

Rare pathogenic structural variants show potential to enhance prostate cancer germline testing for African men

Vanessa Hayes

vanessa.hayes@sydney.edu.au

University of Sydney <https://orcid.org/0000-0002-4524-7280>

Tingting Gong

Human Phenome Institute, Fudan University <https://orcid.org/0000-0001-5907-2445>

Jue Jiang

Garvan Institute of Medical Research <https://orcid.org/0000-0003-0920-8310>

Riana Bornman

University of Pretoria <https://orcid.org/0000-0003-3975-2333>

Kazzem Gheybi

University of Sydney

Phillip Stricker

St. Vincent's Hospital <https://orcid.org/0000-0002-0934-0656>

Joachim Weischenfeldt

BRIC, University of Copenhagen <https://orcid.org/0000-0002-3917-5524>

Shingai Mutambirwa

Sefako Makgatho Health Science University

Article

Keywords: prostate cancer, pathogenic variants, structural variants, health disparity, African ancestry, germline testing

Posted Date: June 13th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4531885/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: There is NO Competing Interest.

Abstract

Prostate cancer (PCa) is highly heritable, with men of African ancestry at greatest risk and associated lethality. Lack of representation in genomic data means germline testing guidelines exclude for African men. Established that structural variations (SVs) are major contributors to human disease and prostate tumourigenesis, their role is under-appreciated in familial and therapeutic testing. Utilising a clinico-methodologically matched African (n = 113) *versus* European (n = 57) deep-sequenced PCa resource, we interrogated 42,966 high-quality germline SVs using a best-fit pathogenicity prediction workflow. We identified 15 potentially pathogenic SVs representing 12.4% African and 7.0% European patients, of which 72% and 86% met germline testing standard-of-care recommendations, respectively. Notable African-specific loss-of-function gene candidates include DNA damage repair *MLH1* and *BARD1* and tumour suppressors *FOXP1*, *WASF1* and *RB1*. Representing only a fraction of the vast African diaspora, this study raises considerations with respect to the contribution of kilo-to-mega-base rare variants to PCa pathogenicity and African associated disparity.

Introduction

Prostate cancer (PCa) is a significant global health burden and a leading cause of male associated cancer deaths¹. With one of the highest heritability rates (estimated 58%), PCa risk shows a great degree of variability², particularly when considering a man's ancestral heritage. In the United States, Black men are at greatest risk for aggressive disease presentation³ and depending on age at diagnosis an over double to triple (< 65 years) the risk for PCa-associated mortality than White Americans^{4, 5}. Contributed by a complex interaction of socioeconomic factors and genetics⁶, inherited risk includes a combination of both common (low-risk with combined genetic risk scores) and rare (high-risk or pathogenic) germline variants^{7, 8}. Revolutionised through advancement of precision oncology, most notably the approval of the poly-(ADP ribose) polymerase (PARP) inhibitors Olaparib⁹ and rucaparib¹⁰ for the treatment of metastatic castrate resistant PCa for patients harbouring rare pathogenic variants in specified DNA repair genes¹¹, has increased the value for germline testing. Furthermore, the National Comprehensive Cancer Network (NCCN) recommends germline testing for all men with metastatic, recurrent or high-risk localized PCa, regardless of family history¹². Although a significant risk factor for aggressive disease, no consensus could be reached for men of African ancestry¹³, while a recent review further highlighted the knowledge gap¹⁴.

The lack of consensus for PCa germline testing in Black men is directly attributed to a lack of available data, compounded by a lack of African-relevant genomic data that captures the true extent of elevated genetic diversity. While consensus has yet to be reached for minority inclusion in the benefits of recent breakthroughs in PCa precision oncology, contradictory studies suggest that Black American patients harbour more¹⁵ and conversely less actionable pathogenic variants than White Americans¹⁶. The picture is no different for Africa, although more recently PCa genomics has reached the continent with the first whole exome (n = 45 Nigerian)¹⁷ and whole genome sequencing studies (n = 113 Black South Africans)¹⁸. Although preliminary, notable differences within Africa are emerging. For example, an elevated frequency of *BRCA1* germline mutations reported for Nigerian patients, reflecting African American data^{17, 19}, is lacking in Southern African cases²⁰. Additionally, we have recently reflected on the lack of the West African exclusive and functionally relevant common PCa susceptibility variants *CHEK2* p.Ile448Ser (rs17886163) and *HOXB13* p.Ter285Lys (rs77179853) in Southern Africa^{21, 22}. Reporting a 2.1-fold age-adjusted increase in aggressive PCa presentation in Black South African *versus* Black American men²³, through deep sequenced interrogation for the 20 most common genes included in PCa germline testing panels using NCCN inclusion criteria (Gleason score \geq 8), we observed a prevalence for rare pathogenic variants of 5.6%²⁰, comparable with a single East African study (5.7%)²⁴ and almost half that reported for non-Africans (11.8%)²⁵. These studies highlight the need for developing African-relevant PCa germline testing panels through African inclusion in genome profiling.

Again, it is well established that Structural Variations (SVs) play a critical role in prostate tumour progression with prognostic and therapeutic potential^{26, 27}, including tumours derived from men of African ancestry^{18, 28}. Yet, irrespective of patient ancestry, little is known with regards to the contribution of germline potentially pathogenic rare SVs. Typically, greater than 50 bases in length, SVs encompassing large deletions (DEL), duplications (DUP), insertions (INS), inversions (INV) and translocations (TRA), are overlooked and/or difficult to resolve using current germline genetic testing assays. While it is well established that SVs play a critical role in diagnostic screening for inherited genetic diseases²⁹, more recently, long-read sequencing has been used to identify potential pathogenic SVs in hereditary cancer syndromes³⁰ and known breast cancer susceptibility genes³¹, however, the impact of rare pathogenic SVs on PCa predisposition, and in turn targeted treatment, remains unknown.

Expanding on our earlier work^{18, 20, 28}, including deep sequenced germline genomes for 113 African (Black South African) and 57 European (4 South African, 53 Australian) PCa patients, through high-quality SV calling and genotyping, comprehensive gene annotation and best-fit pathogenicity prediction workflow, we interrogate for rare potentially pathogenic SVs (PP-SVs). While agreeably a small study size, this resource is not only unique for the African continent, importantly it provides clinically and technically matched non-African data for direct comparative analyses, while the whole genome approach increases sensitivity for SV detection. As such, the study aims to limit spurious findings between the ancestries, while providing a foundation for further efforts across the continent. Identifying candidate PP-SVs highlights the value of whole genome interrogation not only to improve the detection rate for rare pathogenic PCa variants, but importantly begin to contribute to the much-needed emphasis on an all-African inclusion model for germline testing and associated clinical care.

Results

NCCN high-risk characterisation for ancestrally assigned PCa patients

Clinically and technically matched whole genome sequenced germline data (mean coverage 45.9X; range 30.2-97.6X) was derived from 170 PCa patients, ancestrally classified previously using 7,472,833 genome-wide SNVs and population substructure analysis¹⁸. In brief, 113 Black South African patients presented with an African ancestral genetic fraction of > 85%, while the 57 White patients presented with European ancestral genetic fractions of > 90% (4 South African, 52 Australian) and 73.7% European and 26.3% Asian substructure (1 Australian) (**Supplementary Table 1**). Importantly, although mean age was 5-years younger at presentation or surgery, a greater number of European (86%; 49/57) over African patients (72%; 81/113) met current NCCN guidelines for germline testing based on International Society of Urological Pathology (ISUP) Group Grading defined as high-risk localized PCa (ISUP 4/5 or Gleason score ≥ 8). Notably, we have previously provided evidence for the extension of these criteria for Black South African men to include ISUP 3, which would expand our cohort of high-risk Black men to 82% (93/113)²⁰. While Black South Africans present with significantly elevated median and range of prostate specific antigen (PSA) levels (median 244 ng/mL *versus* 9.4), as previously presented^{18, 23}, still the study was biased towards over representation of NCCN guidelines for PSA inclusive high-risk PCa for the European (70.2%; 40/57) over African patients (65/113; 57.5%).

Genome-wide gene-disrupting SV discovery

In this study, we identified and genotyped 42,966 high-quality germline SVs. We found a median of 9,206 SVs (range: 8,891 to 9,708) per-African genome, which is significantly higher than the median of 7,490 (range: 7,309 to 8,050) per-European genome (p-value = 1.1×10^{-26} by Wilcoxon test). In total, we identified 38,668 African derived SVs (18,674 private) and 24,292 European derived SVs (4,298 private) (**Supplementary Table 2**). Including only high-quality genotype calls for allele frequency (AF) estimation left a total of 33,243 high-confidence SVs. Excluding for common SVs, defined as minor allele frequency (MAF) > 5%, a total of 20,982 rare (MAF < 1%) and low-frequency (MAF = 1 to 5%) SVs remained across the ancestries for further annotation (Fig. 1).

