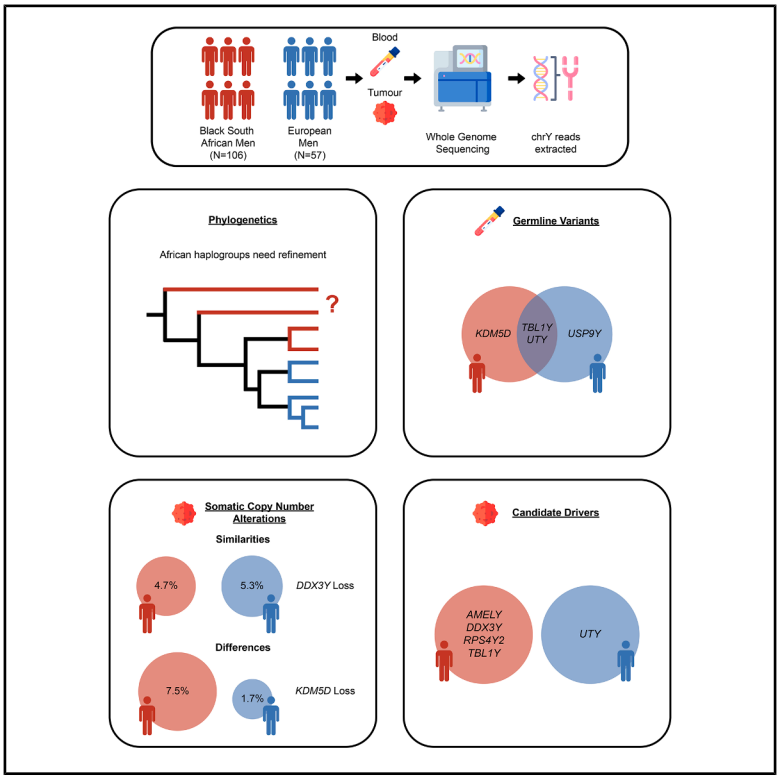
iScience

Article

Y chromosome variation and prostate cancer ancestral disparities

Graphical abstract



Highlights

- Phylogenetic analysis showed Y-haplogroup refinement is needed among African lineages
- Different potentially deleterious germline SNVs in *TBL1Y/UTY* in both populations
- Somatic CN loss in tumor suppressor *DDX3Y* in both African and European tumors
- Treatment resistance-associated *KDM5D* somatic CN loss increased in African tumors

Authors

Pamela X.Y. Soh, Alice Adams, M. S. Riana Bornman, ..., Shingai B. A. Mutambirwa, Weerachai Jaratlerdsiri, Vanessa M. Hayes

Correspondence

vanessa.hayes@sydney.edu.au

In brief

Human genetics; Sequence analysis; Cancer

Soh et al., 2025, iScience *28*, 112437
May 16, 2025
https://doi.org/10.1016/j.isci.2025.112437

CellPress

iScience

CellPress
OPEN ACCESS

Article

Y chromosome variation and prostate cancer ancestral disparities

Pamela X.Y. Soh,[1] Alice Adams,[1,2] M.S. Riana Bornman,[3] Jue Jiang,[1] Phillip D. Stricker,[4] Shingai B.A. Mutambirwa,[5] Weerachai Jaratlerdsiri,[1] and Vanessa M. Hayes[1,3,6,7,*]

[1]Ancestry and Health Genomics Laboratory, Charles Perkins Centre, School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Camperdown, NSW 2006, Australia
[2]Faculty of Science, University of Bath, BA2 7AY Bath, UK
[3]School of Health Systems and Public Health, University of Pretoria, Pretoria, South Africa
[4]St Vincent's Prostate Cancer Research Centre, Sydney, NSW 2010, Australia
[5]Department of Urology, Sefako Makgatho Health Science University, Dr George Mukhari Academic Hospital, Medunsa, Ga-Rankuwa, South Africa
[6]Manchester Cancer Research Centre, University of Manchester, M20 4GJ Manchester, UK
[7]Lead contact
*Correspondence: vanessa.hayes@sydney.edu.au
https://doi.org/10.1016/j.isci.2025.112437

SUMMARY

Prostate cancer (PCa) is marked by significant ancestral bias, with African men disproportionately impacted. However, genome profiling studies have yet to explore the mutational landscape and disparity contribution of the male-determining Y chromosome. Using a cohort of 106 African and 57 European PCa cases, biased toward aggressive presenting primary disease, we performed complete Y chromosome interrogation for inherited and somatic variance. Capturing unexplored early-diverged Y-haplogroup substructure, while European men are 3.1-fold more likely to present with a rare potentially deleterious germline variant, a higher proportion of African patients acquired Y chromosome tumorigenic events (26.4% African, 14% European). While somatic copy number alterations were universally more common to aggressive tumors, besides shared alterations impacting *DDX3Y* and *USP9Y*, African derived tumors were prone to somatic losses associated with *KDM5D*, *PCDH11Y*, and *RBMY*. This much-needed African inclusive study alludes to possible Y chromosome contribution, at least in part, to treatment resistance and worsened mortality rates in African men.

INTRODUCTION

The human Y chromosome (chrY) is small and gene-deficient, while rich in repeat elements and segmental duplications, with chromosomal-wide copy number variation (CNV).[1] As a consequence of extensive structural complexity, chrY is notoriously difficult to sequence,[2,3] and largely ignored in genome-wide disease interrogation studies.[4] Except for the pseudoautosomal region (PAR),[5] the lack of recombination leads to Y-haplotypes, including haplogroup-specific Y-CNVs,[6,7] which are passed on from father to son as patrilineages with high geographic and ethnic specificity.[8] In a similar fashion, prostate cancer (PCa) is highly heritable,[9] such that a diagnosed father and brother with PCa doubles familial associated risk,[10] while ancestry is a significant risk factor for PCa presentation, age at onset, and associated mortality.[11] The commonality of inheritance and ethnic specificity provides the rationale for further interrogation of the "underappreciated" male chromosome.

Typically excluded from genome-wide PCa association studies,[12] targeted chrY explorations have provided a glimpse into geo-ethnic specific associations. While data from men of European, Ashkenazi Jewish,[13] African American,[14] and Korean[15] ancestry showed no significant association between Y-haplotypes and PCa, conversely, studies have identified increased PCa risks for Y-haplogroup DE (DE-M145; Japanese),[16] O3 (O-M122; Japanese),[14] I1c (I-Z17943; Swedish),[17] and microvariant alleles of short tandem repeats in DYS388 (Malaysian), DYS439 (Malaysian),[18] and DYS458 (Portuguese).[19] Additionally, extensive genealogical data from Utah in the United States has identified 7.3% of 1000 unique Y chromosomes to be associated with high risk of PCa.[20] Although presenting with the highest global mortality rates,[21] data from Africa is lacking. In prostate tumor cell lines, chrY gene loss is common in high grade LNCaP and PC3,[22,23] while loss of the entire chrY, a frequent event among cancer types, is rare in PCa but is associated with poor progression-free survival.[24,25] However, the specific contributors of inherited and acquired chrY variation between ethnicities remains unclear.

From our previously generated whole genome data focused on autosomal interpretation,[26] here, we compared blood and tumor chrY data between 106 African and 57 European men with histopathologically confirmed treatment naive primary PCa. Biased toward aggressive disease, patient inclusion required a confirmed genetic ancestry as non-admixed, with data

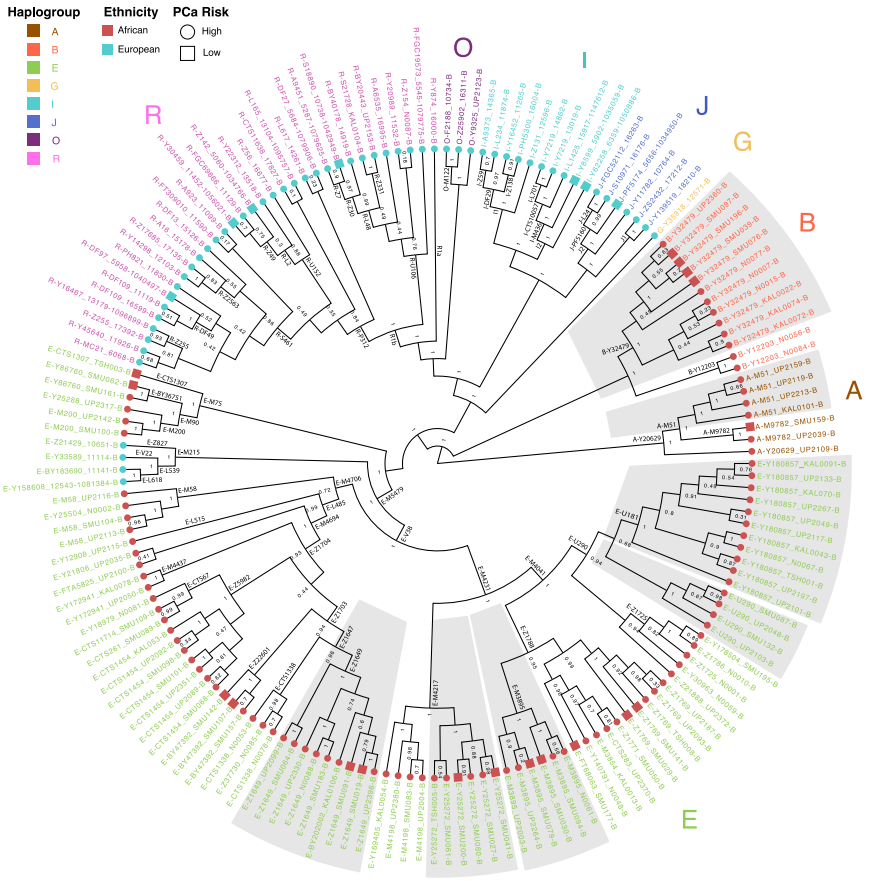


Figure 1. Phylogenetic tree of Y chromosome sequences (midpoint-rooted)

Ethnicities are indicated by the tip colors (light blue = European, red = African), with the tip shape indicating high or low risk prostate cancer (square = low risk, circle = high risk), and sample names are colored according to predicted haplogroups. Values on the branches indicate bootstrap values. Lineages shaded in gray boxes indicate regions of the tree where samples with the same predicted haplogroups may have uncaptured sublineages as there is high bootstrap support for their split.

