Compared with the African American studies, differences from our study include formalin fixation of sample sources (12) or stringency of mutant calling associated with the Complete Genomics Inc. sequencing technology (11), which may, at least in theory, have contributed to the observed elevated TMB in our African samples. Furthermore, the number of small potentially impactful oncogenic driver mutations appears to be significantly increased, with further suggestive evidence that these events initiate tumorigenesis. While European-based studies have reported abundance of acquired complex genomic rearrangements or non-coding alterations over single nucleotide driver mutations (9, 10), *TMPRSS2-ERG* or any *ERG* fusions are absent in our study, while chromoplexic or chromothriptic chained fusion events are lacking. It therefore waits to be validated if large genomic rearrangements are a hallmark of the observed racial disparity, being common to non-African over African derived PCa.

Excluded from our calculated averages, we describe the most hyper-mutated PCa genome, exceeding current reports for localized (maximally 1,068 protein-coding; ref. 51) or metastatic disease (maximally 1,527 protein-coding; ref. 52). Observing copy number losses in mismatch DNA repair genes, *MSH2* ($\log_2$ copy number = -0.75) and *MSH6* (-0.75), as well as an inter-chromosomal translocation between *MSH6* intron 1 (chromosome 2) and chromosome 15, we speculate that alterations in the DNA repair pathway may have contributed to the hyper-mutation (51, 52). Alterations in DNA repair genes do not, however, explain the increased overall TMB observed, with ~70% of the tumors in our study clustering within cancer mutational signatures of unknown cause. Traditionally, PCa has one of the lowest reported TMBs of adult cancers (53) and no known modifiable prostate carcinogenic mutagen (54). This study raises an important question if as yet unknown environmental/carcinogenic and/or genetic factors are contributing to increased TMB in men from Africa.

Although this study is limited by study size, strengths include; (*i*) the first genome sequencing for PCa within men from Africa, (*ii*) patients representing genetically the most diverse ancestral human lineages known to date (including significant KhoeSan ancestral contributions), (*iii*) treatment-naïve tumors, (*iv*) access to fresh tissue minimising formalin fixation associated DNA degradation and subsequent sequencing biases, (*v*) deep whole genome sequencing allowing for confidence in somatic variant calling within a highly heterogenous and non-tumor contaminated environment, and (*vi*) direct pathological, technical and most importantly computationally matched comparative analyses with a European-derived sample source. One should acknowledge additional contrasts to non-African data including; as African patients were recruited at time of diagnosis, it is unknown if metastasis had occurred, while baseline PSA levels are significantly increased within the Black South African population, reporting in a cohort of 302 men pathologically

excluded for prostate cancer 75.4% with PSA levels ≥ 4 µg/L, of which 61.4% presented with PSA levels ≥ 20 µg/L (7).

Given the significant increase in the burden of small somatic mutations (or TMB) among Africa-derived tumors, averaging a 1.8-fold increase with direct African versus European study comparison, one may speculate (especially with clear mismatch repair defects) that higher TMB would predict for responsiveness to immunotherapy within African men with high-risk PCa, as seen for diverse cancer types (55). The potential benefits for immunotherapeutics for advanced PCa within Africa, however, calls for larger study validation, both within South Africa and across the continent. This study highlights the critical need for further African-ancestral studies, not only to elucidate the underlying contribution to racial disparities but importantly for much lacking global inclusivity as we move into an era of precision medicine (56, 57).

Acknowledgments

We are eternally grateful to the patients for their participation, as well as the nurses and urology staff for recruiting and maintaining the Southern African Prostate Cancer Study (SAPCS). We thank the participants and colleagues who contributed to the Southern African Ubuntu Genome Reference Project and the Australian Prostate Cancer Research Centre (APCRC) New South Wales Biobank. We are grateful to The Cancer Genome Atlas (TCGA) and Baca and co-authors (2015), who made their data freely available, and acknowledge the University of Sydney Informatics Hub for providing access to the High Performance Computing infrastructure used in this study.