Further interrogation for gene regions overlapping, we identified 1,857 gene-disruptive SVs, including 1,752 potential Loss-of-Function (pLoF), 52 Copy Gain (CG) and 53 Intragenic Exon DUP (IED) (detailed in **Methods**). Notably, pLoF, CG and IED SVs can have functional impact on genes through either gene inactivation or increased dosage effect³². Conversely, there is no clear or direct coding effect by SVs with other gene impact types, which included in our study 109 partial gene DUP, 22 partial exon DUP, 48 whole-gene INV, 343 promoter SVs, 9,431 intronic SVs and 258 enhancer SVs. As such, the latter SVs were not discussed further. In total, we identified 1,857 (MAF $\leq 5\%$) gene-disruptive SVs of which 1,407 are African-relevant, including 93% (1,314) African-private, and 543 European-relevant, including 83% (450) European-private (**Supplementary Table 2**). There were 93 SVs (5%) shared by both African and European PCa patients. The 1,857 gene-disruptive SVs (1,050 rare in both African and European) underwent further downstream interrogation for potential clinical relevance. Of the 1,857 gene-disruptive SVs, 1,167 were previously reported in dbVar database of SVs, while 690 were absent and as such regarded as novel, of which 513 (74%) are uniquely African (Fig. 1).

Characterising ClinVar verified candidate potentially pathogenic SVs

Of the 1,167 dbVar reported gene-disruptive SVs, 14 (1.2%) were recorded in ClinVar, with three reported as 'pathogenic' or 'likely pathogenic' based on functional prediction consensus. One 2,958 bp likely pathogenic DEL results in loss of exon 7 in *OCA2* (**Supplementary Fig. 1**), a 5,064 bp pathogenic DEL leads to exon 5–7 loss in *PIGN* (**Supplementary Fig. 2**), while a 235 bp likely pathogenic DUP duplicates exon 3 of *SLC3A1* (**Supplementary Fig. 3**). The *OCA2* and *PIGN* DELs were identified in a single African patient each, while the *SLC3A1* DUP presented in two African patients (Table 1).

Although pathogenic in ClinVar, none have been associated with cancer phenotypes and include rather oculocutaneous albinism, multiple congenital anomalies-hypotonia-seizures syndrome and cystinuria, respectively. As such, we searched the literature for plausibility with further ascertainment derived from normal prostate and tumour tissue data sets using GENT2³³. Reported to be downregulated in numerous cancer types (all-type $P < 0.001$, GENT2 T-test), although not significant for PCa, pLoF deletion of the pigmentation gene *OCA2* has been linked not only to Prader-Willi syndrome, but also Prader-Willi associated malignancies³⁴, and melanoma³⁵, with recent studies linking melanoma with increased PCa risk³⁶. Highly expressed in normal prostate tissue with significant upregulation in tumour tissue ($P < 0.001$, GENT2 T-test), *PIGN* functions as a cancer chromosomal instability suppressor gene^{37, 38}. Although at lower levels, *SLC3A1* is also upregulated in PCa ($P < 0.001$, GENT2 T-test), with overexpression in breast cancer associated with tumourigenesis³⁹. These observations taken together provide the rationale for characterising the pLoF *OCA2* and *PIGN* DELs and *SLC3A1* IED as potentially pathogenic SVs (PP-SVs). Notably, all three SVs are reported as rare (irrespective of ancestry) in multiple population-wide studies including gnomAD SV³², 1000 genomes Project (1KGP)^{40, 41} and TOPMed SV⁴² (**Supplementary Data 1**).

Table 1
Candidate potentially pathogenic (PP) SVs identified in 170 PCa patients.

Genes	Gene impact type ¹	chrom1	pos1	chrom2	pos2	SV type	ClinVar / dbVar concordance	MAF African (this study)	MAF European (this study)	MAF African (dbVar) ²	MAF European (dbVar) ²
Potentially Pathogenic SV (PP-SV)											
<i>SLC3A1</i>	IED	chr2	44281377	chr2	44281612	DUP	Likely pathogenic	0.01 ³	0	0.0075	1.3e-04
<i>OCA2</i>	pLoF	chr15	28017719	chr15	28020677	DEL	Likely pathogenic	0.004	0	0.0015	0.001
<i>PIGN</i>	pLoF	chr18	62152637	chr18	62157701	DEL	Pathogenic	0.004	0	0.0013	1.3e-04
<i>SLC7A2</i>	pLoF	chr8	17418976	chr8	17544122	DEL	In dbVar	0.009	0	0.003	0
<i>DNAJC15</i>	pLoF	chr13	43078470	chr13	43079390	DEL	In dbVar	0	0.009	0	1.0e-04
<i>BCL2L11</i>	pLoF	chr2	111122626	chr2	111125901	DEL	novel	0.005	0	NA	NA
<i>BARD1</i>	pLoF	chr2	214768022	chr2	214772899	DEL	novel	0.005	0	NA	NA
<i>COL4A2/ COL4A1</i>	CG	chr13	110294204	chr13	110633815	DUP	In dbVar	0.005	0	1.3e-04	6.3e-06
<i>SLC2A5</i>	IED	chr1	9045605	chr1	9049441	DUP	In dbVar	0	0.009	7.3e-04	0.002
<i>FOXP1</i>	pLoF	chr3	71097066	chr3	74525618	INV	novel	0.009	0	NA	NA
<i>WASF1</i>	pLoF	chr6	108167886	chr6	110172775	INV	In dbVar	0.004	0	9.6e-05	0
<i>MLH1</i>	pLoF	chr3	37000362	chr3	39352689	INV	In dbVar	0.004	0	4e-04	6.4e-06
<i>RB1</i>	pLoF	chr13	48466588	chr13	48473911	INV	In dbVar	0.004	0	1.8e-04	1.3e-05
<i>CTNNA1</i>	pLoF	chr5	138903881	chr19	21614900	TRA	novel	0	0.009	NA	NA
<i>AK8-DST</i>	pLoF	chr9	132876361	chr6	56896165	TRA	novel	0	0.009	NA	NA
PP-SV candidates classified as 'cautionary'											
<i>LTBP1/ BIRC6</i>	CG	chr2	32403832	chr2	33107415	DUP	In dbVar	0	0.009	1.0e-04	0.0018
<i>PHC3- PRKACA</i>	pLoF	chr3	170090742	chr19	14110142	TRA	novel	0.004	0	NA	NA
<i>KCTD3- DST</i>	pLoF	chr1	215567414	chr6	56652607	TRA	novel	0.009	0	NA	NA
<i>PKHD1</i>	pLoF	chr6	51981375	chr15	30874073	TRA	novel	0.009	0	NA	NA

¹ Gene impact type based on gene annotation. pLoF: Potential loss-of-function. CG: Copy gain. IED: Intragenic Exon Duplication.

² The ancestry related MAF in dbVar were based on gnomAD³² or TOPMed⁴² SV study. The detail of all dbVar studies (dbVar study name and ID) and reported allele frequencies were shown in **Supplementary Data 1**.

³ Presenting at low-frequency rather than rare variants within the ancestrally-defined patient cohort.

Characterising candidate potentially pathogenic SVs absent from ClinVar

Among 1,843 SVs with unknown classification in ClinVar or absent from dbVar, we predicted their potential pathogenicity based on four SV impact prediction tools, including StrVCTVRE⁴³, CADD-SV⁴⁴, POSTRE⁴⁵ and PhenoSV⁴⁶. The number of scored SVs by four tools and their types were shown in **Supplementary Fig. 4** and **Supplementary Table 3**. Candidate SVs were required to meet two of the following criteria: StrVCTVRE score ≥ 0.37 , CADD-SV score ≥ 10 , POSTRE score ≥ 0.8 and/or PhenoSV score ≥ 0.5 (**Supplementary Table 4** and **Methods**). Based on this criterion, all three ClinVar identified pathogenic or likely pathogenic SVs and the single SV of uncertain significance were successfully annotated as pathogenic candidates, while conversely our workflow excluded for all 10 ClinVar characterised benign SVs (**Supplementary Table 5**). Using our criteria, 291 SVs were defined as PP-SV candidates (107 DELs, 16 DUPs, 11 INVs and 157 TRAs) disrupting 419 genes. In total 190 candidate SVs were private to African and 88 to European patients, with 13 shared between the ancestries (**Supplementary Table 4**).

To further define cancer-related pathogenic potential, we assessed for the presence of disrupted genes by PP-SV candidates in gene sets derived from the Human Molecular Signature Database (MSigDB) oncogenic signature and hallmark gene sets⁴⁷ and COSMIC Cancer Gene Census (COSMIC CGC) cancer driver genes⁴⁸. Requiring disrupted genes in two of the three cancer gene sets, 58 SVs were defined as cancer-related PP-SV candidates, including 20 DELs, 3 DUPs, 6 INVs and 29 TRAs, disrupting 56 genes. Of the 58 candidates, 23 of them were identified with MAF between 1–5% in either African or European patients, leaving 35 rare PP-SV candidates for further consideration, of which 16 have been reported in dbVar. Two dbVar SVs including TRA disrupting gene *NBEA* and *POLR2C* DEL were reported at low-frequencies (AF = 0.03 and 0.01, respectively) (**Supplementary Data 1**) and were therefore excluded from further analysis. Using our criteria, 33 rare cancer-related PP-SV candidates were identified (**Supplementary Data 2** and Fig. 1), including 15 DELs, 3 DUPs (1 IED and 2 CGs), 5 INVs and 10 TRAs.

