

Molecular Profiles of Prostate Cancer in Men of Middle Eastern Ancestry Identifies Key Differences with Western Populations: A Multiethnic SNP Array Study

Alia Albawardi^{1,2}], Julie Livingstone^{3,4,5}], Saeeda Al Marzooqi^{1,2}, Kathleen E. Houlahan^{3,4,5,6}, Aktham Adnan Ahmad Awwad¹, Ramy A. Abdelsalam^{7,8}, Paul C. Boutros^{3,4,5,9,10}, Tarek A. Bismar^{7,11,12}

¹ Tawam Hospital, Abu Dhabi, United Arab Emirates

² Pathology College of Medicine & Health Sciences, United Arab Emirates University, United Arab Emirates

³ Departments of Human Genetics, University of California, Los Angeles, USA

⁴ Jonsson Comprehensive Cancer Centre, University of California, Los Angeles, USA

⁵ Institute for Precision Health, University of California, Los Angeles, USA

⁶ Department of Medical Biophysics, University of Toronto, Toronto, Canada

⁷ Department of Pathology and Laboratory Medicine, University of Calgary-Cumming School of Medicine and Alberta Precision Labs, Calgary, Alberta, Canada

⁸ Department of Pathology, Mansoura University, Mansoura, Egypt

⁹ Department of Pharmacology & Toxicology, University of Toronto, Toronto, Canada

¹⁰ Department of Urology, University of California, Los Angeles, USA

¹¹ Departments of Oncology, Biochemistry and Molecular Biology, University of Calgary-Cumming School of Medicine, Calgary, Alberta, Canada

¹² Arnie Charbonneau Cancer Institute and Tom Baker Cancer Center, Calgary, Alberta, Canada

] These authors contributed equally

Corresponding Author:

Tarek A. Bismar, MD

Professor, University of Calgary Cumming School of Medicine

Department of Pathology & Laboratory Medicine

Rockyview Hospital Laboratory

Alberta Precision Labs

7007 - 14th Street SW

Calgary, Alberta T2V 1P9

ph.: 403-943-8430

fax: 403-943-3333

e-mail: tabismar@ucalgary.ca

Abstract

Background: Our knowledge of prostate cancer (PCa) genomics mainly reflects European (EUR) and Asian (ASN) populations. Our understanding of the influence of Middle Eastern (ME) and African (AFR) ancestry on the mutational profiles of prostate cancer is limited.

Design, setting and participants: To characterize genomic differences between ME, EUR, ASN and AFR ancestry, fluorescent *in situ* hybridization (FISH) studies for *NKX3-1* deletions and *MYC* amplification was carried out on 42 tumors arising in individuals of ME ancestry. These were supplemented by analysis of copy number profiles of 401 tumors of all ancestries.

Outcome measurement and statistical analysis: FISH results of *NKX3-1* and *MYC* were assessed in the ME cohort and compared to other ancestries. Gene level copy number aberrations (CNAs) for each sample were statistically compared between ancestry groups.

Results and Limitation: *NKX3-1* deletions by FISH were observed in 17/42 (17.5%) prostate tumors arising in men of ME ancestry, while *MYC* amplifications were only observed in 1/42 (0.23%). Using CNAs called from arrays, the incidence of *NKX3-1* deletions was significantly lower in ME vs. other ancestries (20% vs. 52%; $p = 2.3 \times 10^{-3}$). No differences in *MYC* amplifications was noted between the two cohorts using array-based CNAs. Across the genome, tumors arising in men of ME ancestry had fewer CNAs than those of other ancestries ($p = 0.014$). Additionally, the somatic amplification of 21 specific genes was more frequent in tumors arising in men of ME vs. EUR ancestry (Proportions test; $Q < 0.05$). Those included amplifications in the glutathione S-transferase family on chromosome 1 (*GSTM1*, *GSTM2*, *GSTM5*) and the IQ Motif Containing family on chromosome 3 (*IQCF1*, *IQCF2*, *IQCF13*, *IQCF4*, *IQCF5*, *IQCF6*).

Conclusion: Significant differences in genetic background exists between different ancestries. The lower number of somatic CNAs observed in prostate cancer arising in men of ME ancestry may explain the relatively good prognosis in this population. Larger studies are warranted.

Keywords: Prostate Cancer; Middle Eastern Ancestry; Copy Number Aberrations

Introduction

Prostate cancer (PCa) is the most diagnosed non-skin malignancy in men. Risk factors include age [1], family history [2], diet, obesity and ancestry [3]. Genome-wide association studies have identified common genetic polymorphisms on chromosome 8p to be a significant influence on PCa initiation and progression in populations of European (EUR) ancestry (3a).