generation and analysis reliant on a single technical and informatic pipeline. As such, we minimized for inter-study sequencing associated artifacts, while captured through our focus on southern African men, not only the highest global PCa mortality rates,[21] but also the greatest genetic diversity.[27]

RESULTS

ChrY phylogenetics and uncaptured African ancestral sublineages

Phylogenetic analysis of our patient cohort showed broad Y-haplogroup representation including the African predominant E (81.1%, 86/106) and specific A and B, and European-specific R (58%, 33/57), G, I, J and O (Figure 1; Table S1). It was not surprising that our southern African subjects presented with the earliest diverged Y-haplogroups A and B, representing the root-region for contemporary modern humans' paternal evolutionary tree.[7] Furthermore, notable within-chrY phylogenetic sub-branching suggests significant uncaptured Y-haplogroup diversity across southern Africa. This includes four (of seven) currently unknown A-M51, or International Society of Genetic Genealogy (ISOGG) Y-DNA database (https://isogg.org/tree/, version 15.73, last revised 11 July 2020) A1b1b2a sequences, which splits into two additional nodes (bootstrap value = 1, 4 individuals) from A-M9782 (bootstrap value = 1). Eleven (of 13) currently unknown B-Y32479 sequences, a subgroup of B-M150 (descending from ISOGG B2a1a1a1a2), split into two major groups (bootstrap value = 1, 5 and 6 individuals each) with further sub-branching. Haplogroup E showed high bootstrap support for uncaptured African-specific sub-branches within E-Z1649 (9 individuals), E-M4217 (6), E-M3895 (6), E-U290 (4) and E-U181 (11).

Overall, African over European men had a significantly greater number of chrY single nucleotide variants (SNVs, mean 1010 vs. 332) and small insertions and deletions (indels, 106 vs. 53.8), which mirrors the autosomal means (4296320 vs. 3432990 and 442130 vs. 353398, respectively) (Figure 2A; Wilcoxon rank-sum test, $p < 2.22e\text{-}16$ for all). However, unlike autosomal whole genome data where Africans had greater within-population variance (SNVs interquartile range (IQR): 61234 vs. 36715), the opposite is true for chrY variance (SNVs IQR: 14.8 vs. 580) (Figure 2A). The retracted chrY variance observed is consistent with our predominantly Bantu cohort (75.47%, 80/106 from E-V38 lineages), that would have descended from a limited pool of male migrants (and thus limited Y-lineages) into Southern Africa during the Bantu migration into the region approximately 1500 years ago.[28] While the A-M51 sublineages appear to be restricted to indigenous southern African San peoples, and B-M150 is a mix between Bantu and San peoples,[29,30] it is notable that all individuals self-identified ethnolinguistically as southern Bantu. Through autosomal interrogation we further confirm contributing San population fractions for all haplogroup A (mean 0.248, range 0.096–0.495) and haplogroup B (mean 0.153, range 0.0166–0.249) individuals, while 24.4% (21/86) of haplogroup E representative individuals lacked a San ancestral fraction (<0.1%) (Figure S1).

Four European patients shared the E-ancestral sub-branch E-M215 (and sublineage E-M35), with three patients with the E-M78 lineage (sublineage of E-L539) which predominate in Northern Africa, Eastern Africa, the Near East and Europe[31] (Figure 1; Table S1). In addition, these four patients shared a nonsynonymous *USP9Y* variant rs7067496 which was present in all African patients (Figure 2B; Table S2). This variant is the M235 marker for haplogroup F in ISOGG. All seven African

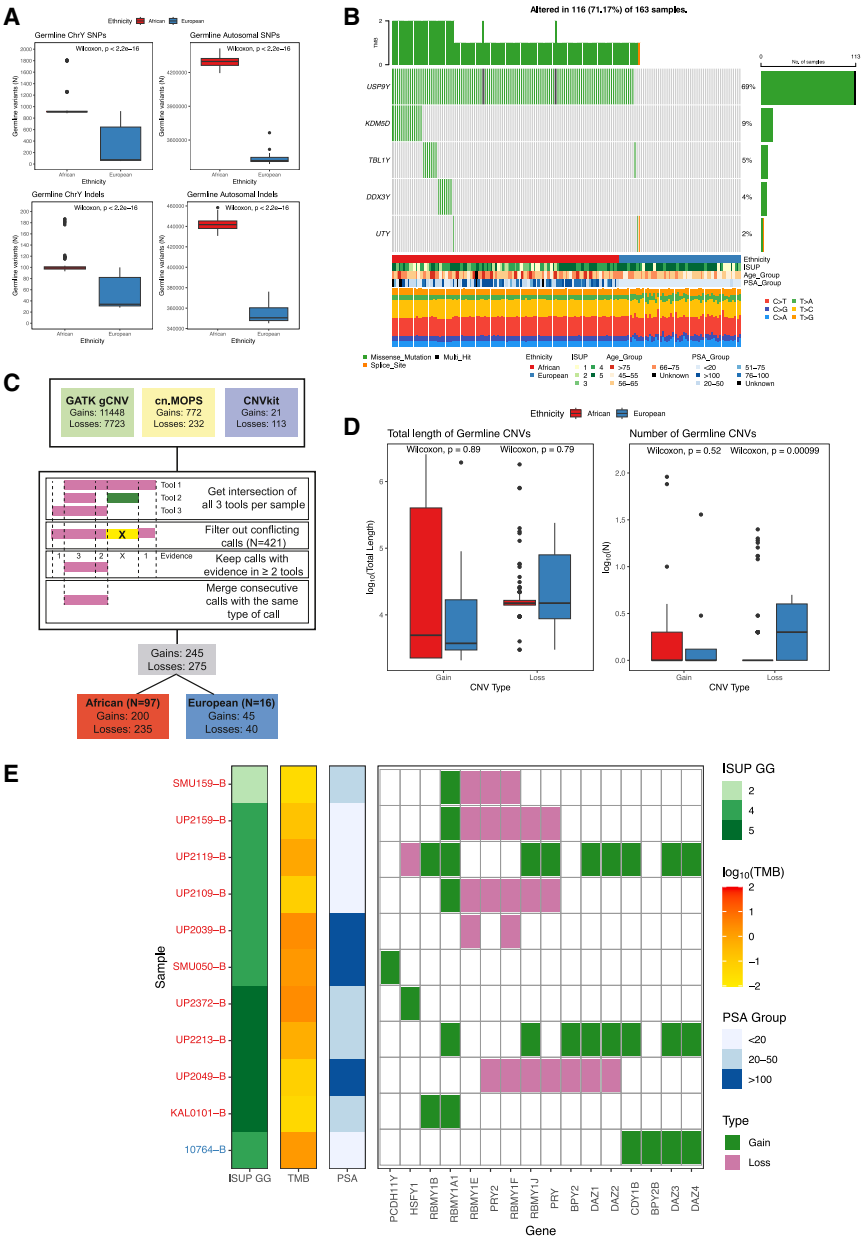


Figure 2. A higher number of germline single nucleotide variants (SNVs) were found in men of African ancestry compared to European ancestry, while among protein-coding genes, germline copy number variation (CNV) was limited in men of European ancestry

(A) Boxplots of the number of germline single nucleotide variants (SNVs) and indels in chrY compared to the autosome[26] by ethnicity. Data are represented by the median (horizontal line in box), the interquartile range (IQR) from 25th to 75th percentiles (box), with whiskers extending to the largest or smallest values no more than 1.5 times the IQR.

(B) Oncoplot displaying the number of non-synonymous germline variants found in the population.

(C) Workflow for the germline CNV analysis, showing the number of calls found by each tool and the method used to merge calls.

(D) The total length and the number of germline CNVs with evidence in at least two tools between ethnicities. Data are represented by the median (horizontal line in box), the IQR from 25th to 75th percentiles (box), with whiskers extending to the largest or smallest values no more than 1.5 times the IQR.

(E) CNV losses and gains identified in protein coding genes. Sample names are on the y-axis, color coded by ethnicity (red, African; blue, European). Clinical features for each sample are indicated by the first three columns, including the International Society of Urological Pathology grade group (ISUP GG), tumor mutational burden in the autosome[26] (TMB; values are in number of mutations per megabase), and prostate specific antigen group (PSA; values are in ng/mL).

patients in haplogroup A shared a nonsynonymous *DDX3Y* variant rs111406208 (not in ISOGG database), while all 13 African patients in haplogroup B shared a nonsynonymous *KDM5D* variant rs35681523 (marker M152 for haplogroup B-Y10267).

No associations between Y-haplogroups and PCa risk

Notable distinctions between the multi-ethnic cohorts include average age at diagnosis for African men being 5.2 years later (66.9 vs. 61.7 years old), while the European cohort was biased toward slightly more advanced disease at surgery defined here as International Society of Urological Pathology group grading (ISUP) 3, 4 or 5 or high-risk PCa (HRPCa, 87.7%, 50/57 vs. 83%, 88/106). As previously presented for the Southern African Prostate Cancer Study (SAPCS)[32,33] and in contrast to the National Cancer Consortium Network (NCCN) European-driven criteria that includes within the HRPCa definition as having a Prostate Specific Antigen (PSA) level >20 ng/mL, our African patients present with mean PSA levels 12.8-fold greater (255.94 ng/mL, range 4.28–4847 ng/mL). Unable to use European-derived PSA criteria to classify HRPCa in our study, we restrict our analyses to ISUP grading.

Testing for the association between major haplogroups of patients with HRPCa (N = 138) versus low-risk PCa (LRPCa; N = 25), we found no significant association between major haplogroups and aggressive PCa risk (p = 0.8, Fisher's exact test) (Table S3). Splitting into ethnicities, there was no significant association between haplogroups and PCa risk in European (LRPCa N = 7, HRPCa N = 50, p > 0.9, Fisher's exact test), nor African patients (LRPCa N = 18, HRPCa N = 88, p = 0.3, Fisher's exact test). Using a 'treeWAS' approach,[34] we further tested for the association between 9448 biallelic genotypes and HRPCa versus LRPCa while correcting for the population structure

within the phylogenetic tree, and found no significant associations even at a relaxed *p*-value threshold of 0.1.