Of the 15 pLoF DELs, 11 were excluded as PP-SVs, with impacting genes showing oncogenic behaviour in multiple cancer types or no strong evidence for their tumour suppressor effects (**Supplementary Table 7**). Conversely, four pLoF DELs were defined as PP-SVs, impacting known tumour suppressors or established DNA damage repair gene (**Supplementary Table 7**). Two of them are known to dbVar, including a *SLC7A2* 125,146 bp DEL identified in two African (**Supplementary Figs. 5**) and a *DNAJC15* 920 bp DEL in a European patient (**Supplementary Figs. 6**). Another two identified PP-SVs are novel pLoF DELs, which identified in a single African patient each, including a *BCL2L1* 3,275 bp (**Supplementary Fig. 7**) and DNA damage repair gene *BARD1* 4,877 bp DEL (Fig. 2A, **Supplementary Fig. 8**).

Of the two dbVar whole-gene DUPs, the *COL4A2* 339,611 bp CG, with breakpoints disrupting *COL4A1* and *NAXD*, observed in a single African patient is defined as a PP-SV (**Supplementary Fig. 9**), as *COL4A2* indicating oncogenic behaviour in gastric and breast cancers (**Supplementary Table 7**). In contrast, the *TTC27* 703,583 bp DUP observed in a single European patient is afforded 'cautionary' PP-SV status (**Supplementary Fig. 10**). Although *TTC27* is absent in three cancer gene databases, the breakpoints disrupt MSigDB and COSMIC CGC genes *BIRC6* and *LTBP1*, resulting in a *LTBP1-BIRC6* gene fusion of unclear effect. Observed in a single European patient, a 3,836 base DUP directly impacts exon 4 of *SLC2A5* (**Supplementary Fig. 11**), which downregulated in PCa ($P < 0.001$, GENT2 T-test) and has been identified an oncogenic behaviour (**Supplementary Table 7**), therefore allocated PP-SV status.

Of the five pLoF INVs, those impacting *MLH1*, *RB1* and *WASF1* are in dbVar, while *FOXP1* and *NSD3* INVs are novel. As *NSD3* has been identified as oncogenic in multiple cancers, the associated INV is classified here as unlikely pathogenic, with all remaining pLoF INVs classified as PP-SVs, as they disrupting known to PCa and Lynch Syndrome predisposing DNA mismatch repair gene *MLH1* and PCa tumour suppressor genes *RB1*, *WASF1*, and *FOXP1* (**Supplementary Table 7**). Identified in a single African patient each (**Supplementary Fig. 12–14**), the three dbVar INVs were reported as rare by the recent TOPMed SV study⁴², in which *WASF1* INV was also identified as African-specific (Table 1 and **Supplementary Data 1**). The novel INV impacting *FOXP1* was identified in two African patients (Fig. 2E, **Supplementary Fig. 15**).

Of the 10 pLoF TRAs, five impacting genes of *GRM8*, *WDR43*, *NPM1*, *NUSAP1* and *MECOM* with oncogenic properties (**Supplementary Table 7**), therefore are classified as unlikely pathogenic. *PKHD1* TRA identified in two African patients received a 'cautionary' PP-SV classification, as identified potential oncogenic in colon cancer, while potential tumour suppressor in colorectal cancer (**Supplementary Table 7**). As *CTNNA1* was known to have tumour suppressor behaviour across multiple tumour types (**Supplementary Table 7**), here we classify the European-specific pLoF *CTNNA1* TRA as a PP-SV (**Supplementary Fig. 16**). The remaining pLoF TRAs result in *PHC3-PRKACA* (1 African patient), *KCTD3-DST* (2 African patients) and *AK8-DST* (1 European patient, **Supplementary Fig. 17**) novel gene fusions. *PHC3-PRKACA* was classified as 'cautionary' PP-SV, as *PHC3* showed potential cancer suppressor effect in PCa, while *PRKACA* appears to portray oncogenic behaviour (**Supplementary Table 7**). Although unknown to PCa, both *DST* and *AK8* have demonstrated tumour suppressor behaviour, conversely, *KCTD3* with an unclear role in cancer (**Supplementary Table 7**). Here we classify *AK8-DST* as a PP-SV, while *KCTD3-DST* is assigned 'cautionary' PP-SV status.

Correlating PP-SVs and 'cautionary' PP-SVs with clinical features

The clinicopathological features of the study cohort has been previously described^{18, 28}. In brief, African patients show a 5-year greater mean age and 25-fold greater PSA level at diagnosis compared to European patients (**Supplementary Table 1**). Based on our previous observations²⁰, high-risk or aggressive PCa were defined as ISUP GG ≥ 3 and conversely, low-risk disease presentation as ISUP GG < 3 . Biased towards aggressive disease presentation (82% African, 86.0% European), it was notable that all four patients with a pathogenic or likely pathogenic SV presented with aggressive disease at diagnosis, 92.9% (13/14) of PP-SV and 83.3% (5/6) cautionary PP-SV presenting patients (Table 2).

Table 2
Clinicopathological features of patients by ethnicity presenting with potentially pathogenic (PP) SVs and cautionary PP-SVs as defined by this study criteria.

Gene name	Pathogenicity	SV type	Patient ID	Ethnicity	Age	PSA	ISUP GG	Family history
<i>SLC3A1</i>	PP-SV	DUP	N0001	African	75	22.9	4	
	Likely Pathogenic		SMU094	African	64	15	4	
<i>OCA2</i>	PP-SV	DEL	N0059	African	79	153	5	
	Likely Pathogenic							
<i>PIGN</i>	PP-SV	DEL	SMU083	African	86	40.5	3	
	Pathogenic							
<i>SLC7A2</i>	PP-SV	DEL	UP2035	African	70	680	5	
			KAL0054	African	64	42.9	5	
<i>DNAJC15</i>	PP-SV	DEL	17135	European	63	7.8	5	
<i>BCL2L11</i>	PP-SV	DEL	KAL0101	African	71	32.3	5	
<i>BARD1</i>	PP-SV	DEL	N0073	African	62	unknown	unknown	
<i>COL4A2/COL4A1</i>	PP-SV	DUP	UP2039	African	71	319	4	
<i>SLC2A5</i>	PP-SV	DUP	11099	European	70	9.9	5	
<i>FOXP1</i>	PP-SV	INV	UP2101	African	57	75	5	
			N0084	African	65	591	4	
<i>WASF1</i>	PP-SV	INV	N0048	African	70	83.3	5	
<i>MLH1</i>	PP-SV	INV	SMU080	African	64	23.3	4	Sister with cervical cancer
<i>RB1</i>	PP-SV	INV	SMU064	African	70	13.7	3	
<i>CTNNA1</i>	PP-SV	TRA	13179	European	59	8.4	5	
<i>AK8-DST</i>	PP-SV	TRA	11452	European	67	11	1	
<i>LTBP1/BIRC6</i>	Cautionary PP-SV	DUP	5287	European	54	4.3	5	
<i>PHC3-PRKACA</i>	Cautionary PP-SV	TRA	SMU061	African	65	12.1	3	Mother with stomach cancer
<i>KCTD3-DST</i>	Cautionary PP-SV	TRA	UP2039	African	71	319	4	
			SMU101	African	70	4.3	3	
<i>PKHD1</i>	Cautionary PP-SV	TRA	N0056	African	70	153	5	
			SMU196	African	47	9.5	1	

Discussion

ClinVar defined pathogenic (or likely pathogenic) SVs disrupting *SLC3A1*, *OCA2* or *PIGN* were observed in 3.5% (4/113) of African patients. Specifically, the *SLC3A1* intragenic exon DUP was identified in two patients presenting with ISUP GG4, while the *OCA2* and *PIGN* pLoF DELs presented in a single patient each with ISUP GG5 and ISUP GG3 PCa, respectively (Table 2). Visually inspecting the three PP-SVs using Integrative Genomic Viewer⁴⁹, *SLC3A1* DUP was found with three supporting read-pairs in sample N0001 (**Supplementary Fig. 3**), and split-reads and more than 40% increase in read depth comparing to ± 10 kb of the SV region in both samples (**Supplementary Table 6**), while *OCA2* and *PIGN* DELs were found with 16 and 6 supporting read-pairs respectively (**Supplementary Fig. 1–2**), and have 44–51% reduction in read depth (**Supplementary Table 6**). Solute carrier family 3 member 1 (*SLC3A1*) is an amino acid transporter, which through heterodimerisation with *SLC7A9* is responsible for cystine reabsorption through cationic and neutral amino acid exchange⁵⁰. Mutations, including SVs, in *SCL3A1* are associated with cystinuria, an inherited disease that results in the formation of cystine stones in the kidney, with disease presentation suggested to require biallelic loss⁵¹. *SCL3A1* over-expression has been associated with enhanced tumorigenesis in breast cancer, while blocking *SCL3A1* has suggestive therapeutic potential³⁹. *OCA2* is a pigmentation gene with inherited mutations associated with oculocutaneous albinism⁵². Polymorphisms have been associated with skin cancers⁵³, as well as clinical response and survival in breast cancer patients having received neoadjuvant chemotherapy⁵⁴. Inherited *PIGN* mutations have been associated with multiple congenital anomalies-hypotonia-seizures syndrome and Frys syndrome, with some mutations related to milder forms of clinical presentation^{55, 56}. Coding for phosphatidylinositol glycan anchor biosynthesis class N, *PIGN* is involved in the

biosynthesis of glycosylphosphatidylinositol, which has been shown to suppress cancer chromosomal instability³⁷ through PIGN complexed spindle assembly checkpoint regulation³⁸, a common phenomenon in solid tumours⁵⁷. Notably, no previous associations have been made between *SCL3A1*, *OCA2* or *PIGN* mutation and PCa.