. Similarly, specific germline polymorphisms have been associated with prostate cancer risk in men of various ancestries [4,5].

In a similar way, several somatic genomic alterations differ in frequency between ancestries, including *TMPRSS2-ERG* genomic rearrangements, *SPINK1* overexpression and *SPOP* single nucleotide variants (SNVs), supporting differences in prostate cancer risk and pathway progression [6,7]. The rate of *PTEN* loss, as assessed by IHC, is lower in men with ME ethnicity and tumors with *PTEN* loss are not enriched for *ERG* loss, as documented in EUR cohorts (7a).

Copy number aberrations (CNAs) are the main mutational drivers of prostate cancer, occurring early in evolution [8], being highly prognostic [9], driving subtype [10] and shaping the tumor transcriptome and proteome [11]. The large majority of prostate cancer genomic data is from men with European or Asian ancestry [12] leaving other ancestry groups understudied. To begin to fill this gap, we have assembled a cohort of publicly available CNA data from 376 patients of diverse ancestries, and new data from 25 men of Middle Eastern ancestry.

Methods

Fluorescence *in situ* Hybridization Patient Cohort

A retrospective cohort of 42 radical prostatectomies and transurethral resection of the prostate, from Tawam Hospital (Al Ain- United Arab Emirates) reflecting ME men, was investigated by FISH for *NKX3-1* deletion and *MYC* amplification. All selected cases had histopathologic diagnosis of prostate adenocarcinoma. The original histopathologic classification was confirmed by one of the study pathologists (A.A, SA or TAB). Formalin-fixed, paraffin-embedded tissue sections were used for interphase fluorescence *in situ* hybridization (FISH). Deparaffinized tissue was treated with 0.2 mol/L HCl for 10 minutes, 2× SSC for 10 minutes at 80°C and digested with Proteinase K (Invitrogen) for 8 minutes. The tissues and BAC probes were co-denatured for 5 minutes at 94°C and hybridized overnight at 37°C. Post-hybridization washing was with 2× SSC with 0.1% Tween 20 for 5 minutes, and fluorescent detection was done using anti-digoxigenin conjugated to fluorescein (Roche Applied Science, Indianapolis, IN) and streptavidin conjugated to Alexa Fluor 594 (Invitrogen). Slides were counterstained and mounted in ProLong Gold Antifade Reagent with 4',6-diamidino-2-phenylindole (Invitrogen). Slides were examined using a Leica DMRA fluorescence microscope (Leica, Deerfield, IL) and imaged with a CCD camera using the CytoVision software system (Applied Imaging, Santa Clara, CA). FISH signals were scored manually (X100 oil immersion) in morphologically intact and non-overlapping nuclei by pathologists, and a minimum of 100 cancer cells from each site were recorded. Cancer sites with very weak or no signals were recorded as insufficiently hybridized. All BACs were obtained from the BACPAC Genomics (Emeryville, CA), and probe locations were verified by hybridization to metaphase spreads of normal peripheral lymphocytes. For detection of gene deletion or amplification, the following probes were used: for *NKX3-1*, RP11-325C22 (green) and for *MYC*, RP11-1136L8 (red). BAC DNA was isolated using a QIAFilter Maxi Prep kit (Qiagen, Valencia, CA), and probes were synthesized using digoxigenin- or biotin-nick translation mixes (Roche

Applied Science).

OncoScan SNP Array Patient Cohort

We assessed, 22 out of the 42 cohort samples used for FISH analysis and three additional samples, using the OncoScan SNP microarray. Representative slide(s) from each case corresponding to the formalin-fixed paraffin-embedded block(s), were identified to obtain tissue cores for OncoScan SNP microarray profiling. For molecular analysis, approximately 8-10 punches of 1.5 mm, were collected in Eppendorf tubes from regions with highest percentage of invasive carcinoma, non-tumoral, and involved regional lymph nodes if any.

Ethical approval was obtained from Al Ain Medical District Human Research Ethics Committee # 15/118 (2015-4246). This study was supported by a grant from Sheikh Hamdan Bin Rashid Al Maktoum Award for Medical Sciences (MRG/35/2017). Patient selection, tissue collection and sample processing for the 376 remaining non ME samples were performed as described in a previous publication [10].