Potentially deleterious germline chrY variants

A total of 13 nonsynonymous germline SNVs were detected through ANNOVAR, including 6 in *USP9Y*, 2 each in *KDM5D*, *TBL1Y* and *UTY*, and 1 in *DDX3Y* (Figure 2B; Table S2). None of these variants had an available ClinVar prediction. Notably, the three variants common (allele frequencies (AF) > 0.66) to our Southern Africans (rs373532788 *TBL1Y*, rs111406208 *DDX3Y*, rs35681523 *KDM5D*), are rare in the largely west African ancestral gnomAD populations (AF < 0.012), while the single variant common to our European cohort (rs7067496 *USP9Y*; AF = 0.07) is fixed in Southern Africans (AF = 1). Filtering for variants that were damaging/possibly damaging in at least one of SIFT, SIFT4G, PolyPhen2 HDIV or PolyPhen2 HVAR, which predicts the functional impact of variants through evolutionary conservation and changes to protein structure and are not dependent on other prediction tools,[35,36] followed by filtering out variants predicted to be benign by InterVar, eight potentially deleterious variants (PDVs) were left. Variants were considered unknown (38.5%, 5/13; 3 African and 2 European) if not found in the gnomAD v4.1 database, which has genomic variants from over 800,000 individuals across the globe.[37]

Five European patients presented with a rare PDV each and included *TBL1Y* (known chrY:7074584 A>G; present in both African/African American and European samples in gnomAD at AF < 6.1e-05), three *USP9Y* (unknown chrY:12842454 C>T, unknown chrY:12859400 A>C, and known rs766658730 C>G) and *UTY* (known rs200431840 G>A). These five European patients had ISUP grades of 4 or 5, with PSA levels ranging from 8.6 to 17 ng/mL, and were diagnosed between 58 and 72 years old. Two unknown African-specific rare PDVs identified include *UTY* (chrY:13355115 T>C) and *KDM5D* (chrY:19744477 T>A). The African patient with the *UTY* PDV was diagnosed at 62 years old, 4.9 years earlier than the mean African age of diagnosis, with a PSA of 1232.8 ng/mL and ISUP grade of 5. The African patient with the *KDM5D* PDV was diagnosed at 75 years, with a PSA of 51 ng/mL and ISUP grade of 4. Notably, the common Southern African *TBL1Y* (rs373532788, AF = 0.066) variant showed deleterious predictions from PolyPhen2 HDIV, PolyPhen2 HVAR, MutationTaster, fathmm-MKL coding, and with high CADD (22.2) and DANN (0.997) scores. While this study is not designed for common variant interrogation, the potential for this candidate PDV to contribute to PCa risk in this population warrants further investigation.

Germline chrY CNVs are common to African patients

We utilized three tools for germline CNV analysis, namely GATK gCNV,[38] cn.MOPS[39] and CNVkit[40] (Figure 2C). Requiring two-caller concurrence, we identified 200 gains and 235 losses in 97 (91.5%) African and 45 gains and 50 losses in 16 (28%) European patients (Figure S2). While there were no significant differences between the total length of CNV gains or losses between ethnicities, men of African ancestry had significantly fewer number of losses (Figure 2D, $p = 0.00099$, Wilcoxon rank-sum test).

CNVs impacted 15 and 4 protein-coding and 42 and 14 RNA genes in African and European samples, respectively (Figures S3 and S4). Notably, 83% (88/106) of African men presented with CN-loss spanning the lncRNA *TTTY22,* which included all patients presenting with haplogroups A ($n = 7$) and E-V38 ($n = 80$), and a single E-M75 patient, while 74.5% (79/106) showed loss of lncRNA *ENSG00000228379*, including 6/7 haplogroup A, 72/80 E-V38 and one E-M75 (Figure S3). While a single European patient showed CN-loss in both RNA genes, another presented with *TTTY22* CN-gain. Excluding these likely ancestral events, CNVs were present in 40 genes in 13.2% (14/106) of African and 12 genes in 14% (8/57) of European patients.

Notably, no losses were detected among protein-coding genes in Europeans, however, a single HRPCa patient showed CN-gain in 4 genes: *BPY2B*, *CDY1B*, *DAZ3* and *DAZ4* (Figures 2E and S4). Among African patients, CNVs in protein-coding genes were present in only one LRPCa patient (1/27; 3.7%), and nine HRPCa patients (9/79; 11.4%). All seven chrY's from haplogroup A had at least one germline CNV, with six carrying a gain in *RBMY1A1*, and four loss in *RBMY1E* and *RBMY1F*. Three haplogroup E African derived chrY's showed germline CNVs including gain of *HSFY11*, gain of *PCDH11Y*, and CN-loss spanning *PRY2*, *RBMY1F*, *RBMY1J*, *PRY*, *BPY2*, *DAZ1*, and *DAZ2*. The most frequent losses in African patients were in *RBMY1F* (5 patients), *PRY2* (4 patients) and *RBMY1E* (4 patients), while the most common gain was in *RBMY1A1* (6 patients).

Somatic copy number alterations disproportionately impact African tumors

Unlike germline small variants, the number of chrY somatic small variants were not significantly different between ethnicities, contrasting against the autosome, where there was a significantly greater number of somatic SNVs in Africans than Europeans ($p = 0.0024$) (Figure 3A). Nonsynonymous mutations were found in the genes *AMELY*, *DDX3Y*, *RPS4Y2*, *TBL1Y* in African tumors (5 samples: 4 HRPCa 1 LRPCa), and in *UTY* in one European HRPCa tumor (Figure 3B; Table S4). The tumor sample from the African patient UP2330, which had extremely high tumor mutational burden (TMB) of 174.85 mutations/Mb in the autosome, had a nonstop variant in *AMELY* as well as a splice site variant in *DDX3Y*.

Using the intersection and consensus between GATK gCNV[38] and CNVkit[40] for somatic copy number alterations (SCNAs), we identified 51 and 2 gains and 431 and 110 losses in 39 African and 21 European derived tumors, respectively, and representing 36.8% of each cohort (Figures 3C and S5). There were no significant differences ($p > 0.05$) between the total length and number of SCNAs between ethnicities (Figure 3D). SCNAs were present in 45 and 17 protein-coding and 121 and 41 RNA genes in African and European tumors, respectively (Figures S6 and S7). The most frequent losses in African tumors impacted the RNA genes *ENSG00000236951*, *ENSG00000291034*, *RBMY2FP*, *RNU6-1314P* and *TTTY5* (10/106 samples each; 9.4%), while the most common loss in European tumors was also in *ENSG00000291034* (6/57 samples; 10.5%). CN-gain was only observed in three RNA genes in European tumors (one sample each), compared to gains in 78 RNA genes in African tumors (mean 2.46 samples per gene, range 1–3) (Figure S6).

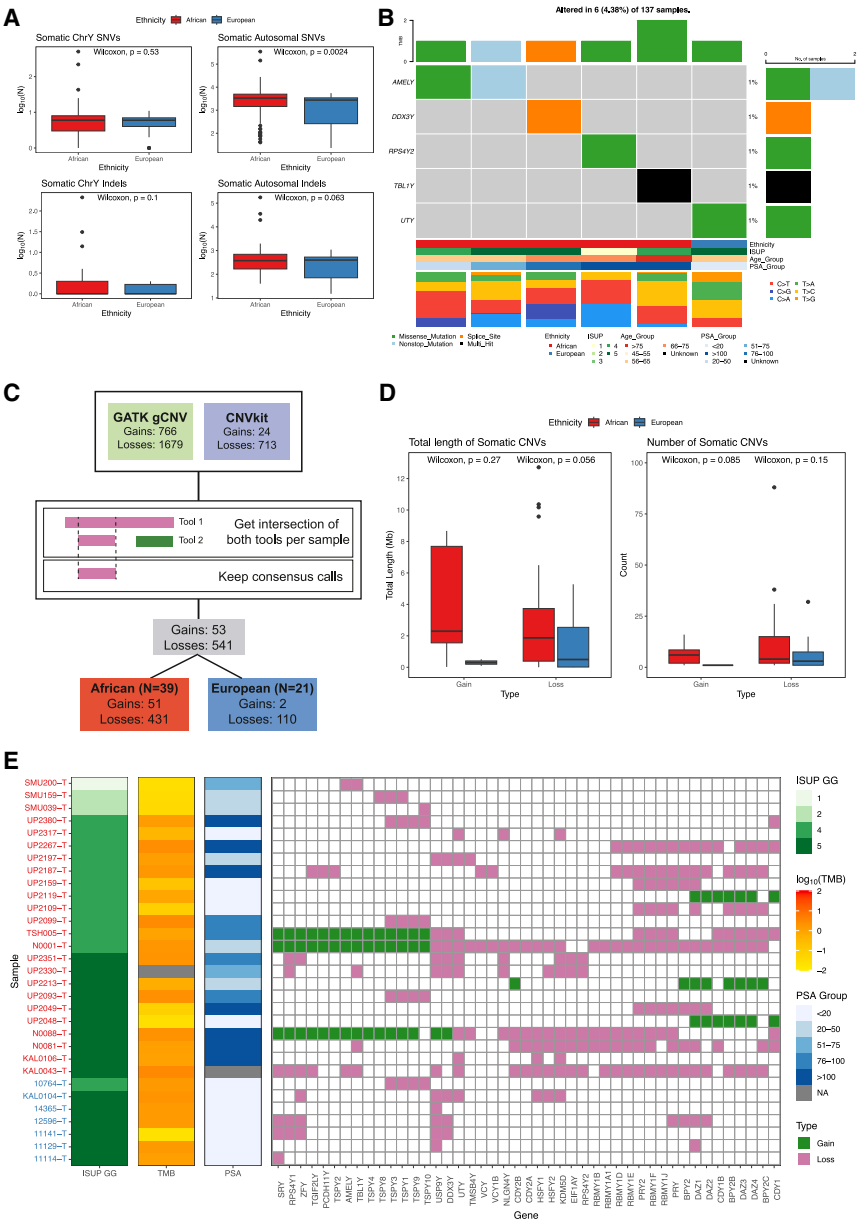


Figure 3. Somatic nonsynonymous variants were exceedingly rare and copy number gains were absent in European tumors, however copy number losses were common to both African (19.8%) and European (12.3%) tumors

(A) Boxplots of the number of somatic single nucleotide polymorphisms (SNVs) and indels in chrY compared to the autosome[26] by ethnicity. Data are represented by the median (horizontal line in box), the interquartile range (IQR) from 25th to 75th percentiles (box), with whiskers extending to the largest or smallest values no more than 1.5 times the IQR.