As our study is biased towards under-represented African patients, it is highly plausible that the majority of SVs detected are unlikely to be represented in ClinVar. As such, it is critical that we developed a best-fit workflow for PP-SV prediction. The four SV impact prediction tools used in this study were chosen based on the criteria of easy-to-use (either web-based or packed as software), providing pathogenicity scores or labels, accepting multiple SVs and covering all SV types. However, there are multiple factors to be taken into consideration when using SV impact prediction tools to establish potential pathogenicity, as different tools have limitations in applicable SV types, regions or diseases, as well as different scoring systems. While all tools can predict the impact of DELs and DUPs, StrVCTVRE is limited to DELs and DUPs in exonic regions. Besides predicting the simpler SVs, CADD-SV is capable of annotating INSVs and POSTRE annotates INSVs and TRAs, while PhenoSV is able to predict the impact of all these three types. POSTRE doesn't work for all diseases or phenotypes. Therefore, combining multiple tools is necessary to cover all SV types and increase the confidence level. Another factor is the choice of threshold to establish pathogenicity. POSTRE and PhenoSV defines the threshold of pathogenicity, but StrVCTVRE and CADD-SV are limited to scores and calling for thresholds to be established depending on individual study aims. In this study, we have decided the thresholds based on tools' validated results from database (90% sensitivity in ClinVar by StrVCTVRE⁴³ and top 10% in gnomAD by CADD-SV⁴⁴). When combining results from multiple tools, we found the requirement of passing thresholds of all four tools identified two PP-SV candidates (out of 1,843 SVs) (**Supplementary Table 4**), with notable failure to identify the three ClinVar pathogenic/likely pathogenic SVs (**Supplementary Table 5**). As such, PP-SV candidate classification in this study required an SV to pass thresholds of at least two impact prediction tools, with disrupted genes requiring further clarification as hallmark or drivers in cancer gene databases (MSigDB and COSMIC CGC).

Using our described workflow, 12 SVs were predicted as PP-SVs, identified in 7.0% (4/57) of European and 8.8% (10/113) of African patients, bringing the total of African patients presenting with a potential pathogenic SV to 12.4% (14/113). Remarkably, five of our African-specific PP-SVs included well-known pathogenic cancer genes and/or PCa tumour suppressor genes, including DNA damage response genes. Most notably, the DNA mismatch repair tumour suppressor gene *MLH1* commonly mutated in Lynch Syndrome, including cases with PCa⁵⁸, is a known candidate gene in PCa germline testing panels²⁰. While PCa patients presenting with pathogenic *MLH1* mutations were reported to have significantly higher disease burden for African Americans²⁴, here we found a dbVar known *MLH1* pLoF INV with around 11 supporting short read-pairs (**Supplementary Fig. 12**) in a 64 year old African male presenting with ISUP GG4 at diagnosis. Not recognised as a PCa germline testing panel gene, *FOXP1* is an established PCa tumour suppressor driver gene, with CN loss increasing cell proliferation and migration, and poor prognosis⁵⁹. Recently, we showed *FOXP1* to be equally impacted by predominantly CN loss in African compared with European derived tumours (20% of 183 tumours)¹⁸. Here we found a germline inverted duplication impacting *FOXP1* with around 18 supporting read-pairs in two African patients (**Supplementary Fig. 15**). Notably, one African patient (UP2101) presented 10 years earlier than the cohort average receiving an ISUP GG5 diagnosis. Loss of the *BRAC1* associated RING domain-1 (*BARD1*) DNA damage repair gene has been found to induce homologous recombination deficiency and increase the sensitivity to PARP inhibitor in PCa cell lines⁶⁰. Here the novel *BARD1* exon 5 DEL, supported by 10 read-pairs and with around 50% reduction in read depth comparing to ± 10 kb of the DEL region (**Supplementary Fig. 8, Supplementary Table 6**), was identified in a 62-year-old African PCa patient with unknown pathology. While a paediatric cancer predisposing tumour suppressor gene commonly mutated in retinoblastoma and to a lesser extent osteosarcoma⁶¹, and less common as an adult cancer predisposing gene⁶², *RB1* is recognised as one of five most prevalent somatically mutated genes in metastatic cancers⁶³, with *RB1* loss in prostate tumours associated with poor patient outcomes⁶⁴. To the best of our knowledge, this is the first report of a germline potentially pathogenic *RB1* PCa variant, which includes a pLoF INV of exon 24 with three supporting read-pairs (**Supplementary Fig. 13**) in a single ISUP GG3 diagnosed African patient. Lastly, the tumour suppressor gene *WASF1* with loss associated with aggressive or metastatic lethal PCa⁶⁵. Identifying a potentially pathogenic INV previously reported at MAF of 9.6e-05 in Africans and resulting in *NR2E1-WASF1* fusion was identified in a single African patient presenting at 70 years of age with ISUP GG5 PCa, showed 14 supporting read-pairs (**Supplementary Fig. 14**).

Other notable PP-SV DELs impacting tumour suppressor genes unknown to PCa, includes *SLC7A2* and *DNAJC15*. Knockdown of *SLC7A2* has been shown to promote viability, invasion and migration of ovarian cancer⁶⁶ and enhance proliferation of non-small-cell lung cancer cells⁶⁷, while *DNAJC15* has tumour suppressor behaviour in breast cancer⁶⁸. Identified in two African patients presenting with ISUP GG5 disease, loss of *SLC7A2* exons 1 and 2, supported by 10 read-pairs and with around 50% reduction in read depth (**Supplementary Fig. 5 and Supplementary Table 6**) has previously been reported in African populations at MAF of 0.03 (**Supplementary Data 1**). Specific to Europeans (MAF = 1.0e-04), loss of *DNAJC5* exon 4 supported by 13 read-pairs and with around 50% reduction in read depth (**Supplementary Fig. 6 and Supplementary Table 6**), was identified in a single European patient presenting for surgery at age 63 years with ISUP GG5 disease. While not associated with PCa, the loss of *BCL2L11* and *CTNNA1* has been identified to leading tumourigenesis and promoting invasion and metastasis of multiple cancers^{69,70}. Here the novel *BCL2L11* pLoF DEL on exon 2 with more than 20 supporting read-pairs and with around 50% reduction in read depth (**Supplementary Fig. 7 and Supplementary Table 6**) was identified in a single African patient presenting at age 71 years with ISUP GG5 PCa, while the novel pLoF TRA interrupting *CTNNA1* with more than 20 supporting read-pairs (**Supplementary Fig. 16**) in a single European patient presenting at age 59 years with ISUP GG5 PCa. Another identified novel potentially pathogenic inter-chromosomal TRA with around 18 supporting read-pairs (**Supplementary**

Fig. 17) leading to a *AK8-DST* fusion in a single European patient (ISUP GG1, 67 years). Although no associations have been made between PCa, higher expression of *DST* has been identified to promote pathogenesis and development of breast cancer, while *AK8* downregulation has been found to promote migration and invasion of uterine carcinosarcoma⁷¹.

Two known PP-SVs identified to potentially increase gene dosage of well-known oncogenes *COL4A2* and *SLC2A5*, through whole-gene duplication and intra-genic exon duplication respectively. Although not associated with PCa, *COL4A2* loss has been identified to inhibit triple-negative breast cancer cell proliferation and migration⁷² and its mutations as risk factor for familial cerebrovascular disease⁷³, while inactivation of *SLC2A5* has been found to inhibit cell proliferation and migration in multiple cancer cell lines⁷⁴. The whole *COL4A2* DUP with more than 20 supporting read-pairs and more than 50% gain in read depth (**Supplementary Fig. 9** and **Supplementary Table 6**) was identified in a single African patient (ISUP GG4, 71 years) and the exon 4 DUP in *SLC2A5* with more than 20 supporting read-pairs and more than 50% increase in read depth (**Supplementary Fig. 11** and **Supplementary Table 6**) was identified in a single European patient (ISUP GG5, 70 years).