SNP Microarray Data Generation and CNA Calling

SNP microarrays were performed with 200 ng of DNA on Affymetrix OncoScan FFPE Express 3.0 arrays as previously described [10]. The genotypes of 217,439 SNPs were extracted from the OncoScan OSCHP files and converted to VCF format. Gene level CNAs for each patient were identified by overlapping copy number segments, with RefGene (2014-07-15) annotation, using BEDTools (v2.17.0) [13]. Percent genome altered (PGA) was calculated for each sample by dividing the number of base pairs that were involved in all copy number segments by the total length of the genome.

Inferring Ancestry

The HGDP-CEPH dataset (n = 1,042) was downloaded [14], subsetted to SNPs that overlap with the OncoScan SNP array (n = 63,320), and converted to VCF format. HGDP-CEPH VCFs were merged with the OncoScan cohort using VCFtools [15]. SNPs in linkage equilibrium with each

other were pruned using the -indep command in PLINK [16]. Principal Component Analysis (PCA) was implemented using PLINK 1.9 (--pca) in the entire dataset, as well as within, the European population and Middle Eastern population separately.

The ADMIXTURE v 1.3.0 algorithm [17] was used to infer the ancestry of individuals based on the ancestry proportion given k ancestral populations with the PLINK BED file as input. Unsupervised ADMIXTURE analysis was run with k -fold cross validation with iterations of k from 2 to 8. This assignment was used to compare CNA profiles between ancestry populations.

Statistical Analysis

Specific statistical tests used are indicated in the figure legends or appropriate methods section and were performed within the R statistical environment (v3.3.1). Visualization in R was performed with the BPG package (v5.9.2)[18].

Results

By FISH, 16% (7/42) of prostate tumors arising in men of ME ancestry harbored an *NKX3-1* deletion. These results were confirmed by SNP microarray, where only 20% (5/25) of tumors from men with Middle Eastern ancestry had a deletion of *NKX3-1*. By comparison, fully 52% (186/346) of tumors arising in men of European ancestry harbored a deletion of *NKX3-1* ($p = 2.3 \times 10^{-3}$). Similarly, by FISH and ~2.5% (1/42) of tumors arising in men of ME ancestry had *MYC* amplification: this rate was higher by SNP microarray at 16% (4/25), and comparable to the 18% (65/346) rate in men of European ($p = 0.94$). **Fig. S1** shows an example of *NKX3-1* deletion detected by FISH.

To expand these intriguing results, we collected a dataset of 401 patients with sporadic, localized, treatment naive disease. Each patient in this cohort had whole genome copy-number profiling of the index lesion of their tumor. This cohort included 25 newly generated tumors from men of Middle Eastern Ancestry (**Table 1**; **Table S1**). Patients either underwent image-guided radiotherapy (IGRT) or surgery (radical prostatectomy), and the histologically most representative region was molecularly profiled. There was no difference in cellularity or other metrics as a function of processing batch (two-way ANOVA; $P = 0.15$; **Fig. S2**), as measured by ASCAT [19].

First, we inferred the genetic ancestry of this joint dataset. We integrated our data with the Human Genome Diversity Panel (HGDP-CEPH), which contains samples from 51 different global populations [14]. To determine the ancestry fraction for each individual in the present cohort, we used the software ADMIXTURE [17] (**Fig. 1A**; **Fig. S3**). HGDP-CEPH samples previously assigned an ancestry group were not re-assigned, independent of the admixture results. Four primary admixture populations were apparent (Africa - Red; America - Yellow, East Asia - Green and Oceania - Aquamarine). The other populations were a combination of these primary populations, including samples from men with Middle Eastern ancestry, who showed a combination of African, Central South Asian and European genotypic features.

Three Middle Eastern samples in our dataset contained a larger than average proportion of the African primary admixture component (**Fig. S4**). We performed principal component analysis (PCA) using 63,320 SNPs to investigate genetic diversity (**Fig. 1B**). PC1 and PC2 explained 30% and 27% of genetic variance, respectively, with PC1 distinguishing East from West populations while PC2 divided African/Middle Eastern populations from non-African populations. PCA of specific geographical regions showed sub-structures within the larger populations, including within specific regions of Europe and the Middle East (**Fig 1C-D**). As previously reported [20,21], Middle Eastern and European populations have divergent genetic ancestry.

To explore differences in somatic CNA profiles between different ancestry groups, we compared the copy number profiles of tumors from men with Middle Eastern ancestry (MEA; $n = 25$), African ancestry (AFR; $n = 14$), East Asian ancestry (EAS; $n = 11$) and European ancestry (EUR; $n = 346$; **Fig. 2A; Table S2**). Tumors identified as arising in men of Native American ancestry ($n = 5$) were removed due to the low sample-size. All ancestry groups showed large inter-tumoral heterogeneity: with tumors from men with European ancestry had between 0 - 284 CNA, tumors from men with African ancestry 19 - 217 CNAs, tumors from men with East Asian ancestry 4 - 209 CNAs and tumors from men with Middle Eastern ancestry 1 - 166 CNAs.