(B) Oncoplot displaying the number of non-synonymous somatic variants found in the population.

(C) Workflow for the somatic CNV analysis, showing the number of calls found by each tool and the method used to merge calls.

(D) The total length and number of somatic CNVs found with evidence in both tools between ethnicities. Data are represented by the median (horizontal line in box), the IQR from 25th to 75th percentiles (box), with whiskers extending to the largest or smallest values no more than 1.5 times the IQR.

(E) Somatic CNV losses and gains identified in protein coding genes. Sample names are on the y-axis, color coded by ethnicity (red, African; blue, European). Clinical features for each sample are indicated by the first three columns, including the International Society of Urological Pathology grade group (ISUP GG), tumor mutational burden in the autosome[26] (TMB; values are the $\log_{10}$ number of mutations per megabase), and prostate specific antigen group (PSA; values are in ng/mL).

Among protein-coding genes, while no gains were detected in European tumors, there was an elevated frequency of SCNAs for African tumors, though not significant (22.6% vs. 12.3%, *p*-value = 0.143, Fisher's test). *USP9Y* loss was the most common SCNA (5 patients, 8.7%). For African tumors, gains were detected in 3 samples for 18 genes each, whereas the most common loss impacted *PRY*, *PRY2*, and *RBMY1F* (10 patients for all, 9.4%) (Figures 3E and S7). LRPCa tumors had few SCNAs, representing 11% (3/27) of African tumors only, with losses in 1–3 protein-coding genes. For HRPCa tumors, SCNAs were more frequent in African (21/106; 19.8%) over European tumors (7/57, 12.3%), although not significant ($p = 0.2792$, Fisher's exact test). However, African tumors displayed SCNAs in a higher number of protein-coding genes (mean = 12.9, SD = 11.3, range 3–42), compared to European tumors (mean = 4.14, SD = 3.08, range 1–9), more than 3-fold higher despite the nearly 2-fold difference in sample size.

Lack of significant loss of chrY (LOY)

As an estimate for Y ploidy, the mean depth in X-degenerate regions of chrY were divided by half the mean depth of whole-genome coverage as previously reported.[26] We identified a median Y ploidy of 0.842 and 0.841 in the blood of African and European men, respectively, with minimum values of 0.535 and 0.558 in these populations (Figure S8). Y ploidy was not significantly associated with age in African men ($R^2 = 0.01$, *p*-value = 0.237), while age was a significant effect in European men but also similarly had high variability ($R^2 = 0.07$, *p*-value = 0.04) (Figure S9). Conversely, median somatic Y ploidy was 0.831 and 0.847, with minimum 0.479 and 0.639 in African and European men, respectively (Figure S10). While there was no significant association with age for both ancestries (Figure S11), when calculating the difference between tumor and blood Y ploidy per sample we observed no significant LOY (Figure S12). However, a single African patient (UP2351) presenting at age 66 year with

ISUP grade 5 diagnosed PCa and PSA 92 ng/mL showed a greater than 4 times SD from the mean of ploidy loss in the tumor (0.36), suggesting at least partial somatic LOY.

DISCUSSION

In this study, we sought to explore the landscape of germline and somatic Y chromosome variance associated with PCa ancestral disparities within an African-inclusive cohort.

Y-haplogroups

While some Y-haplogroups have been associated with increased PCa risk,[16–20] others have not.[13–15] Limited by our case-only design and sample size for risk association analyses, we found no correlation with high-risk disease. However, the identification of unknown African-specific chrY sub-branches will enable future studies focused on inclusion across the African diaspora, including modern human's earliest diverged paternal lineages (haplogroups A and B). Additionally, we provide evidence for the possible incorrect assignment of chrY F-M235 as haplogroup defining, as it was present in all African patients (haplogroups A, B, and E) and four European E-M35 patients.

The haplogroup-specific variation at *TTTY22* lncRNA, commonly experiencing CN-loss in E-P177 (E1b), is rare in all other haplogroups.[41] Here we found *TTTY22* loss in the majority of our African samples, including all seven haplogroup A Y chromosomes from the A1b lineage (A-M51, A-M9782, A-Y20629) absent from the previous study,[41] and in a single sample presenting with the likely non-Bantu E-M75 (E2)[30] lineage. We also refine *TTTY22* loss in E-P177 as a marker for sub-lineage E-V38 (Bantu African) but not E-M215 (European). Overall, these results show that *TTTY22* loss is likely ancestral to modern humans, with gain of this lncRNA a more recent event in human evolutionary divergence.

Germline variants

Seeking to identify inherited chrY PCa predisposing variants in this "gene-poor chromosome", of the 13 nonsynonymous variants, only the likely benign *USP9Y* rs7067496 was shared between the ancestries. Its commonality to all African participants again suggests this variant to be ancestral. Notably, European men were 3.1-fold more likely than African men to present with a rare PDV (prevalence 8.7% vs. 2.8%) and more likely to present with a PDV impacting *USP9Y* or *TBL1Y*, rather than *UTY* or *KDM5D*. This difference can be attributed to the "Out-of-Africa" bottleneck, where populations that have migrated out of Africa accumulate a higher number of deleterious mutations due to less time for purifying selection compared to African populations.[42] Of the eight PDVs identified, four are unknown (including two African) and three are found in *USP9Y* (prevalence of 5.3% of Europeans). While *USP9Y* overexpression has been shown to inhibit lung cancer tumorigenesis,[43] USP9Y fusion to TTTY15 appears to be a common event in prostate tumors of Chinese ancestral patients,[44] suggesting a possible mechanism for PDV functionality through altering USP9Y expression or TTTY15 fusion. Conversely, *TBL1Y* has not been reported to play a role in cancers but rather cardiac development and hearing loss, however, gene expression is primarily in the prostate and cochlea in adults.[45,46] *KDM5D*, a histone H3 lysine 4 demethylase, can modify gene expression resulting in aggressive PCa, interacts with the androgen receptor (AR) and can alter sensitivity to docetaxel.[47,48] *UTY* is another lysine demethylase that plays a role a prostate differentiation by mediating the interaction between NKX3.1 and G9a, with the authors suggesting a disruption to this network could potentially result in PCa predisposition.[49] Further work will be required to investigate whether these African PDVs alter gene function or epigenetic regulation.

While CNV is known to vary between populations,[50] individual specific CNV events were infrequent in our study. A single European HRPCa patient presenting with germline CN-gains impacting *CDY1B*, *DAZ3*, and *DAZ4*, also showed patient-specific gain in *BPY2B*. Notably, loss of *BPY2B* has been associated with male infertility.[51] In African patients, CNVs included a gain in *RBMY1A1* (5.7% of patients), and losses in *RBMY1F*, *RBMY1E*, *RBMY1J*, *PRY*, and *PRY2* (2.8–4.7%). *RBMY1* genes are known to be prone to CNV and structural organization, particularly CN-gains,[52] so it is notable that loss predominates for three of these family protein coding genes. While both the *RBMY1* and *PRY* genes have been suggestively linked to infertility,[53,54] and *RBMY1* has no known effects in cancer, intriguingly, *PRY* has been linked to apoptosis of spermatids and spermatozoa.[53]

Somatic variants

Here, we found nonsynonymous variants impacting *AMELY*, *DDX3Y*, *RPS4Y2*, and *TBL1Y* in five African tumors (4.7%), and *UTY* in a single European tumor (1.8%), highlighting the potential for these chrY genes as ancestry-specific cancer driver candidates. The higher frequency in the African tumors provides potential rationale for the inclusion of chrY in health disparity studies.

In contrast to somatic SNVs, prostate tumor SCNAs have been reported in a single study identifying a high frequency (14–52%) of gene losses impacting *SRY*, *ZFY*, *BPY1*, *KDM5D*, *RBM1*, and *BPY2*.[22] Additionally, for PCa cell lines losses in *NLGN4Y*, *TMSB4Y*, *TSPY1*, and *TTTY13* have been observed for LNCaP and PC3, with further *DDX3Y*, *TTTY15*, *USP9Y*, and *UTY* losses in PC3 only.[23] Studies have also noted differential expression of the chrY genes in PCa cell lines and regulation can be affected by androgen.[55,56] Here, SCNAs were notably absent from European LRPCa and rare in African low-risk tumors (up to three genes in three patients). Observing no distinct pattern of SCNAs among HRPCa, while gains were limited to African tumors, the most frequent protein-coding events were losses in *PRY*, *PRY2*, *RBMY1F*, *RBMY1J*, *UTY*, *DAZ1*, and *KDM5D* (>7.5% of African patients), with notable lack of CNs for *PRY2*, *RBMY1F*, and *RBMY1J* in European tumors. Overall, SCNAs were present in almost three times as many genes in African tumors (166 RNA and protein-coding genes) than European tumors (58 genes), partially owing to the larger African cohort, but may suggest differing transcriptional profiles between the ancestries.