Grant Support

This work was supported in part by the Cancer Association of South Africa (CANSA) and National Research Foundation (NRF) South Africa to M.S.R. Bornman and V.M. Hayes, University of Limpopo Research Fund to P.A. Venter and V.M. Hayes, as well as the Garvan Foundation Rod Will Great Xpectations Fundraiser, Australia, to V.M. Hayes. W. Jaratlerdsiri was supported by the Australian Prostate Cancer Research Centre NSW (APCRC-NSW) Australia, and D.C. Petersen and E.K.F. Chan by Movember Australia and the Prostate Cancer Foundation Australia (PCFA) as part of the Movember Revolutionary Team Award (MRTA) to the Garvan Institute of Medical Research program on prostate cancer bone metastasis (ProMis) with V.M. Hayes as one of the Team Leads. V.M. Hayes is supported by the University of Sydney Foundation and Petre Foundation, Australia.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015; 136:E359-E86.

2. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA Cancer J Clin. 2014; 64:9-29.

3. Park SY, Haiman CA, Cheng I, Park SL, Wilkens LR, Kolonel LN, et al. Racial/ethnic differences in lifestyle-related factors and prostate cancer risk: the Multiethnic Cohort Study. Cancer Causes Control. 2015; 26:1507-15.

4. Powell IJ. The precise role of ethnicity and family history on aggressive prostate cancer: a review analysis. Arch Esp Urol. 2011; 64:711-9.

5. Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, et al. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. JAMA. 2016; 315:68-76.

6. Adeloye D, David RA, Aderemi AV, Iseolorunkanmi A, Oyedokun A, Iweala EE, et al. An Estimate of the Incidence of Prostate Cancer in Africa: A Systematic Review and Meta-Analysis. PLoS One. 2016; 11:e0153496.

7. Tindall EA, Monare LR, Petersen DC, van Zyl S, Hardie RA, Segone AM, et al. Clinical presentation of prostate cancer in black South Africans. Prostate. 2014; 74:880-91.

8. The-Cancer-Genome-Atlas-Network. The molecular taxonomy of primary prostate cancer. Cell. 2015; 163:1011-25.

9. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated evolution of prostate cancer genomes. Cell. 2013; 153:666-77.

10. Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, et al. Genomic hallmarks of localized, non-indolent prostate cancer. Nature. 2017; 541:359-64.

11. Lindquist KJ, Paris PL, Hoffmann TJ, Cardin NJ, Kazma R, Mefford JA, et al. Mutational landscape of aggressive prostate tumors in African American men. Cancer Res. 2016; 76:1860-8.

12. Huang FW, Mosquera JM, Garofalo A, Oh C, Baco M, Amin-Mansour A, et al. Exome Sequencing of African-American Prostate Cancer Reveals Loss-of-Function ERF Mutations. Cancer Discov. 2017.

13. Petersen DC, Libiger O, Tindall EA, Hardie RA, Hannick LI, Glashoff RH, et al. Complex patterns of genomic admixture within southern Africa. PLoS Genet. 2013; 9:e1003309.

14. Hayes VM, Bornman MSR. Prostate Cancer in Southern Africa: Does Africa Hold Untapped Potential to Add Value to the Current Understanding of a Common Disease? J Glob Oncol. 2017.

15. Kalsbeek A, Chan EFK, Grogan J, Petersen DC, Jaratlerdsiri W, Gupta R, et al. Mutational load of the mitochondrial genome predicts pathology features and biochemical recurrence in prostate cancer Aging (Albany NY). 2016; 8:2702-12.

16. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2015; 17:10-2.

17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics. 2009; 25:1754-60.

18. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013; 11:11.0.1-33.

19. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 2016; 32:292-4.

20. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. Ann Oncol. 2015; 26:64-70.

21. McCrow JP, Petersen DC, Louw M, Chan EK, Harmeyer K, Vecchiarelli S, et al. Spectrum of mitochondrial genomic variation and associated clinical presentation of prostate cancer in South African men. Prostate. 2016; 76:349-58.

22. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint 2012:arXiv:1207.3907 [q-bio.GN]

23. Al-Olama AA, Kote-Jarai Z, Berndt SI, Conti DV, Schumacher F, Han Y, et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. Nat Genet. 2014; 46:1103-9.

24. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14:178-92.

25. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013; 31:213-9.

26. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012; 22:568-76.

27. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics. 2012; 28:1811-7.

28. Jaratlerdsiri W, Chan EKF, Petersen DC, Yang C, Croucher PI, Bornman MSR, et al. Next generation mapping reveals novel large genomic rearrangements in prostate cancer. Oncotarget. 2017; 8:23588-602.

29. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014; 15:R84.

30. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: Rapid detection of structural variants and indels for clinical sequencing applications. Bioinformatics. 2016; 32:1220-2.

31. Govind SK, Zia A, Hennings-Yeomans PH, Watson JD, Fraser M, Anghel C, et al. ShatterProof: operational detection and quantification of chromothripsis. BMC Bioinformatics. 2014; 15:78.

32. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011; 12:R41.

33. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput Biol. 2016; 12:e1004873.

34. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010; 18:11-22.

35. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38:e164.

36. Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, et al. Oncotator: cancer variant annotation tool. Hum Mutat. 2015; 36:E2423-E9.

37. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: Integrative Exploration of Genetic Variation and Genome Annotations. PLoS Comput Biol. 2013; 9:e1003153.

38. Lochovsky L, Zhang J, Fu Y, Khurana E, Gerstein M. LARVA: an integrative framework for large-scale analysis of recurrent variants in noncoding annotations. Nucleic Acids Res. 2015; 43:8123-34.

39. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017; 45:D777-D83.

40. Gehring JS, Fischer B, Lawrence M, Huber W. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. Bioinformatics. 2015; 31:3673-5.

41. Mao Y, Chen H, Liang H, Meric-Bernstam F, Mills GB, Chen K. CanDrA: cancer-specific driver missense mutation annotation with optimized features. PLoS One. 2013; 8:e77945.

42. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013; Chapter 7:Unit7.20.

43. Deshwar AG, Vembu S, Yung CK, Jang GH, Stein L, Morris Q. PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. Genome Biol. 2015; 16:35.

44. Wedge DC, Gundem G, Mitchell T, Woodcock DJ, Martincorena I, Ghori M, et al. Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. Nat Genet. 2018; 50:682-92.

45. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, et al. The long tail of oncogenic drivers in prostate cancer. Nat Genet. 2018; 50:645-51.

46. Orrenius S, Zhivotovsky B, Nicotera P. Regulation of cell death: the calcium-apoptosis link. Nat Rev Mol Cell Biol. 2003; 4:552-65.

47. Lalonde E, Ishkanian AS, Sykes J, Fraser M, Ross-Adams H, Erho N, et al. Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year

biochemical recurrence of prostate cancer: a retrospective cohort study. Lancet Oncol. 2014; 15:1521-32.

48. Hieronymus H, Schultz N, Gopalan A, Carver BS, Chang MT, Xiao Y, et al. Copy number alteration burden predicts prostate cancer relapse. Proc Natl Acad Sci U S A. 2014; 111:11139-44.

49. Easwaran H, Tsai HC, Baylin SB. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. Mol Cell. 2014; 54:716-27.

50. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko A, et al. The genomic complexity of primary human prostate cancer. Nature. 2011; 470:214-20.

51. Kumar A, Coleman I, Morrissey C, Zhang X, True LD, Gulati R, et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nat Med. 2016; 22:369-78.

52. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, et al. Integrative clinical genomics of advanced prostate cancer. Cell. 2015; 161:1215-28.

53. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499:214-8.

54. Leitzmann MF, Rohrmann S. Risk factors for the onset of prostatic cancer: age, location, and behavioral correlates. Clin Epidemiol. 2012; 4:1-11.

55. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, et al. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. Mol Cancer Ther. 2017; 16:2598-608.

56. Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, Szolovits P, et al. Genetic Misdiagnoses and the Potential for Health Disparities. N Engl J Med. 2016; 375:655-65.

57. Spratt DE, Chan T, Waldron L, Speers C, Feng FY, Ogunwobi OO, et al. Racial/Ethnic Disparities in Genomic Sequencing. JAMA Oncol. 2016; 2:1070-4.

Table 1. Clinical presentation, patient ancestry, sequencing statistics and genomic variation

	UP2003	UP2039	UP2099	UP2113	UP2116	UP2133
Ethnicity	Coloured	Coloured	Zulu	Pedi	Swati	Zulu
Clinical presentation at diagnosis						
Age (years)	76	71	76	88	99	58
PSA levels (ng/ml)	3459	319	100	123	65	100
Gleason score	9	8	8	8	10	9
Family history PCa	yes	no	no	no	no	no
Ancestral fractions[a]						
African: Bantu/KhoeSan	0.421/0.417	0.622/0.378	0.868/0.132	0.944/0.056	0.975/0.025	0.868/0.132
Non-African: Eur/Asian	0.138/0.024	0	0	0	0	0
WGS statistics						
Blood/Tumor coverage	43X/86X	40X/75X	42X/82X	58X/86X	41X/87X	43X/81X
Tumor purity	0.42	0.68	0.78	0.41	0.44	0.43
Germline variations (total number)						
SNVs	4,447,982	4,463,245	4,459,492	4,361,491	4,398,876	4,465,481
ins	535,378	547,399	539,284	568,665	533,522	569,188
del	569,071	580,344	572,637	600,646	565,696	605,795
Somatic variants (total number)[b]						
SNV	11,463	8,017	7,939	138,990	6,946	11,463
ins	1,006	844	1,077	7,401	1,003	1,471
del	1,760	1,182	1,362	23,151	1,326	1,448
Nonsynonymous	95	52	42	1,507	56	79
TMB per Mbp	4.6	3.3	3.4	55	3	4.7
SV-DEL (bases impacted)	124 (296,889,562)	7 (31,812,041)	46 (198,385,225)	27 (7,785,288)	56 (125,131,631)	54 (315,444,459)
SV-DUP (bases impacted)	16 (310,358,766)	2 (50,355)	9 (134,176,076)	7 (13,696,461)	6 (98,051,676)	234 (320,826,012)
SV-INV (bases impacted)	12 (262,977,083)	1 (287,931)	6 (500,592,107)	0 (0)	2 (79,698,508)	4 (3,017,735)
SV-CTX	19	1	17	1	2	20
Chained event[c]	0	0	0	0	0	1 (4)
SCNA-Gains/Losses	2,325/3,979	3,466/6,184	6,759/4,784	1,863/2,524	3,009/5,130	11,593/13,848
TCGA molecular subtype[d]						
Classification (mutation)	*SPOP* (F133I)	UC	UC	*FOXA1* (D226G)	UC	UC
Oncogenic driver mutations						
SNV: TransFIC/CanDrA[e]	28/21(8)	18/11(2)	15/11(5)	613/237(113)	8/20(4)	25/14(6)
SV-DEL: High Impact[f]	21(9)	2(0)	16(7)	1(1)	12(8)	16(6)
SCNA recurrent focal amplifications						
CCND1 (11q13.3)	high gain	het loss	some gain	high gain	some gain	high gain
MYC (8q24.21)	high gain	some gain	high gain	diploid	high gain	high gain
SCNA recurrent focal deletions						
PTEN (10q23.31)	diploid	het loss	some gain	diploid	some gain	het loss
TP53 (17p13.1)	het loss	het loss	some gain	diploid	some gain	hom del
FANCD2 (3p25.3)	het loss	het loss	diploid	diploid	diploid	hom del
SPOPL (2q22.1)	hom del	het loss	some gain	high gain	het loss	diploid
MAP3K7 (6q15)	het loss	some gain	het loss	diploid	hom del	het loss
CHD1 (5q15)	hom del	het loss	some gain	diploid	some gain	het loss