Using short-read sequencing data for SV calling and genotyping remains a potential limitation, appreciating that SVs in difficult-to-sequence regions may have been overlooked⁷⁵. To ensure the highest possible accuracy of SV detection and population allele frequency estimation, we required high-confidence calls from two SV callers and high-quality genotype calls at both the population- and individual-level, while all PP-SVs were visually inspected. Due to lack of available expression data, we were unable to validate the direct impact of identified PP-SVs and cautionary PP-SVs. Further guidelines related to criteria for pathogenic SV identification using short read sequencing technologies and/or long read sequencing approaches are required, making these methods accessible for routine germline testing.

Conclusion

Here we have described a first-of-its-kind pathogenicity investigation of SVs in PCa patients with ancestry disparity. We observed three ClinVar-defined pathogenic or likely pathogenic PP-SVs (*SLC3A1*, *OCA2* and *PIGN*) and 12 predicted PP-SVs, including seven known SVs (*SLC7A2*, *DNAJC15*, *COL4A2*, *SLC2A5*, *WASF1*, *MLH1* and *RB1*), and five novel SVs (*BCL2L11*, *BARD1*, *FOXP1*, *CTNNA1* and *AK8-DST*), suggesting that inherited SVs may constitute an under-appreciated contribution to PCa pathogenicity. Furthermore, the identification of African-private (eight known and three novel) and European-private (two known and two novel) PP-SVs allows for further speculation with regards to associated racial disparities, while improving the detection rate for PCa germline testing with SV inclusivity, and in turn raising limitations for African inclusion and associated clinical care.

Methods

WGS data generation

To avoid technical and analytical biases, all samples (whole blood) were processed (beginning at DNA extraction), data generated and analysed within a single laboratory using a single computational pipeline, as previously described^{18, 28}. In brief, whole-genome sequencing data were generated using Illumina HiSeq X Ten (21 cases) or NovoSeq (149 cases) instruments with 2×150 cycle paired-end mode at the Kinghorn Centre for Clinical Genomics (Garvan Institute of Medical Research, Australia). Following the BROAD's best practice recommendations for "data pre-processing for variant discovery", sequencing reads were aligned to GRCh38 reference genome with alternative contigs using scalable FASTQ-to-BAM (v2.0) workflow with default settings⁷⁶. The mean depth of coverage for all samples were 45.9X (range 30.2-97.6X).

Structural variant calling and high-confidence SV filtering

Germline SVs were called using Manta (v1.6.0)⁷⁷ and GRIDSS (v2.13.3)^{78, 79}. SV types reported by Manta included DEL, tandem DUP, INS and adjacent breakends (BNDs) for a fusion junction with inverted sequence or in an inter-chromosomal rearranged genome. Pairs of BND in inverted junction were annotated as inversions (INV). Pairs of BND in different chromosomes were annotated as inter-chromosomal translocations (TRA). Conversely, GRIDSS reports BND for all fusion junctions resulting from any SV event. Simple SV types, defined as DEL, DUP, INS, INV and TRA, were assigned based on the strands and ALT field in VCF (modified from GRIDSS accompanied R script: simple-event-annotation.R). To obtain high-confidence SV call set, we integrated call sets from Manta and GRIDSS and generated concordant call set for each genome. Two SV calls were considered as concordant if they were reported as "PASS" by one of the two callers and have matching SV type and reported breakpoint positions within 200bp of each other. *Bedtools pairtopair*⁸⁰ was used to compare two call sets.

Population-level genotyping and high-confidence genotype call filtering

We used Graphtyper2 (v2.7.5)⁸¹ to re-genotype SVs for all samples. Following published guidelines, we merged all high-confidence SV set per-sample (individual VCFs) using svimmer (<https://github.com/DecodeGenetics/svimmer>) with default parameters. The individual VCFs were in format of Manta VCFs, as Manta provides detailed information on the exact breakpoint sequence, which is the essential information required by Graphtyper2. We extracted all SVs with "aggregate" model as suggested, and obtained 57,096 SVs with "PASS" in FILTER field in VCF. To further filtering SV genotype calls on a per-sample basis, we required more than 50% genotype calls as "PASS" (PASS_ratio ≥ 0.5 in INFO field), resulting in 42,966 SVs.

To further filtering SV genotype calls on a per-sample basis, we set SV genotype as missing if genotype filter tag (FT) is not “PASS” for all SVs, except BND. For BND, as FT tag is not available, we set BND genotype with genotype quality (GQ) < 20 as missing. We then excluded SVs with genotype missingness rate > 20% in either African or European genomes, resulting in 33,340 SVs. We further removed 97 SVs with allele frequency of 100%, indicating the difference of sample genomes to reference genome. The allele frequency of each SV was then calculated based on the high-quality genotype calls only.

Gene annotation and functional impact of SVs

All SVs were annotated against gene regions from the Ensembl human gene annotation file (GRCh38 assembly, release 108). As multiple transcripts can be available for a single gene, the Ensembl Canonical transcript was used (<http://www.ensembl.org/info/genome/genebuild/canonical.html>). By comparing the position of SV breakpoint with gene regions using bedtools⁸⁰, we examined nine gene overlapping categories with gnomAD³², including potential Loss of Function (pLoF), Copy Gain (CG), Intragenic Exon DUP (IED), partial gene DUP, whole-gene INV, UTR SVs, promoter SVs, intronic SVs and intergenic SVs. In addition, we defined partial-exon DUP as both breakpoints contained within the same gene, while neither both within exons (pLoF) nor fully overlapped at least one exon (IED). Promoters were defined as 1kb window before each transcription start site on the transcribed strand. We labelled SVs as enhancer-disruptive if at least one breakpoint was contained within a gene’s enhancer, by comparing to GeneHancer⁸² regulatory elements regions. GeneHancer regulatory elements and gene interactions “double elite” subset was downloaded from UCSC Table *geneHancerInteractionsDoubleElite* [last updated 15/01/2019] from GeneHancer track for GRCh38. The transcript structure plots were generated based on Ensembl human gene annotation (GRCh38 assembly, release 108) using R package ggtranscript (v0.99.3)⁸³. The sequencing depth of DEL or DUP regions and their ± 10 kb regions were calculated using samtools (v1.6) *depth* command⁸⁴.

Short-read data detect the SV signatures from aligned reads around the SV breakpoints and is hard to capture the whole large SVs⁸⁵. Therefore, we restricted the disrupted genes of SVs greater than 1Mbp to be genes overlapped by SV breakpoints for downstream analysis.

Identification of dbVar concordance and novel SVs

The NCBI’s database of human genomic structural variation (dbVar) [last updated 30/10/2023]⁸⁶ were used to identify dbVar concordance and novel SVs. The dbVar database included a total of 6,476,337 unique SVs, including 86,686 SVs with interpretations of their significance to disease in ClinVar database⁸⁷. Structural variants concordant to dbVar SVs were defined as having both breakpoints within 200 bases of dbVar defined SV breakpoints. The ancestry related variant allele frequency of SVs (**Supplementary Data 1**) were derived from dbVar pages of SVs or VCFs uploaded by different dbVar studies to dbVar’s FTP site.

Pathogenicity prediction

The pathogenicity of SVs were predicted through prediction tools StrVCTVRE⁴³, CADD-SV⁴⁴, POSTRE⁴⁵ and PhenoSV⁴⁶. StrVCTVRE only scores the deleteriousness of DEL and DUP overlapping one or more exons, CADD-SV scores DEL, DUP and INS, POSTRE predicts the impact of DEL, DUP, INV and TRA, and PhenoSV works for all five SV types. As POSTRE only accepts genome coordinates on reference genome Hg19, the *liftOver* function from *rtracklayer* package in R was used to lift SV coordinates from Hg38 to Hg19. As suggested by StrVCTVRE, the ClinVar 90% sensitivity threshold (0.37) was used to define potentially pathogenic SVs. The scaled CADD-SV scores range from 0 (potentially benign) to 48 (potentially pathogenic), indicating the position of the input SV within the gnomAD-SV score distribution. The threshold of 10 for CADD-SV score was used to establish potential pathogenicity, corresponding to top 10% score observed in gnomAD-SV. The threshold of 0.8 and 0.5 for POSTRE and PhenoSV score respectively was used in this study, which is the threshold of pathogenicity labelling defined by POSTRE and PhenoSV.

The hallmark gene sets and oncogenic signature gene sets were downloaded from the Human Molecular Signature Database (MSigDB v2023.1)⁴⁷. The MSigDB oncogenic signature gene sets included genes representing signatures of cellular pathways which are often dysregulated in cancer. Cancer-driver genes were downloaded from COSMIC Cancer Gene Census (GRCh38 COSMIC v98, downloaded 26/09/2023).

Abbreviations

PCa
Prostate cancer
SV
Structural variation
DEL
Deletion
INS
Insertion
DUP
Duplication
INV
Inversion

TRA
Translocation
WGS
Whole genome sequencing
PSA
Prostate Specific Antigen
BND
Breakend
ISUP GG
International Society of Urological pathology Group Grading
pLoF
potential Loss of Function
CG
Copy Gain
IED
Intragenic Exon Duplication
PPSV
potentially pathogenic SV
NCCN
National Comprehensive Cancer Network
1KGP
1000 genomes Project Phase3
COSMIC CGC
COSMIC Cancer Gene Census.