Additionally, we found *DDX3Y* and *USP9Y* CN-loss to impact both African (5 and 6 tumors, respectively) and European tumors (3 and 5, respectively). However, one African tumor also showed a CN-gain of *DDX3Y* and *USP9Y*. *DDX3Y* along with its paralog on the X chromosome, *DDX3X*, are involved in RNA regulation with roles in neurogenesis, and can function as tumor suppressors or oncogenes that have been implicated in other cancers.[57] Recent

work has also found that DDX3Y is stabilized by USP9Y, and tumor suppressive effects were observed in lung cancer cells when both DDX3Y and USP9Y were overexpressed.[43]

Conversely, among many of the differences between both ancestries, notable differences were in *KDM5D*, *PCDH11Y* and *RBMY* genes. As highlighted above, *KDM5D* has been associated with aggressive PCa and resistance to docetaxel.[47,48] Here, we identified eight HRPCa African tumors and one HRPCa European tumor to carry a somatic loss in *KDM5D*. Furthermore, three African tumors had CN-gain in *PCDH11Y*, while one African tumor had CN-loss, and alterations in this gene were absent among European tumors. Expression of PCDH11Y has been noted to induce Wnt signaling and promote androgen-resistant malignant growth in the LNCaP cell line.[58] Lastly, *RBMY* genes losses were common in African tumors but absent from the European tumors. *RBMY* genes have shown a role in hepatocellular carcinoma and is potentially involved in AR activity regulation.[59,60] Collectively, the differences in SCNA between these genes in African versus European tumors point toward pathways for treatment-resistant tumors in African patients, suggesting that these genes may contribute to worsened mortality rates in African men.[21]

While LOY, and/or partial LOY, appears to be a frequent event in cancers,[24] yet rare in PCa, conversely, its presence has been associated with poor PCa survival.[24,25] Observing no significant ancestry-derived differences in inherited or somatic LOY, a single African outlier showed partial LOY in his aggressive presenting tumor. Besides seminal work by Qi et al. having explored LOY across many cancers using largely European-derived resources (>80% European in The Cancer Genome Atlas (TCGA) cohort),[24] future and larger African inclusive studies are required to determine if ancestral differences in complete or partial LOY contribute, at least in part, to the observed clinical disparities.

Conclusions

In conclusion, identifying as yet unknown chrY haplogroup substructure representing modern humans' earliest diverged paternal lineages, while we found no association between haplogroups and clinical features, European patients were 3.1-fold more likely to present with a rare PDV with ethnically driven gene specificity. In turn, we identified both commonalities (*DDX3Y* and *USP9Y*) and differences (*KDM5D*, *PCDH11Y*, and *RBMY* genes) in acquired chrY variation between African and European patients. Conversely, while European tumors lacked CN-gains, prevalence for SCNA in protein-coding genes was elevated for African tumors (22.6% vs. 12.3%). Identifying here disparities in the patients' inherited profile and acquisition of tumor-derived variance in men of African and European ancestries calls for further interrogation and inclusion of the understudied chrY in genomic studies aimed at developing an ancestrally inclusive model for PCa precision medicine.

Limitations of the study

While African inclusive, yet conscious to provide a technically, analytically, and as close to possible clinicopathologically matched non-African data source, there are several limitations that need to be highlighted. Compared with non-African focused efforts, this study is small making it hard to provide definitive conclusions. While biased toward aggressive disease, the late presentation of African patients makes it difficult to prioritize early-onset disease for the identification of PDVs, while lack of PCa awareness in the study region[61] would further limit prioritization for family history, a known PCa risk factor. Furthermore, lack of publicly available known African-relevant pathogenic chrY variants current databases, as well as lack of population and regionally matched genomic data, further limits our ability to adequately predict functionality and association. Additionally, due to the nature of merging CNVs across results from different tools, we were also unable to report specific numbers of gains and losses due to variation between callers, however, we attempted to capture the presence of a CNV that was in concordance among the callers.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Professor Vanessa M. Hayes (vanessa.hayes@sydney.edu.au).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data: DNA sequence data have been deposited at the European Genome-Phenome Archive (EGA), and the accession number is listed in the key resources table. They are available upon request if access is granted. This paper uses existing, publicly available data from gnomAD v3.1.2. The link to the dataset is listed in the key resources table.
- Code: This paper does not report original code.
- All other items: Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We acknowledge, most importantly, the patients and the clinical staff who contributed to the Southern African Prostate Cancer Study (SAPCS) and the St Vincent's Hospital Sydney resources in South Africa and Australia, respectively, as well as all the additional authors who contributed to generating or interrogating the published whole genome data resource. We acknowledge the Sydney Informatic Hub at the University of Sydney for providing critical computational infrastructure. Genomic sequencing was supported by the National Health and Medical Research Council (NHMRC) of Australia through a Project Grant (APP1165762 to V.M.H.) and Ideas Grants (APP2001098 to V.M.H. and M.S.R.B.; APP2010551 to V.M.H.). Further analytics was supported by a U.S.A. Congressionally Directed Medical Research Programs (CDMRP) Prostate Cancer Research Program (PCRP) HEROIC Consortium Award (PC210168 and PC230673, HEROIC PCaPH Africa1K to V.M.H. and M.S.R.B., which includes co-leads Professors Gail Prins, University of Illinois at Chicago, U.S.A., and Mungai Peter Ngugi, University of Nairobi, Kenya), a U.S.A. National Institute of Health (NIH) National Cancer Institute (NCI) Award (1R01CA285772-01 to V.M.H.) and a U.S.A. Prostate Cancer Foundation (PCF) Challenge Award (2023CHAL4150 to V.M.H.). V.M.H. was further supported by the Petre Foundation via the University of Sydney Foundation, Australia.

AUTHOR CONTRIBUTIONS

P.X.Y.S. and V.M.H. conceived and designed the study. P.X.Y.S., A.A., J.J., and V.M.H. provided methodological support and/or performed analyses, while P.X.Y.S. led the formal investigations. M.S.R.B., P.D.S., S.B.A.M., W. J., and V.M.H. contributed key resources. P.X.Y.S., A.A., and V.M.H., wrote

and edited the manuscript, P.X.Y.S. generated the figures, while V.M.H. provided supervision, project administration, and funding. All authors reviewed and approved the final manuscript.

DECLARATION OF INTERESTS

V.M.H. is a Member of Active Surveillance Movember Committee and received an honorarium from The Korean Urological Oncology Society for 2024 Annual Conference as a guest speaker.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Patient cohort and genomic data
  - Y-chromosome haplogroup and phylogenetic analyses
  - Autosomal substructure
  - Annotations and SNV predicted effects
  - Copy number variation (CNV) analysis
  - Loss of chrY (LOY) analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.112437.

Received: September 12, 2024
Revised: December 5, 2024
Accepted: April 10, 2025
Published: April 15, 2025

REFERENCES

1. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature *423*, 825–837. https://doi.org/10.1038/nature01722.
2. Rhie, A., Nurk, S., Cechova, M., Hoyt, S.J., Taylor, D.J., Altemose, N., Hook, P.W., Koren, S., Rautiainen, M., Alexandrov, I.A., et al. (2023). The complete sequence of a human Y chromosome. Nature *621*, 344–354. https://doi.org/10.1038/s41586-023-06457-y.
3. Hallast, P., Ebert, P., Loftus, M., Yilmaz, F., Audano, P.A., Logsdon, G.A., Bonder, M.J., Zhou, W., Höps, W., Kim, K., et al. (2023). Assembly of 43 human Y chromosomes reveals extensive complexity and variation. Nature *621*, 355–364. https://doi.org/10.1038/s41586-023-06425-6.
4. Wilson, M.A. (2021). The Y chromosome and its impact on health and disease. Hum. Mol. Genet. *30*, R296–R300. https://doi.org/10.1093/hmg/ddab215.
5. Helena Mangs, A., and Morris, B.J. (2007). The Human Pseudoautosomal Region (PAR): Origin, Function and Future. Curr. Genomics *8*, 129–136. https://doi.org/10.2174/138920207780368141.
6. Massaia, A., and Xue, Y. (2017). Human Y chromosome copy number variation in the next generation sequencing era and beyond. Hum. Genet. *136*, 591–603. https://doi.org/10.1007/s00439-017-1788-5.
7. Poznik, G.D., Xue, Y., Mendez, F.L., Willems, T.F., Massaia, A., Wilson Sayres, M.A., Ayub, Q., McCarthy, S.A., Narechania, A., Kashin, S., et al. (2016). Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nat. Genet. *48*, 593–599. https://doi.org/10.1038/ng.3559.
8. Hammer, M.F., Spurdle, A.B., Karafet, T., Bonner, M.R., Wood, E.T., Novelletto, A., Malaspina, P., Mitchell, R.J., Horai, S., Jenkins, T., and Zegura, S.L. (1997). The geographic distribution of human Y chromosome variation. Genetics *145*, 787–805. https://doi.org/10.1093/genetics/145.3.787.
9. Hjelmborg, J.B., Scheike, T., Holst, K., Skytthe, A., Penney, K.L., Graff, R.E., Pukkala, E., Christensen, K., Adami, H.O., Holm, N.V., et al. (2014). The heritability of prostate cancer in the Nordic Twin Study of Cancer. Cancer Epidemiol. Biomarkers Prev. *23*, 2303–2310. https://doi.org/10.1158/1055-9965.EPI-13-0568.
10. Nair-Shalliker, V., Bang, A., Egger, S., Yu, X.Q., Chiam, K., Steinberg, J., Patel, M.I., Banks, E., O'Connell, D.L., Armstrong, B.K., and Smith, D.P. (2022). Family history, obesity, urological factors and diabetic medications and their associations with risk of prostate cancer diagnosis in a large prospective study. Br. J. Cancer *127*, 735–746. https://doi.org/10.1038/s41416-022-01827-1.
11. Zeng, H., Xu, M., Xie, Y., Nawrocki, S., Morze, J., Ran, X., Shan, T., Xia, C., Wang, Y., Lu, L., et al. (2023). Racial/ethnic disparities in the cause of death among patients with prostate cancer in the United States from 1995 to 2019: a population-based retrospective cohort study. EClinicalMedicine *62*, 102138. https://doi.org/10.1016/j.eclinm.2023.102138.
12. Wang, A., Shen, J., Rodriguez, A.A., Saunders, E.J., Chen, F., Janivara, R., Darst, B.F., Sheng, X., Xu, Y., Chou, A.J., et al. (2023). Characterizing prostate cancer risk through multi-ancestry genome-wide discovery of 187 novel risk variants. Nat. Genet. *55*, 2065–2074. https://doi.org/10.1038/s41588-023-01534-4.
13. Wang, Z., Parikh, H., Jia, J., Myers, T., Yeager, M., Jacobs, K.B., Hutchinson, A., Burdett, L., Ghosh, A., Thun, M.J., et al. (2012). Y chromosome haplogroups and prostate cancer in populations of European and Ashkenazi Jewish ancestry. Hum. Genet. *131*, 1173–1185. https://doi.org/10.1007/s00439-012-1139-5.
14. Paracchini, S., Pearce, C.L., Kolonel, L.N., Altshuler, D., Henderson, B.E., and Tyler-Smith, C. (2003). A Y chromosomal influence on prostate cancer risk: the multi-ethnic cohort study. J. Med. Genet. *40*, 815–819. https://doi.org/10.1136/jmg.40.11.815.
15. Kim, W., Yoo, T.K., Kim, S.J., Shin, D.J., Tyler-Smith, C., Jin, H.J., Kwak, K.D., Kim, E.T., and Bae, Y.S. (2007). Lack of association between Y-chromosomal haplogroups and prostate cancer in the Korean population. PLoS One *2*, e172. https://doi.org/10.1371/journal.pone.0000172.
16. Ewis, A.A., Lee, J., Naroda, T., Sano, T., Kagawa, S., Iwamoto, T., Shinka, T., Shinohara, Y., Ishikawa, M., Baba, Y., and Nakahori, Y. (2006). Prostate cancer incidence varies among males from different Y-chromosome lineages. Prostate Cancer Prostatic Dis. *9*, 303–309. https://doi.org/10.1038/sj.pcan.4500876.
17. Lindstrom, S., Adami, H.O., Adolfsson, J., and Wiklund, F. (2008). Y chromosome haplotypes and prostate cancer in Sweden. Clin. Cancer Res. *14*, 6712–6716. https://doi.org/10.1158/1078-0432.CCR-08-0658.
18. Nargesi, M.M., Ismail, P., Razack, A.H.A., Pasalar, P., Nazemi, A., Oshkoor, S.A., and Amini, P. (2011). Linkage between Prostate Cancer Occurrence and Y-Chromosomal DYS Loci in Malaysian Subjects. Asian Pac. J. Cancer Prev. *12*, 1265–1268.
19. Carvalho, R., Pinheiro, M.F., and Medeiros, R. (2010). Localization of candidate genes in a region of high frequency of microvariant alleles for prostate cancer susceptibility: the chromosome region Yp11.2 genetic variation. DNA Cell Biol. *29*, 3–7. https://doi.org/10.1089/dna.2009.0905.
20. Cannon-Albright, L.A., Farnham, J.M., Bailey, M., Albright, F.S., Teerlink, C.C., Agarwal, N., Stephenson, R.A., and Thomas, A. (2014). Identification of specific Y chromosomes associated with increased prostate cancer risk. Prostate *74*, 991–998. https://doi.org/10.1002/pros.22821.
21. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., and Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. *74*, 229–263. https://doi.org/10.3322/caac.21834.