[a] Ancestral fractions based on STRUCTURE analysis; [b] Number of SNVs/indels detected using MuTect and that of SVs (hg38) from high-confidence results of somatic callers; [c] In brackets is the number of inter-chromosomal translocations involved; [d] A total of seven TCGA subtypes; [e] overlapping results in brackets; [f] Oncogenic potentials reported in MalaCard in brackets. **ABBREVIATIONS:** PSA, prostate specific antigen; ins, insertion <50 bp; del, deletion <50 bp; TMB, small Tumor Mutational Burden; DEL, deletion >50 bp; DUP, duplication; INV, inversion; CTX, inter-chromosomal translocation; SCNA, somatic copy number alteration; UC, unclassified; het loss, heterozygous loss; hom del, homozygous deletion

FIGURE LEGENDS

Figure 1. Spectrum of somatic genomic variation in high-risk PCa in African and European men. **A,** Circos plots depicting total tumor mutational burden within six African-derived SAPCS tumors. The outermost (first) track represents autosome ideograms (chromosomes 1 to Y, centromeres as red line), with the pter-qter orientation in a clockwise direction (length in Mbp). The second track represents somatic copy number gains (red) and losses (blue). The third and fourth tracks show histograms of the number of somatic SNVs and indels observed per Mbp, respectively. Vertical scales of these two tracks are 0 – 25 variants per Mbp, except for 0 – 130 in the UP2113 circos plot. The innermost area shows breakpoints of acquired genomic rearrangements (>900 bp), with their links colored according to the following variant types: deletions (red), tandem duplications (blue), inversions (yellow), and interchromosomal translocations (black). Somatic variation identified in the SAPCS was further compared with European ($BACA_{EUR}$, $TCGA_{EUR}$ and $APCRC_{EUR}$) and African American cohorts ($TCGA_{AA}$) with high-risk PCa for, **B**, protein coding mutations, **C**, predicted deleterious mutations using PolyPhen-2, and **D**, oncogenic driver mutations classified using a machine learning algorithm and COSMIC true datasets in the CanDrA package. A single or double asterisk indicates a significant *P*-value less than 0.05 or 0.01, respectively. Red diamonds suggest mean values. Horizontal lines in a box plot (top to bottom) indicate the third quartile, median, and first quartile, respectively.

Figure 2. Somatic mutational signatures and mutational recurrence in African high-risk PCa. **A,** Somatic signature annotation for the high-risk prostate tumors, each of which has the frequency of mutations corresponding to different COSMIC signatures (1 – 30) assigned and defined as African (SAPCS, black square), African American (TCGA, black circle) or European (remainder). **B,** SAPCS tumors versus existing significantly mutated genes or driver genes among a thousand prostate cancers (44, 45). Percentage of the mutated genes is calculated using total number of mutations divided by number of tumors (n=1,013).

Figure 3. Predicted SAPCS tumor evolution. **A,** Phylogenies revealing relationships between sample clone and subclone and their segregation of somatic mutations for each African SAPCS patient. Arrows indicate the number of small somatic mutations and copy number alterations (CNA) at either clonal or subclonal stage. Cancer cell fractions (CCF = 0 – 1) for each clone or subclone are shown in gray circles. Abbreviations include: DHS, DNase I hypersensitive site; TFBS, transcription factor binding site; UTR, untranslated region; and CN, copy number. Significant differences of molecular aberrations acquired at the clonal and subclonal stage, **B,** for the number of somatic SNVs and indels and **C,** for the number of copy number alterations.