Declarations

Ethics approval and consent to participate

Irrespective of country of origin, all individuals provided informed consent to participate in the study. Conforming to the principles of the Helsinki Declaration, South African patients were recruited as part of the Southern African Prostate Cancer Study (SAPCS) with approval granted by the University of Pretoria Faculty of Health Sciences Research Ethics Committee (HREC, with US Federal wide assurance FWA00002567 and IRB00002235 IORG0001762; #43/2010). In Australia, participant recruitment was approved by the St Vincent's HREC (#SVH/12/231). Samples were shipped to the Garvan Institute of Medical Research in accordance with institutional Material Transfer Agreements (MTAs), as well as additional Republic of South Africa Department of Health Export Permit (National Health Act 2003; J1/2/4/2 #1/12). This study was approved by the St. Vincent's HREC (#SVH/15/227) for genomic interrogation. Additional IRB review and approval for genomic interrogation was granted by the Human Research Protection Office of the US Army Medical Research and Development Command E02371 (TARGET Africa) and E03280 (HEROIC PCaPH Africa1K).

Consent for publication

Not applicable

Availability of data and materials

The sequence data were obtained and accessible through Jaratlerdsiri et al¹⁸, in the European Genome-Phenome Archive (EGA; <https://ega-archive.org>) under overarching accession EGAS00001006425 and including the Southern African Prostate Cancer Study (SAPCS) Dataset (EGAD00001009067) and Garvan/St Vincent's Prostate Cancer Database (EGAD00001009066). The dbVar SV sites and their variant allele frequencies were downloaded from <https://www.ncbi.nlm.nih.gov/dbvar>. The ENSEMBL gene set was downloaded from <https://www.ensembl.org>. The MSigDB gene sets are available at <https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp#H>. The gene set in COCMIC CGC was downloaded from <https://cancer.sanger.ac.uk/census>. The scripts for sequence read alignment and quality control are available at GitHub (<https://github.com/Sydney-Informatics-Hub/Bioinformatics>). All computational code for SV callset comparison and integration are available at GitHub (<https://github.com/tgong1/StructuralVariantUtil>)⁸⁸.

Competing Interests

The authors declare that they have no competing interests.

Funding

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.; APP2APP2010551 to V.M.H.). Further analytics was supported by a U.S.A. Congressionally Directed Medical Research Programs (CDMRP) Prostate Cancer Research Program (PCRP) Idea Development Award (PC200390, TARGET Africa to V.M.H.) and HEROIC Consortium Award (PC210168, HEROIC PCaPH Africa1K to V.M.H. and M.S.R.B., as well as 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

Conception and design: T.G. and V.M.H.; Financial support: V.M.H.; Methodology: T.G. and J.J.; Formal analysis: T.G., J.J. and V.M.H.; Data curation: W.J., K.G. and J.J.; Participant recruitment, clinical data and sample collection: P.D.S., S.B.A.M. and M.S.R.B.; Supervision: V.M.H.; Writing-review, figures and editing: T.G. and V.M.H. Critical technical review: J.W. All authors have read and approved the final manuscript.

Acknowledgements

We are forever grateful to the patients and their families who have contributed to this study; without their contribution, this research would not be possible. We acknowledge the contributions of the many clinical staff across the SAPCS (South Africa) and the St Vincent's Hospital Sydney (Australia), who over many years have recruited patients and provided samples to these critical bioresources, as well as the additional authors who contributed to the initial published genome profiling project (Jaratlerdsiri *et al.*, 2022)¹⁸.

References

1. Bray F, *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **n/a**.
2. Hjelmborg JB, *et al.* The Heritability of Prostate Cancer in the Nordic Twin Study of Cancer. *Cancer Epidemiology, Biomarkers & Prevention* **23**, 2303-2310 (2014).
3. Smith ZL, Eggener SE, Murphy AB. African-American Prostate Cancer Disparities. *Current Urology Reports* **18**, 81 (2017).
4. Taitt HE. Global Trends and Prostate Cancer: A Review of Incidence, Detection, and Mortality as Influenced by Race, Ethnicity, and Geographic Location. *Am J Mens Health* **12**, 1807-1823 (2018).
5. Giaquinto AN, Miller KD, Tossas KY, Winn RA, Jemal A, Siegel RL. Cancer statistics for African American/Black People 2022. *CA: A Cancer Journal for Clinicians* **72**, 202-229 (2022).
6. Mahal BA, *et al.* Prostate Cancer Racial Disparities: A Systematic Review by the Prostate Cancer Foundation Panel. *European Urology Oncology* **5**, 18-29 (2022).
7. Darst BF, Sheng X, Eeles RA, Kote-Jarai Z, Conti DV, Haiman CA. Combined Effect of a Polygenic Risk Score and Rare Genetic Variants on Prostate Cancer Risk. *European Urology* **80**, 134-138 (2021).
8. Seibert TM, *et al.* Genetic Risk Prediction for Prostate Cancer: Implications for Early Detection and Prevention. *European Urology* **83**, 241-248 (2023).
9. de Bono J, *et al.* Olaparib for Metastatic Castration-Resistant Prostate Cancer. *New England Journal of Medicine* **382**, 2091-2102 (2020).
10. Abida W, *et al.* Rucaparib in Men With Metastatic Castration-Resistant Prostate Cancer Harboring a BRCA1 or BRCA2 Gene Alteration. *Journal of Clinical Oncology* **38**, 3763-3772 (2020).
11. Lozano R, *et al.* Genetic aberrations in DNA repair pathways: a cornerstone of precision oncology in prostate cancer. *British Journal of Cancer* **124**, 552-563 (2021).
12. National Comprehensive Cancer Network Clinical Guidelines in Oncology (NCCN Guidelines®): Prostate Cancer (Version 4.2023).
13. Giri VN, *et al.* Implementation of Germline Testing for Prostate Cancer: Philadelphia Prostate Cancer Consensus Conference 2019. *Journal of Clinical Oncology* **38**, 2798-2811 (2020).
14. Briggs LG, *et al.* Racial Differences in Germline Genetic Testing for Prostate Cancer: A Systematic Review. *JCO Oncology Practice* **19**, e784-e793 (2023).
15. Mahal BA, *et al.* Racial Differences in Genomic Profiling of Prostate Cancer. *New England Journal of Medicine* **383**, 1083-1085 (2020).
16. Valle LF, *et al.* Actionable Genomic Alterations in Prostate Cancer Among Black and White United States Veterans. *Oncologist* **28**, e473-e477 (2023).
17. White JA, *et al.* Whole-exome Sequencing of Nigerian Prostate Tumors from the Prostate Cancer Transatlantic Consortium (CaPTC) Reveals DNA Repair Genes Associated with African Ancestry. *Cancer Res Commun* **2**, 1005-1016 (2022).
18. Jaratlerdsiri W, *et al.* African-specific molecular taxonomy of prostate cancer. *Nature* **609**, 552-559 (2022).