22. Perinchery, G., Sasaki, M., Angan, A., Kumar, V., Carroll, P., and Dahiya, R. (2000). Deletion of Y-chromosome specific genes in human prostate cancer. J. Urol. *163*, 1339–1342.

23. Seim, I., Jeffery, P.L., Thomas, P.B., Nelson, C.C., and Chopin, L.K. (2017). Whole-Genome Sequence of the Metastatic PC3 and LNCaP Human Prostate Cancer Cell Lines. G3 (Bethesda) *7*, 1731–1741. https://doi.org/10.1534/g3.117.039909.

24. Qi, M., Pang, J., Mitsiades, I., Lane, A.A., and Rheinbay, E. (2023). Loss of chromosome Y in primary tumors. Cell. https://doi.org/10.1016/j.cell.2023.06.006.

25. Stahl, P.R., Kilgué, A., Tennstedt, P., Minner, S., Krohn, A., Simon, R., Krause, G.V., Izbicki, J., Graefen, M., Sauter, G., et al. (2012). Y chromosome losses are exceedingly rare in prostate cancer and unrelated to patient age. Prostate *72*, 898–903. https://doi.org/10.1002/pros.21492.

26. Jaratlerdsiri, W., Jiang, J., Gong, T., Patrick, S.M., Willet, C., Chew, T., Lyons, R.J., Haynes, A.M., Pasqualim, G., Louw, M., et al. (2022). African-specific molecular taxonomy of prostate cancer. Nature *609*, 552–559. https://doi.org/10.1038/s41586-022-05154-6.

27. Chan, E.K.F., Timmermann, A., Baldi, B.F., Moore, A.E., Lyons, R.J., Lee, S.S., Kalsbeek, A.M.F., Petersen, D.C., Rautenbach, H., Förtsch, H.E.A., et al. (2019). Human origins in a southern African palaeo-wetland and first migrations. Nature *575*, 185–189. https://doi.org/10.1038/s41586-019-1714-1.

28. Schlebusch, C.M., and Jakobsson, M. (2018). Tales of Human Migration, Admixture, and Selection in Africa. Annu. Rev. Genomics Hum. Genet. *19*, 405–428. https://doi.org/10.1146/annurev-genom-083117-021759.

29. Naidoo, T., Xu, J., Vicente, M., Malmström, H., Soodyall, H., Jakobsson, M., and Schlebusch, C.M. (2020). Y-Chromosome Variation in Southern African Khoe-San Populations Based on Whole-Genome Sequences. Genome Biol. Evol. *12*, 1031–1039. https://doi.org/10.1093/gbe/evaa098.

30. Barbieri, C., Hübner, A., Macholdt, E., Ni, S., Lippold, S., Schröder, R., Mpoloka, S.W., Purps, J., Roewer, L., Stoneking, M., and Pakendorf, B. (2016). Refining the Y chromosome phylogeny with southern African sequences. Hum. Genet. *135*, 541–553. https://doi.org/10.1007/s00439-016-1651-0.

31. Cruciani, F., La Fratta, R., Santolamazza, P., Sellitto, D., Pascone, R., Moral, P., Watson, E., Guida, V., Colomb, E.B., Zaharova, B., et al. (2004). Phylogeographic analysis of haplogroup E3b (E-M215) y chromosomes reveals multiple migratory events within and out of Africa. Am. J. Hum. Genet. *74*, 1014–1022. https://doi.org/10.1086/386294.

32. Tindall, E.A., Monare, L.R., Petersen, D.C., van Zyl, S., Hardie, R.A., Segone, A.M., Venter, P.A., Bornman, M.S.R., and Hayes, V.M. (2014). Clinical presentation of prostate cancer in black South Africans. Prostate *74*, 880–891. https://doi.org/10.1002/pros.22806.

33. Gheybi, K., Jiang, J., Mutambirwa, S.B.A., Soh, P.X.Y., Kote-Jarai, Z., Jaratlerdsiri, W., Eeles, R.A., Bornman, M.S.R., and Hayes, V.M. (2023). Evaluating Germline Testing Panels in Southern African Males With Advanced Prostate Cancer. J. Natl. Compr. Canc. Netw. *21*, 289–296.e3. https://doi.org/10.6004/jnccn.2022.7097.

34. Collins, C., and Didelot, X. (2018). A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. PLoS Comput. Biol. *14*, e1005958. https://doi.org/10.1371/journal.pcbi.1005958.

35. Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P.C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. *40*, W452–W457. https://doi.org/10.1093/nar/gks539.

36. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249. https://doi.org/10.1038/nmeth0410-248.

37. Chen, S., Francioli, L.C., Goodrich, J.K., Collins, R.L., Kanai, M., Wang, Q., Alföldi, J., Watts, N.A., Vittal, C., Gauthier, L.D., et al. (2024). A genomic mutational constraint map using variation in 76,156 human genomes. Nature *625*, 92–100. https://doi.org/10.1038/s41586-023-06045-0.

38. Van der Auwera, G., and O'Connor, B.D. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra, 1st Edition (O'Reilly Media, Inc.).

39. Klambauer, G., Schwarzbauer, K., Mayr, A., Clevert, D.A., Mitterecker, A., Bodenhofer, U., and Hochreiter, S. (2012). cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. Nucleic Acids Res. *40*, e69. https://doi.org/10.1093/nar/gks003.

40. Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput. Biol. *12*, e1004873. https://doi.org/10.1371/journal.pcbi.1004873.

41. Shi, W., Massaia, A., Louzada, S., Banerjee, R., Hallast, P., Chen, Y., Bergström, A., Gu, Y., Leonard, S., Quail, M.A., et al. (2018). Copy number variation arising from gene conversion on the human Y chromosome. Hum. Genet. *137*, 73–83. https://doi.org/10.1007/s00439-017-1857-9.

42. Henn, B.M., Botigué, L.R., Peischl, S., Dupanloup, I., Lipatov, M., Maples, B.K., Martin, A.R., Musharoff, S., Cann, H., Snyder, M.P., et al. (2016). Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc. Natl. Acad. Sci. USA *113*, E440–E449. https://doi.org/10.1073/pnas.1510805112.

43. Xiu, L., Ma, B., and Ding, L. (2024). Antioncogenic roles of USP9Y and DDX3Y in lung cancer: USP9Y stabilizes DDX3Y by preventing its degradation through deubiquitination. Acta Histochem. *126*, 152132. https://doi.org/10.1016/j.acthis.2023.152132.

44. Zhu, Y., Ren, S., Jing, T., Cai, X., Liu, Y., Wang, F., Zhang, W., Shi, X., Chen, R., Shen, J., et al. (2015). Clinical utility of a novel urine-based gene fusion TTTY15-USP9Y in predicting prostate biopsy outcome. Urol. Oncol. *33*, 384.e9–384.e20. https://doi.org/10.1016/j.urolonc.2015.01.019.

45. Di Stazio, M., Collesi, C., Vozzi, D., Liu, W., Myers, M., Morgan, A., D Adamo, P.A., Girotto, G., Rubinato, E., Giacca, M., and Gasparini, P. (2019). TBL1Y: a new gene involved in syndromic hearing loss. Eur. J. Hum. Genet. *27*, 466–474. https://doi.org/10.1038/s41431-018-0282-4.

46. Meyfour, A., Ansari, H., Pahlavan, S., Mirshahvaladi, S., Rezaei-Tavirani, M., Gourabi, H., Baharvand, H., and Salekdeh, G.H. (2017). Y Chromosome Missing Protein, TBL1Y, May Play an Important Role in Cardiac Differentiation. J. Proteome Res. *16*, 4391–4402. https://doi.org/10.1021/acs.jproteome.7b00391.

47. Komura, K., Jeong, S.H., Hinohara, K., Qu, F., Wang, X., Hiraki, M., Azuma, H., Lee, G.S.M., Kantoff, P.W., and Sweeney, C.J. (2016). Resistance to docetaxel in prostate cancer is associated with androgen receptor activation and loss of KDM5D expression. Proc. Natl. Acad. Sci. USA *113*, 6259–6264. https://doi.org/10.1073/pnas.1600420113.