Author Manuscript Published OnlineFirst on September 14, 2018; DOI: 10.1158/0008-5472.CAN-18-0254
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 1

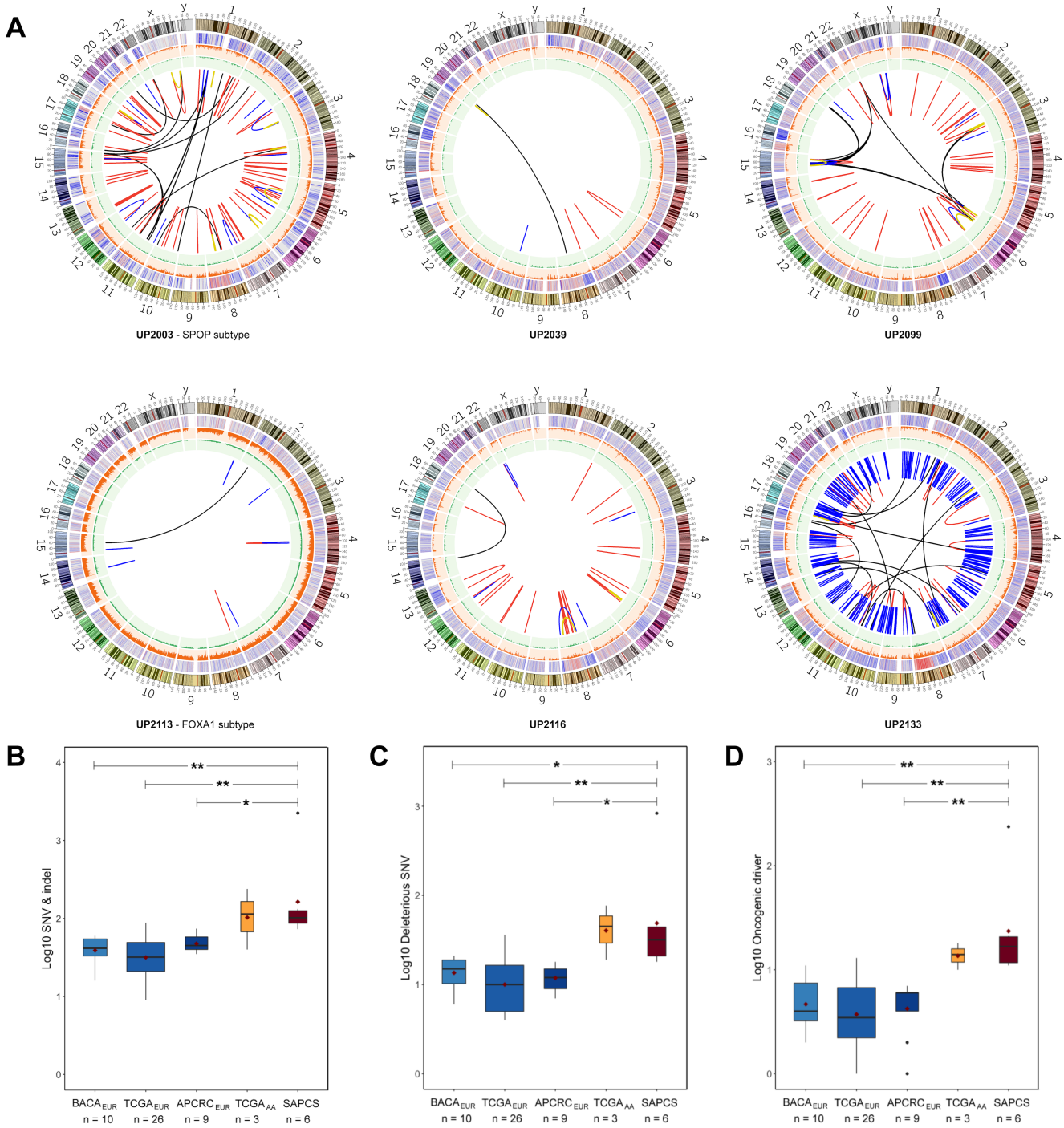


Figure 2