19. Giri VN, Hartman R, Pritzlaff M, Horton C, Keith SW. Germline Variant Spectrum Among African American Men Undergoing Prostate Cancer Germline Testing: Need for Equity in Genetic Testing. *JCO Precision Oncology*, e2200234 (2022).
20. Gheybi K, *et al.* Evaluating Germline Testing Panels in Southern African Males With Advanced Prostate Cancer. *Journal of the National Comprehensive Cancer Network* **21**, 289-296.e283 (2023).
21. Soh PXY, Hayes VM. Common Genetic Variants Associated with Prostate Cancer Risk: The Need for African Inclusion. *European Urology* **84**, 22-24 (2023).
22. Soh PXY, *et al.* Prostate cancer genetic risk and associated aggressive disease in men of African ancestry. *Nature Communications* **14**, 8037 (2023).
23. Tindall EA, *et al.* Clinical presentation of prostate cancer in Black South Africans. *Prostate* **74**, 880-891 (2014).
24. Matejcic M, *et al.* Pathogenic Variants in Cancer Predisposition Genes and Prostate Cancer Risk in Men of African Ancestry. *JCO Precision Oncology*, 32-43 (2020).
25. Pritchard CC, *et al.* Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *New England Journal of Medicine* **375**, 443-453 (2016).
26. The Cancer Genome Atlas Research N. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011-1025 (2015).
27. Ryan MJ, Bose R. Genomic Alteration Burden in Advanced Prostate Cancer and Therapeutic Implications. *Front Oncol* **9**, 1287-1287 (2019).
28. Gong T, *et al.* Genome-wide interrogation of structural variation reveals novel African-specific prostate cancer oncogenic drivers. *Genome Medicine* **14**, 100 (2022).
29. Pagnamenta AT, *et al.* Structural and non-coding variants increase the diagnostic yield of clinical whole genome sequencing for rare diseases. *Genome Medicine* **15**, 94 (2023).
30. Thibodeau ML, *et al.* Improved structural variant interpretation for hereditary cancer susceptibility using long-read sequencing. *Genet Med* **22**, 1892-1897 (2020).
31. Dixon K, *et al.* Defining the heterogeneity of unbalanced structural variation underlying breast cancer susceptibility by nanopore genome sequencing. *Eur J Hum Genet* **31**, 602-606 (2023).
32. Collins RL, *et al.* A structural variation reference for medical and population genetics. *Nature* **581**, 444-451 (2020).
33. Park S-J, Yoon B-H, Kim S-K, Kim S-Y. GENT2: an updated gene expression database for normal and tumor tissues. *BMC Medical Genomics* **12**, 101 (2019).
34. Pellikaan K, *et al.* Malignancies in Prader-Willi Syndrome: Results From a Large International Cohort and Literature Review. *The Journal of Clinical Endocrinology & Metabolism* **108**, e1720-e1730 (2023).
35. De Summa S, *et al.* The Genetic Germline Background of Single and Multiple Primary Melanomas. *Front Mol Biosci* **7**, 555630 (2020).
36. Cole-Clark D, Nair-Shalliker V, Bang A, Rasiah K, Chalasani V, Smith DP. An initial melanoma diagnosis may increase the subsequent risk of prostate cancer: Results from the New South Wales Cancer Registry. *Sci Rep* **8**, 7167 (2018).
37. Burrell RA, *et al.* Replication stress links structural and numerical cancer chromosomal instability. *Nature* **494**, 492-496 (2013).
38. Teye EK, *et al.* PIGN spatiotemporally regulates the spindle assembly checkpoint proteins in leukemia transformation and progression. *Scientific Reports* **11**, 19022 (2021).
39. Jiang Y, *et al.* Cysteine transporter SLC3A1 promotes breast cancer tumorigenesis. *Theranostics* **7**, 1036-1046 (2017).
40. Auton A, *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
41. Byrsk-Bishop M, *et al.* High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426-3440.e3419 (2022).
42. Jun G, *et al.* Structural variation across 138,134 samples in the TOPMed consortium. *bioRxiv*, (2023).
43. Sharo AG, Hu Z, Sunyaev SR, Brenner SE. StrVCTVRE: A supervised learning method to predict the pathogenicity of human genome structural variants. *The American Journal of Human Genetics* **109**, 195-209 (2022).
44. Kleinert P, Kircher M. A framework to score the effects of structural variants in health and disease. *Genome Res* **32**, 766-777 (2022).
45. Sánchez-Gaya V, Rada-Iglesias A. POSTRE: a tool to predict the pathological effects of human structural variants. *Nucleic Acids Research* **51**, e54-e54 (2023).
46. Xu Z, Li Q, Marchionni L, Wang K. PhenoSV: interpretable phenotype-aware model for the prioritization of genes affected by structural variants. *Nature Communications* **14**, 7805 (2023).
47. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).
48. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nature Reviews Cancer* **18**, 696-705 (2018).
49. Robinson JT, Thorvaldsdottir H, Wenger AM, Zehir A, Mesirov JP. Variant Review with the Integrative Genomics Viewer. *Cancer Res* **77**, e31-e34 (2017).

50. Nagamori S, *et al.* Novel cystine transporter in renal proximal tubule identified as a missing partner of cystinuria-related plasma membrane protein rBAT/SLC3A1. *Proceedings of the National Academy of Sciences* **113**, 775-780 (2016).
51. Martell HJ, Wong KA, Martin JF, Kassam Z, Thomas K, Wass MN. Associating mutations causing cystinuria with disease severity with the aim of providing precision medicine. *BMC Genomics* **18**, 550 (2017).
52. Okulicz JF, Shah RS, Schwartz RA, Janniger CK. Oculocutaneous albinism. *J Eur Acad Dermatol Venereol* **17**, 251-256 (2003).
53. Roberts MR, Asgari MM, Toland AE. Genome-wide association studies and polygenic risk scores for skin cancer: clinically useful yet? *British Journal of Dermatology* **181**, 1146-1155 (2019).
54. Li XP, *et al.* OCA2 rs4778137 polymorphism predicts survival of breast cancer patients receiving neoadjuvant chemotherapy. *Gene* **651**, 161-165 (2018).
55. Fleming L, *et al.* Genotype-phenotype correlation of congenital anomalies in multiple congenital anomalies hypotonia seizures syndrome (MCAHS1)/PIGN-related epilepsy. *Am J Med Genet A* **170a**, 77-86 (2016).
56. Jezela-Stanek A, Mierzevska H, Szczepanik E. Vertical nystagmus as a feature of PIGN-related glycosylphosphatidylinositol biosynthesis defects. *Clin Neurol Neurosurg* **196**, 106033 (2020).
57. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature* **396**, 643-649 (1998).
58. Haraldsdottir S, *et al.* Prostate cancer incidence in males with Lynch syndrome. *Genet Med* **16**, 553-557 (2014).
59. Cai H, *et al.* In Vivo Application of CRISPR/Cas9 Revealed Implication of Foxa1 and Foxp1 in Prostate Cancer Proliferation and Epithelial Plasticity (2022).
60. Dillon KM, *et al.* PALB2 or BARD1 loss confers homologous recombination deficiency and PARP inhibitor sensitivity in prostate cancer. *npj Precision Oncology* **6**, 49 (2022).
61. Zhang J, *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med* **373**, 2336-2346 (2015).
62. Huang K-I, *et al.* Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell* **173**, 355-370.e314 (2018).
63. Robinson DR, *et al.* Integrative clinical genomics of metastatic cancer. *Nature* **548**, 297-303 (2017).
64. Abida W, *et al.* Genomic correlates of clinical outcome in advanced prostate cancer. *Proceedings of the National Academy of Sciences* **116**, 11428-11436 (2019).
65. Sowalsky AG, *et al.* Loss of Wave1 gene defines a subtype of lethal prostate cancer. *Oncotarget; Vol 6, No 14*, (2015).
66. Sun T, Bi F, Liu Z, Yang Q. SLC7A2 serves as a potential biomarker and therapeutic target for ovarian cancer. *Aging (Albany NY)* **12**, 13281-13296 (2020).
67. Jiang S, *et al.* Lower SLC7A2 expression is associated with enhanced multidrug resistance, less immune infiltrates and worse prognosis of NSCLC. *Cell Communication and Signaling* **21**, 9 (2023).
68. Alessandrini F, Pezzè L, Menendez D, Resnick MA, Ciribilli Y. ETV7-Mediated DNJC15 Repression Leads to Doxorubicin Resistance in Breast Cancer Cells. *Neoplasia* **20**, 857-870 (2018).
69. Huang J, *et al.* The Role of CTNNA1 in Malignancies: An Updated Review. *J Cancer* **14**, 219-230 (2023).
70. Zhang H, *et al.* Onco-miR-24 regulates cell growth and apoptosis by targeting BCL2L1 in gastric cancer. *Protein Cell* **7**, 141-151 (2016).
71. Huang R, *et al.* Co-expression Analysis of Genes and Tumor-Infiltrating Immune Cells in Metastatic Uterine Carcinosarcoma. *Reproductive Sciences* **28**, 2685-2698 (2021).
72. JingSong H, *et al.* siRNA-mediated suppression of collagen type iv alpha 2 (COL4A2) mRNA inhibits triple-negative breast cancer cell proliferation and migration. *Oncotarget* **8**, 2585-2593 (2017).
73. Verbeek E, *et al.* COL4A2 mutation associated with familial porencephaly and small-vessel disease. *European Journal of Human Genetics* **20**, 844-851 (2012).
74. Groenendyk J, *et al.* Loss of the fructose transporter SLC2A5 inhibits cancer cell migration. *Front Cell Dev Biol* **10**, 896297 (2022).
75. Porubsky D, Eichler EE. A 25-year odyssey of genomic technology advances and structural variant discovery. *Cell*, (2024).
76. Sadsad R, Samaha G, Chew T. Fastq-to-bam @ NCI-Gadi [Internet].). WorkflowHub (2021).
77. Chen X, *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-1222 (2016).
78. Cameron DL, *et al.* GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res* **27**, 2050-2060 (2017).
79. Cameron DL, *et al.* GRIDSS2: comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing. *Genome Biology* **22**, 202 (2021).
80. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
81. Eggertsson HP, *et al.* GraphTyper2 enables population-scale genotyping of structural variation using pangenome graphs. *Nature Communications* **10**, 5402 (2019).

82. Fishilevich S, *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database* **2017**, bax028 (2017).

83. Gustavsson EK, Zhang D, Reynolds RH, Garcia-Ruiz S, Ryten M. ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2. *Bioinformatics* **38**, 3844-3846 (2022).

84. Danecek P, *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).

85. Gong T, Hayes VM, Chan EKF. Detection of somatic structural variants from short-read next-generation sequencing data. *Briefings in Bioinformatics* **22**, bbaa056 (2021).

86. Lappalainen I, *et al.* DbVar and DGVa: public archives for genomic structural variation. *Nucleic Acids Res* **41**, D936-941 (2013).