48. Komura, K., Yoshikawa, Y., Shimamura, T., Chakraborty, G., Gerke, T.A., Hinohara, K., Chadalavada, K., Jeong, S.H., Armenia, J., Du, S.Y., et al. (2018). ATR inhibition controls aggressive prostate tumors deficient in Y-linked histone demethylase KDM5D. J. Clin. Investig. *128*, 2979–2995. https://doi.org/10.1172/JCI96769.

49. Dutta, A., Le Magnen, C., Mitrofanova, A., Ouyang, X., Califano, A., and Abate-Shen, C. (2016). Identification of an NKX3.1-G9a-UTY transcriptional regulatory network that controls prostate differentiation. Science (New York, N.Y.) *352*, 1576–1580. https://doi.org/10.1126/science.aad9512.

50. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006). Global variation in copy number in the human genome. Nature *444*, 444–454. https://doi.org/10.1038/nature05329.

51. Chen, S., Zhang, Q., Chu, L., Chang, C., Chen, Y., Bao, Z., Peng, W., Zhang, L., Li, S., Liu, C., et al. (2022). Comprehensive copy number analysis of Y chromosome-linked loci for detection of structural variations and diagnosis of male infertility. J. Hum. Genet. *67*, 107–114. https://doi.org/10.1038/s10038-021-00973-3.

52. Shi, W., Louzada, S., Grigorova, M., Massaia, A., Arciero, E., Kibena, L., Ge, X.J., Chen, Y., Ayub, Q., Poolamets, O., et al. (2019). Evolutionary and functional analysis of RBMY1 gene copy number variation on the

human Y chromosome. Hum. Mol. Genet. *28*, 2785–2798. https://doi.org/10.1093/hmg/ddz101.

53. Stouffs, K., Lissens, W., Verheyen, G., Van Landuyt, L., Goossens, A., Tournaye, H., Van Steirteghem, A., and Liebaers, I. (2004). Expression pattern of the Y-linked PRY gene suggests a function in apoptosis but not in spermatogenesis. Mol. Hum. Reprod. *10*, 15–21. https://doi.org/10.1093/molehr/gah010.
54. Venables, J.P., Elliott, D.J., Makarova, O.V., Makarov, E.M., Cooke, H.J., and Eperon, I.C. (2000). RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing. Hum. Mol. Genet. *9*, 685–694. https://doi.org/10.1093/hmg/9.5.685.
55. Dasari, V.K., Goharderakhshan, R.Z., Perinchery, G., Li, L.-C., Tanaka, Y., Alonzo, J., and Dahiya, R. (2001). Expression Analysis of Y Chromosome Genes in Human. J. Urol. *165*, 1335–1341. https://doi.org/10.1016/s0022-5347(01)69895-1.
56. Lau, Y.-F.C., and Zhang, J. (2000). Expression analysis of thirty one Y chromosome genes in human prostate cancer. Mol. Carcinog. *27*, 308–321. https://doi.org/10.1002/(sici)1098-2744(200004)27:4<308::Aid-mc9>3.0.Co;2-r.
57. Gadek, M., Sherr, E.H., and Floor, S.N. (2023). The variant landscape and function of DDX3X in cancer and neurodevelopmental disorders. Trends Mol. Med. *29*, 726–739. https://doi.org/10.1016/j.molmed.2023.06.003.
58. Terry, S., Queires, L., Gil-Diez-de-Medina, S., Chen, M.W., de la Taille, A., Allory, Y., Tran, P.L., Abbou, C.C., Buttyan, R., and Vacherot, F. (2006). Protocadherin-PC promotes androgen-independent prostate cancer cell growth. Prostate *66*, 1100–1113. https://doi.org/10.1002/pros.20446.
59. Chua, H.H., Tsuei, D.J., Lee, P.H., Jeng, Y.M., Lu, J., Wu, J.F., Su, D.S., Chen, Y.H., Chien, C.S., Kao, P.C., et al. (2015). RBMY, a novel inhibitor of glycogen synthase kinase 3beta, increases tumor stemness and predicts poor prognosis of hepatocellular carcinoma. Hepatology *62*, 1480–1496. https://doi.org/10.1002/hep.27996.
60. Tsuei, D.J., Lee, P.H., Peng, H.Y., Lu, H.L., Su, D.S., Jeng, Y.M., Hsu, H.C., Hsu, S.H., Wu, J.F., Ni, Y.H., and Chang, M.H. (2011). Male germ cell-specific RNA binding protein RBMY: a new oncogene explaining male predominance in liver cancer. PLoS One *6*, e26948. https://doi.org/10.1371/journal.pone.0026948.
61. Hayes, V.M., Patrick, S.M., Shirinde, J., Jaratlerdsiri, W., Nenzhelele, M., Radzuma, M.B., Gheybi, K., Mokua, W., Oyaro, M.O., Moreira, D.M., et al. (2024). Health Equity Research Outcomes and Improvement Consortium Prostate Cancer Health Precision Africa1K: Closing the Health Equity Gap Through Rural Community Inclusion. J. Urol. Oncol. *22*, 144–149. https://doi.org/10.22465/juo.244800340017.
62. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.
63. Poznik, G.D., Henn, B.M., Yee, M.C., Sliwerska, E., Euskirchen, G.M., Lin, A.A., Snyder, M., Quintana-Murci, L., Kidd, J.M., Underhill, P.A., and Bustamante, C.D. (2013). Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. Science (New York, N.Y.) *341*, 562–565. https://doi.org/10.1126/science.1237619.
64. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.
65. Ralf, A., Montiel González, D., Zhong, K., and Kayser, M. (2018). Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data. Mol. Biol. Evol. *35*, 1291–1294. https://doi.org/10.1093/molbev/msy032.
66. Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., et al. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at bioRxiv. https://doi.org/10.1101/201178.
67. Edgar, R.C. (2022). Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. Nat. Commun. *13*, 6968. https://doi.org/10.1038/s41467-022-34630-w.
68. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics (Oxford, England) *35*, 4453–4455. https://doi.org/10.1093/bioinformatics/btz305.
69. Knaus, B.J., and Grünwald, N.J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. Mol. Ecol. Resour. *17*, 44–53. https://doi.org/10.1111/1755-0998.12549.
70. Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics (Oxford, England) *35*, 526–528. https://doi.org/10.1093/bioinformatics/bty633.
71. Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7. https://doi.org/10.1186/s13742-015-0047-8.
72. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109.
73. Behr, A.A., Liu, K.Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). pong: fast analysis and visualization of latent clusters in population genetic data. Bioinformatics (Oxford, England) *32*, 2817–2823. https://doi.org/10.1093/bioinformatics/btw327.
74. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. https://doi.org/10.1093/nar/gkq603.
75. Li, Q., and Wang, K. (2017). InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. Am. J. Hum. Genet. *100*, 267–280. https://doi.org/10.1016/j.ajhg.2017.01.004.
76. Patwardhan, M.N., Wenger, C.D., Davis, E.S., and Phanstiel, D.H. (2019). Bedtoolsr: An R package for genomic data analysis and manipulation. J. Open Source Softw. *4*, 1742. https://doi.org/10.21105/joss.01742.
77. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput. Biol. *9*, e1003118. https://doi.org/10.1371/journal.pcbi.1003118.
78. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York).
79. Soh, P.X.Y., Mmekwa, N., Petersen, D.C., Gheybi, K., van Zyl, S., Jiang, J., Patrick, S.M., Campbell, R., Jaratlerdseri, W., Mutambirwa, S.B.A., et al. (2023). Prostate cancer genetic risk and associated aggressive disease in men of African ancestry. Nat. Commun. *14*, 8037. https://doi.org/10.1038/s41467-023-43726-w.
80. Coutelier, M., Holtgrewe, M., Jäger, M., Flöttman, R., Mensah, M.A., Spielmann, M., Krawitz, P., Horn, D., Beule, D., and Mundlos, S. (2022). Combining callers improves the detection of copy number variants from whole-genome sequencing. Eur. J. Hum. Genet. *30*, 178–186. https://doi.org/10.1038/s41431-021-00983-x.
81. Gabrielaite, M., Torp, M.H., Rasmussen, M.S., Andreu-Sánchez, S., Vieira, F.G., Pedersen, C.B., Kinalis, S., Madsen, M.B., Kodama, M., Demircan, G.S., et al. (2021). A Comparison of Tools for Copy-Number Variation Detection in Germline Whole Exome and Whole Genome Sequencing Data. Cancers (Basel) *13*, 6283. https://doi.org/10.3390/cancers13246283.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
DNA sequence data	Jaratlerdsiri et al.[26]	EGA: EGAS00001006425
gnomAD v3.1.2 HGDP + 1KGP subset	Karczewski et al.[62] (gnomAD)	https://gnomad.broadinstitute.org/downloads#v3-hgdp-1kg
San reference population DNA sequence data	W.J., unpublished data	N/A
GRCh38 strict callability mask, version 20160622	1000 Genomes Project; Poznik et al.[63]	https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/working/20160622_genome_mask_GRCh38/
Reference hg38 chrY fasta	UCSC	https://hgdownload.cse.ucsc.edu/goldenpath/hg38/chromosomes/chrY.fa.gz
Canonical annotations GENCODE V46	UCSC	https://genome.ucsc.edu/cgi-bin/hgTables
Software and algorithms		
bcftools v1.17	Danecek et al.[64]	https://www.htslib.org/download/
samtools v1.6	Danecek et al.[64]	https://www.htslib.org/download/
Y-leaf v3.0	Ralf et al.[65]	https://github.com/genid/Yleaf
GATK v4.2.0.0 HaplotypeCaller	Poplin et al.[66]	https://github.com/broadinstitute/gatk
vcf-tab-to-fasta perl script	Jinfeng Chen	https://github.com/JinfengChen/vcf-tab-to-fasta
Muscle5 v5.1	Edgar[67]	https://github.com/rcedgar/muscle
RAxML-ng v1.0.3	Kozlov et al.[68]	https://github.com/amkozlov/raxml-ng
FigTree v1.4.4	Rambaut	https://github.com/rambaut/figtree/
Adobe Illustrator 2023	Adobe Inc.	https://www.adobe.com/au/products/illustrator.html
R package 'treeWAS'	Collins and Didelot[34]	https://github.com/caitiecollins/treeWAS/
R package 'vcfR'	Knaus and Grunwald[69]	https://github.com/knausb/vcfR
R package 'ape'	Paradis and Schliep[70]	https://cran.r-project.org/web/packages/ape/index.html
PLINK v2.0	Chang et al.[71]	https://www.cog-genomics.org/plink/2.0/
ADMIXTURE v1.3.0	Alexander et al.[72]	https://dalexander.github.io/admixture/
pong v1.5	Behr et al.[73]	https://dalexander.github.io/admixture/
ANNOVAR	Wang and Hakonarson[74]	https://annovar.openbioinformatics.org
InterVar (online database, version 13 June 2022)	Li and Wang[75]	https://wintervar.wglab.org
GATK v4.4.0.0 gCNV	Poplin et al.[66]	https://github.com/broadinstitute/gatk
R package 'cn.MOPS'	Klambauer et al.[39]	https://bioconductor.org/packages/cn.mops/
CNVkit v0.9.10	Talevich et al.[40]	https://github.com/etal/cnvkit
R package 'bedtoolsr'	Patwardhan et al.[76]	https://github.com/PhanstielLab/bedtoolsr
R package 'IRanges'	Lawrence et al.[77]	https://bioconductor.org/packages/IRanges
R package 'ggplot2'	Wickham[78]	https://ggplot2.tidyverse.org