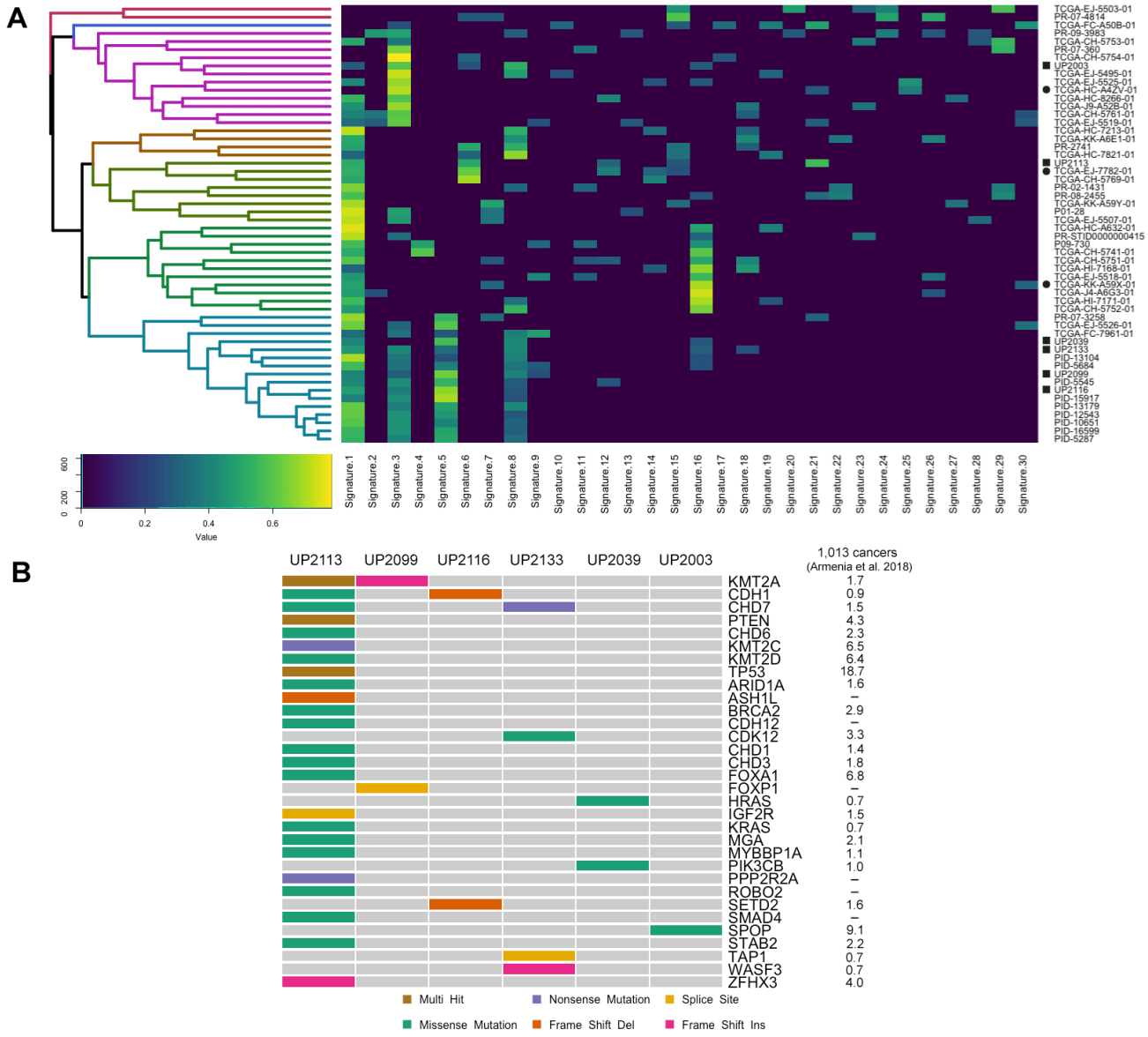
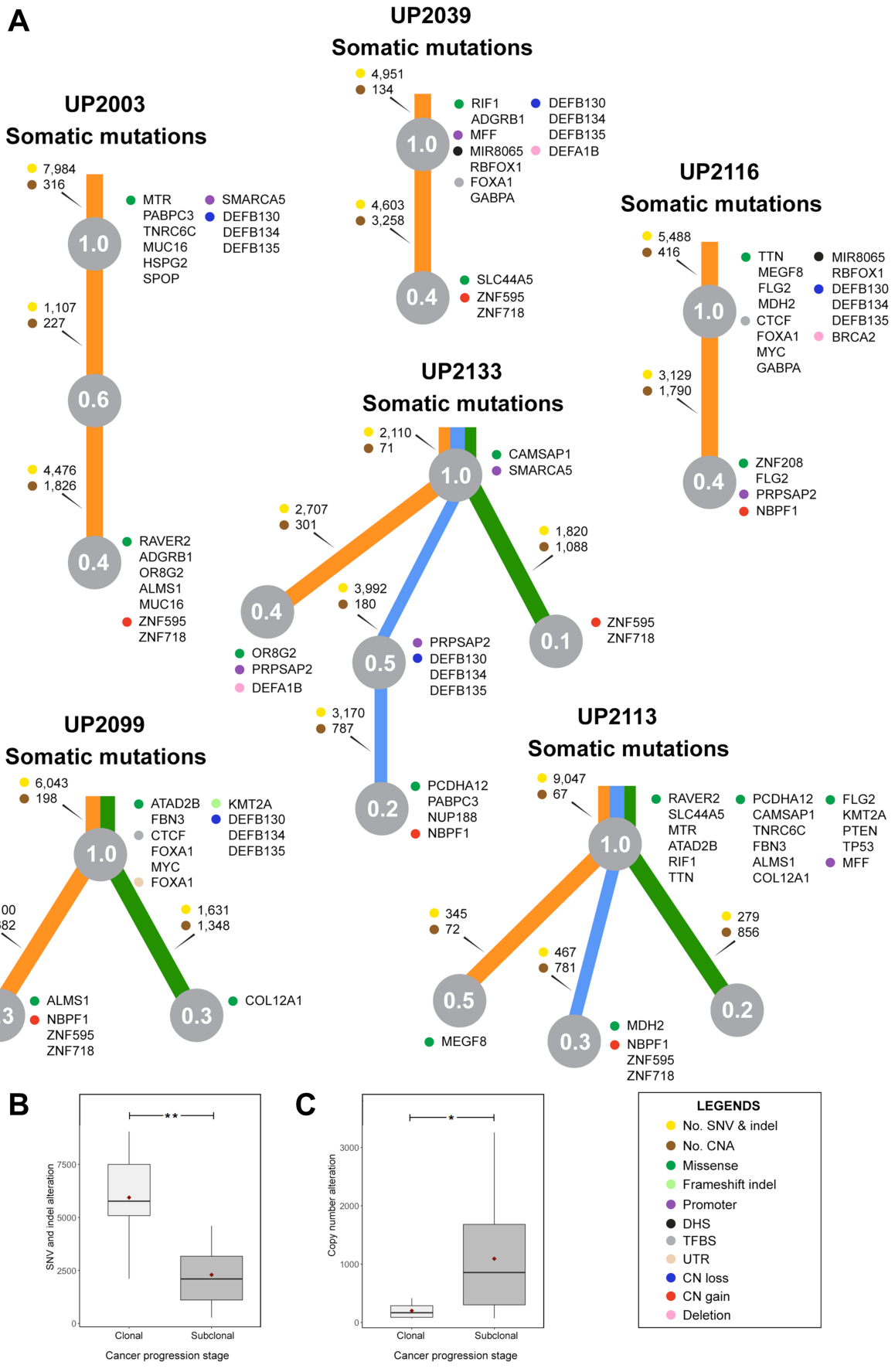


Figure 3



AACR American Association for Cancer Research

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Whole Genome Sequencing Reveals Elevated Tumor Mutational Burden and Initiating Driver Mutations in African Men with Treatment-Naive, High-Risk Prostate Cancer

Weerachai Jaratlerdsiri, Eva K.F. Chan, Tingting Gong, et al.

Cancer Res Published OnlineFirst September 14, 2018.