87. Landrum MJ, *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research* **46**, D1062-D1067 (2018).

88. Gong T. StructuralVariantUtil. *GitHub*, (<https://github.com/tgong1/StructuralVariantUtil> (2022)).

Figures

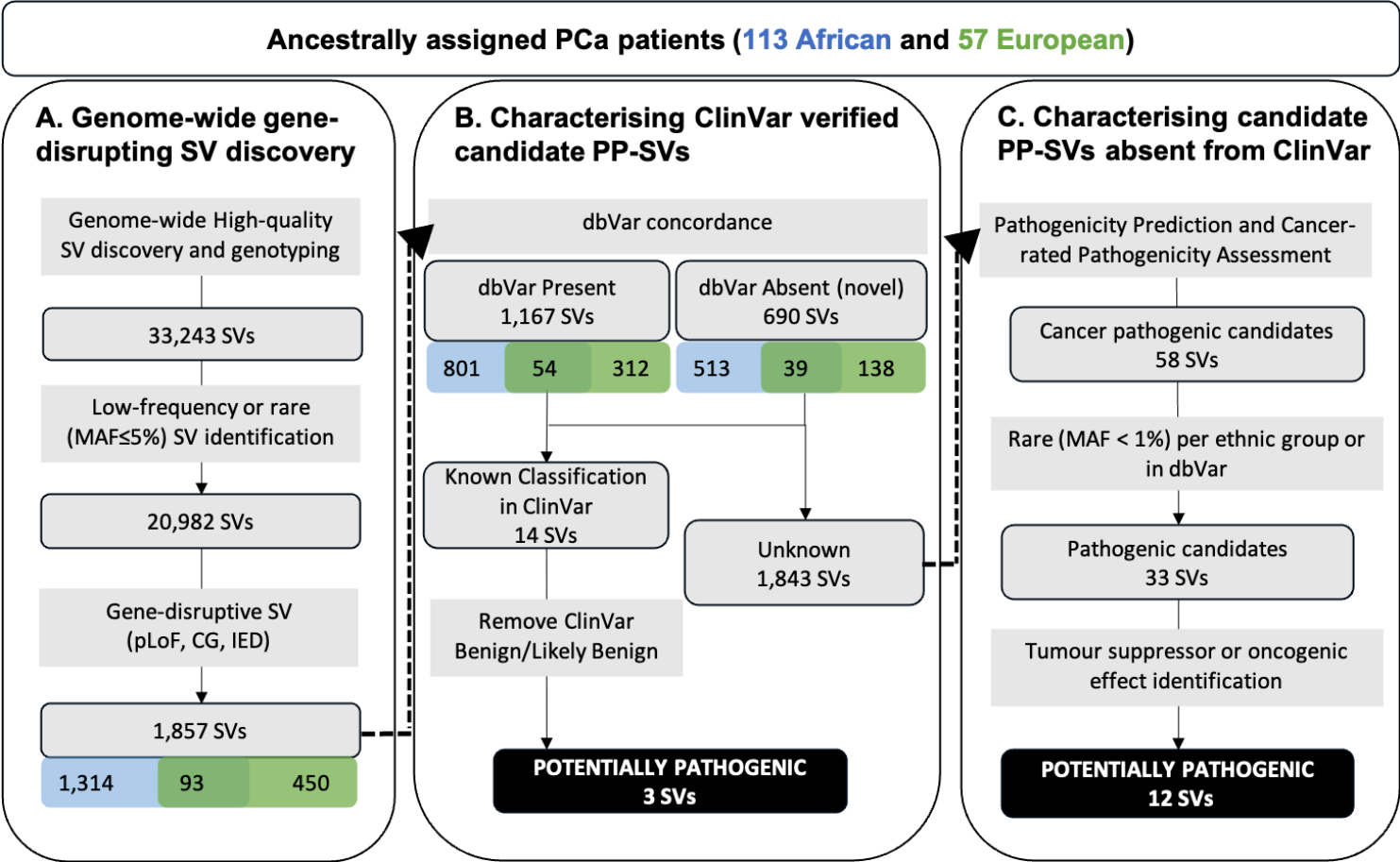


Figure 1

Workflow of PCa potentially pathogenic SV (PP-SV) identification. The detailed criteria to predict the potential pathogenicity were shown in Supplementary Table 3. The identification of tumour suppressor or oncogenic effect for disrupted genes by pathogenic candidates and the related literatures were shown in Supplementary Table 7.

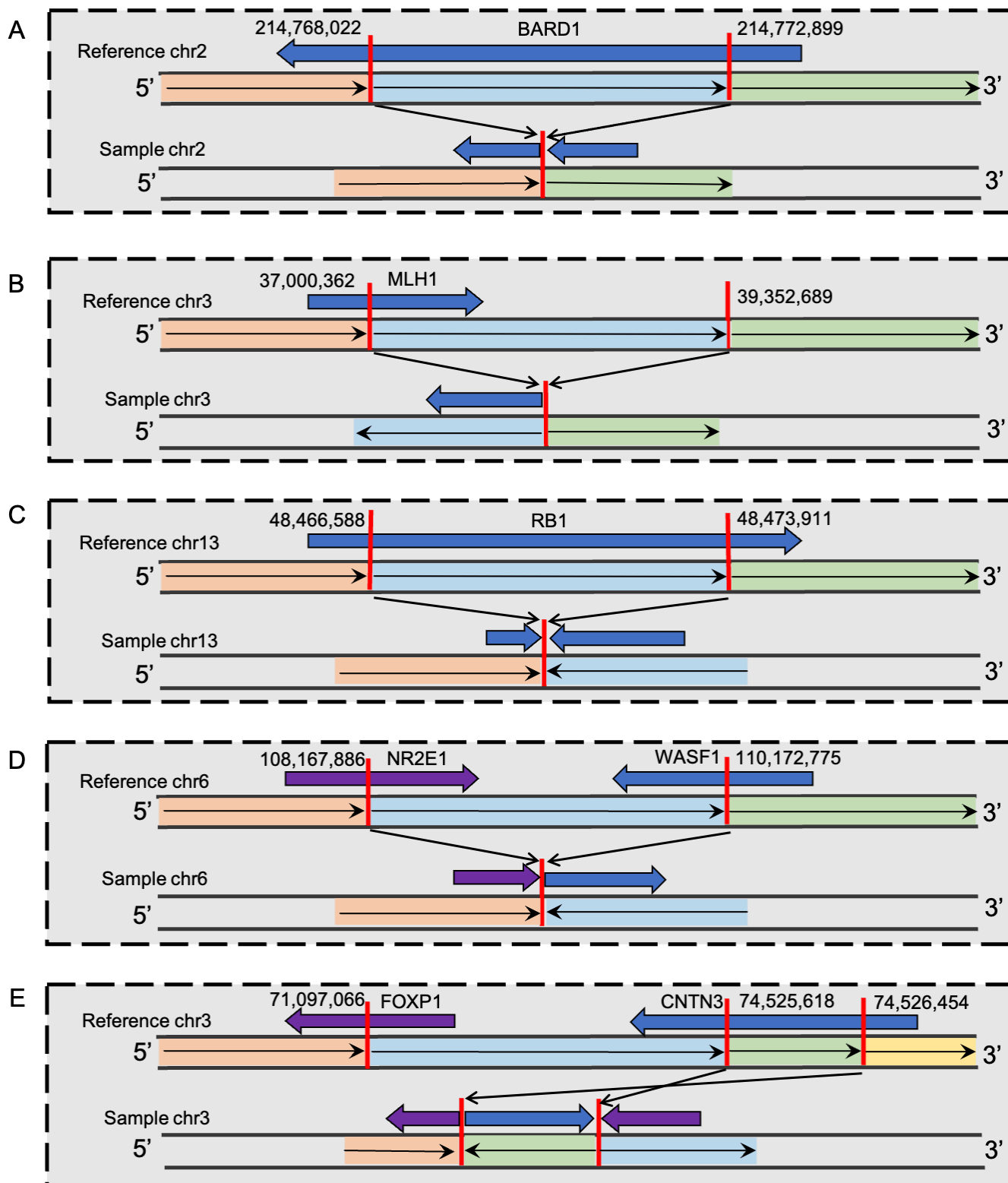


Figure 2

African-specific PP-SVs disrupting well-known pathogenic cancer genes and/or PCa tumour suppressor genes, including DNA damage response genes. (A) 4,877 base pLoF deletion on DNA damage repair gene *BARD1*. (B) pLoF INV impacting PCa DNA mismatch repair gene *MLH1*. (C) pLoF INV impacting PCa tumour suppressor *RB1*. (D) pLoF INV impacting PCa tumour suppressor gene *WASF1*. (E) pLoF INV impacting PCa tumour suppressor gene *FOXP1*. More details of SV region and/or breakpoints on impacted genes and visual inspection of sequencing reads using Integrative Genomic Viewer⁴⁹ are shown in Supplementary Fig. 8, 12, 13, 14 and 15, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarydata1.xlsx](#)
- [Supplementarydata2.xlsx](#)
- [PCaSVpathogenicNatCommsSUPP.docx](#)