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The primary publication from Jaratlerdsiri et al.[26] contains all patient information and genomic data used for this study, including a total 163 male patients of European (n = 57) and African ancestry (n = 106).

METHOD DETAILS

Unless otherwise stated, RStudio v4.3.1 was used to generate plots (using 'ggplot2' package[78]).

Patient cohort and genomic data

DNA was extracted from blood and prostate tumour tissue samples from 163 male patients with histopathologically diagnosed PCa (mean age 65.2 ± 8.12) of European (n = 57, including 53 from Australia and 4 from South Africa) and African ancestry (n = 106, from South Africa). Processing of 150 bp paired end Illumina NovaSeq data into analysis-ready bam files (aligned to the hg38 reference genome), as well as ancestry classification, was conducted previously.[26] Aggressive disease was defined using the International Society of Urological Pathology Grade Group (ISUP GG) into high- (ISUP GG $\geq$ 3; HRPCa) and low-risk (ISUP GG < 3; LRPCa) PCa. Prostate serum antigen (PSA) levels were described previously.[26]

Y-chromosome haplogroup and phylogenetic analyses

Y chromosome reads were extracted from germline analysis-ready bam files using samtools v1.6.[64] Y-haplogroups were predicted using Y-leaf v3.0 using bam files as input.[65] Germline variants were called using GATK's v4.2.0.0 program HaplotypeCaller[66] in haploid mode to produce vcf files. As a large portion of the Y chromosome is inaccessible, the GRCh38 strict callability mask (20160622 version) from 1000 Genomes Project (1KGP) was used to filter variants (https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/working/20160622_genome_mask_GRCh38/).[63]

Using SNVs present in these Y-chromosome callable regions, the vcf file was converted to fasta format using perl scripts from https://github.com/JinfengChen/vcf-tab-to-fasta (accessed 20 Sep 2023), then aligned using Muscle5 v5.1.[67] RAxML-ng v1.0.3 was then used to create a maximum likelihood phylogenetic tree, using the GTRGAMMA model with 200 bootstraps with the Felsenstein bootstrap and transfer bootstrap expectation options.[68] FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) was used for visualisation using midpoint rooting. Adobe Illustrator 2023 was used to further colour and annotate the figure.

The R package 'treeWAS'[34] was used to test for associations between PCa risk and biallelic variants while accounting for population structure in a phylogenetic tree. The vcf file stated above was filtered to keep 9448 biallelic variants, then the R package 'vcfR'[69] was used to read the file into RStudio and to extract genotypes. The genotypes were then converted to a matrix (base R function). The best tree file from RAxML-ng's output was read into R using the 'ape' package.[70] The treeWAS function was run with default settings with the best tree, the genotype matrix, and a vector of phenotypes for each sample (1 = HRPCa, 0 = LRPCa). As the default p-value threshold is 0.01, the function was also tested using relaxed p-value thresholds of 0.05 and 0.1.

Autosomal substructure

A set of 77,372 genome-wide exomic variants that we previously successfully used to discern ancestral differences within-Africa[79] were extracted from the autosomal data of our cohort[26] using bcftools v1.17,[64] along with reference populations of Asian (Han Chinese, CHB), European (Utah, USA residents with North and Western European ancestry, CEU), East African (Luhya Kenyan, LWK) and West African (Yoruba Nigerian, YRI) ancestry (N=20 each, randomly selected) from the gnomAD v3.1.2 HGDP and 1KGP subset.[62] An additional 20 individuals with known San ancestry were added to the reference population (W.J., unpublished data). The data was merged with bcftools v1.17[64] and converted to PLINK[71] format, subsequently SNVs that were not in dbSNP156 and those that were fixed were removed, leaving 64,654 autosomal SNVs. Unsupervised ADMIXTURE v1.3.0 analysis was performed on this dataset for K=2 to K=10 with five-fold cross-validation (CV), with ten replicates each,[72] and runs were evaluated for concordance using pong v1.5.[73] While K=3 provided the lowest CV error rate at a mean 0.253, here, the average fractions at K=4 was reported instead as it displays the split between South Africa and East/West Africa ancestries, with only slightly higher mean CV error rate of 0.255.

Annotations and SNV predicted effects

ANNOVAR was used to annotate variants for databases RefSeq gene from UCSC (refGene; 20211019 version), dbSNP (avsnp150; 20170929 version), REVEL (revel; 20161205 version), ClinVar (clinvar_2022123; 20230105 version), and dbNSFP (dbnsfp42a; 20210710 version).[74] Nonsynonymous variants were queried on InterVar (last update 13 June 2022) to determine pathogenicity.[75]

Copy number variation (CNV) analysis

As other studies have shown,[80,81] the best approach to calling CNVs in whole genome sequencing data is to combine several callers. However, many of these callers do not adjust for the haploid nature of chrY and even fewer are able to detect somatic copy number alterations (SCNAs) in chrY. As such, we selected tools that can handle chrY data, including two tools GATK v4.4.0.0 gCNV[38] and cn.MOPS v1.46.0[39] that were suggested to be among the top four callers by Gabrielaite and colleagues,[81] as well as CNVkit v0.9.10.[40]

GATK's gCNV was run with default parameters according to online tutorials provided by GATK.[38] For the germline analysis, read counts were first collected according to a default bin length of 1000 for each sample (PreprocessIntervals, CollectReadCounts), including a hg38 chrY fasta file downloaded from UCSC (https://hgdownload.cse.ucsc.edu/goldenpath/hg38/chromosomes/chrY.fa.gz). Then, the chrY callability mask was used to annotate and filter intervals (AnnotateIntervals, FilterIntervals). Contig ploidy was then determined (DetermineGermlineContigPloidy), followed by CNV calling in cohort mode (GermlineCNVCaller). Finally, copy number segments and sample results were consolidated with PostprocessGermlineCNVCalls.

For the somatic analysis, read counts were collected from the tumour bam files, then a panel of normal (PoN) was created from the germline read counts (CreateReadCountPanelOfNormals). Tumour read counts were then standardised and denoised against the PoN (DenoiseReadCounts). Next, the germline vcf of SNVs were converted to an interval list, and reference and alternate allele counts

were individually collected for tumour and blood bam files (CollectAllelicCounts). Contiguous copy ratios were then grouped into copy number segments (ModelSegments), specifying tumour and normal allelic counts. Finally, copy number netural, amplified or deleted segments were called with CallCopyRatioSegments and plotted with PlotModeledSegments.

For cn.MOPS, the callability mask was read into R v4.2.2 and converted into a GRanges object. All chrY blood bam files were then read into R using the function ''getSegmentReadCountsFromBam'', while specifying the callability mask GRanges object. Germline CNV calling was then conducted by using the ''haplocn.mops'' function with default parameters, which conducts poisson normalisation across all samples and does ''DNAcopy'' circular binary segmentation. Copy numbers were then calculated using the ''calcIntegerCopyNumbers'' function, then returning the CNVs and CNV regions (CNVR) using the ''cnvs'' and ''cnvr'' functions respectively. As cn.MOPS does not currently have a haploid pipeline for somatic CNVs, this was not conducted.

For CNVkit, germline CNV calling was conducted using the batch pipeline on blood chrY bam files, specifying the whole genome method (-m wgs), Hidden-Markov Model germline segmentation (–segment-method hmm-germline), a target average size of 1000 (to match GATK gCNV's default bin length), specifying the callability mask as an access file, the hg38 chrY fasta from UCSC, and known canonical annotations (GENCODE V46) downloaded from UCSC's Table Browser (including primary tables kgXref and knownAttrs). For somatic CNV calling, the same batch pipeline was used with the same parameters except tumour chrY bam files were used as input, with blood chrY bam files specified under ''–normal", and the tumour version of the Hidden-Markov Model segmentation was used (–segment-method hmm-tumour).

For each sample, we used the multi-intersect function from the 'bedtoolsr' package[76] to get the intersection between calls across the tools. Conflicting calls were then filtered out, and calls were kept if there was evidence of a CNV in at least two tools for the germline analysis, while for the somatic analysis, calls found in both CNVkit and GATK gCNV were kept. Lastly, consecutive calls with the same type of call (gain or loss) were merged using the reduce function from the 'IRanges' package.[77]

Loss of chrY (LOY) analysis

Using X-degenerate regions as defined previously,[3] the 'depth' function from samtools v1.6 was used to calculate depth of coverage at each base position. For each sample, an estimated Y chromosome ploidy was calculated by dividing the mean coverage across these X-degenerate regions by half the mean autosomal coverage as reported in our previous publication.[26]

QUANTIFICATION AND STATISTICAL ANALYSIS

Wilcoxon rank sum and Fisher's exact tests were conducted in R v4.3.1. Statistical details of these tests are provided in the figure or table legends, or in the text. A p-value < 0.05 was considered statistically significant.