The median tumor arising in a man of European ancestry contained 18 amplifications and 29 deletions. Tumors arising in men of East Asian ancestry had similar numbers of CNAs (20 amplifications + 33 deletions; $p = 0.16$; Mann-Whitney test). By contrast, tumors arising in men of Middle Eastern ancestry had fewer CNAs (7 amplifications + 11 deletions; $p = 0.014$; Mann-Whitney test) while those arising in men of African ancestry had the most CNAs (45 amplifications + 27 deletions; $p = 0.016$; Mann-Whitney test). As expected, the smaller number of CNAs in tumors from men with Middle Eastern ancestry led to lower global genomic instability, measured as the percentage of genome altered (PGA; **Fig 2B**). Tumors from men with Middle Eastern ancestry had lower PGA (median = 1.93%) than those arising in men of European ancestry

(median = 6.45%; $p = 1.07 \times 10^{-2}$; Mann-Whitney test). There was no difference in PGA between tumors from men with African ancestry despite the larger number of aberrations (median = 9.72%) or tumors from men with East Asian ancestry (median = 6.83%).

We assigned each tumor to prostate cancer subtype based on their gene level CNA profile (**Fig. 2A & C**) [9]. These subtypes (S1-S4) were defined by specific genomic aberrations (S1 - chromosome 7 amplification, S2 - 8p deletion, 8q amplification, S3 - 8p deletion, 16p deletion, S4 - quiet profile). All genetic ancestry groups had a large proportion of tumors in S4, indicative of quiet CNA profiles (**Fig 2A**). Despite the differences in global genomic instability, the proportion of tumor subtypes did not differ between the different genetic ancestry groups, at least at the statistical power afforded by existing cohorts (Pearson's X^2 test; $p = 0.16$; **Fig. 2C**).

Understanding the interplay between genetic ancestry and CNAs is important for population specific biomarker identification. To this end, we investigated recurrent gene level CNAs in tumors between men with Middle Eastern ancestry, African ancestry, and East Asian ancestry to tumors from men with European ancestry (**Table S2**). We overlapped CNA segments with gene annotation and compared the proportion of amplifications and deletions separately between the two populations. Overall, when investigated globally, there were no differences in the proportion of amplifications and deletions in known prostate cancer driver genes across ancestry populations at this statistical power, beyond the FISH-discovered *NKX3-1* result above (Proportion test; $Q > 0.05$; **Fig. S5; Table S3**).

Globally, 21 genes had more frequent amplifications in tumors from men with Middle Eastern ancestry compared to tumors from men with European ancestry (Proportions test; $Q < 0.05$; **Fig. 2D**). All of these amplifications occurred at a higher rate in tumors from men with Middle Eastern ancestry. These included amplifications in the glutathione S-transferase family of genes on chromosome 1 (*GSTM1*, *GSTM2*, *GSTM5*) and the IQ Motif Containing family of genes on chromosome 3 (*IQCF1*, *IQCF2*, *IQCF13*, *IQCF4*, *IQCF5*, *IQCF6*). Interestingly, *PARP3*, an

enzyme which is required for DNA repair and maintenance of genomic instability, was amplified in a higher proportion of tumors from men with Middle Eastern ancestry. We did not observe any differences in the proportion of deletions between these two populations (Proportion test; $Q > 0.05$).

There were 34 genes in which the proportion of tumors with an amplification and 40 genes in which the proportion of deletions differed between tumors from men with African ancestry compared to tumors from men with European ancestry (Proportions test; $Q < 0.05$; **Fig. 2D**). These include amplifications in the Keratin Associated Protein 10 family of genes on chromosome 21 (*KRTAP10-1* - *KRTAP10-11*). Intriguingly, there was a higher proportion of *BRCA2* amplifications in tumors from men with African ancestry (21%; 3/14) than in tumors from men with European ancestry (1.1%; 4/346). This may account for the larger number of CNAs in this cohort of samples.

There were 148 genes in which the proportion of tumors with an amplification differed between men with East Asian ancestry compared to tumors from men with European ancestry (Proportions test; $Q < 0.05$; **Fig. 2D**). This includes a region on chromosome 7 spanning 26 genes that is amplified in 88% (9/11) of tumors of East Asian ancestry in our cohort. These data provide us with evidence that genes are mutated at different frequencies in individuals of different ancestry, providing further strong support to the idea that germline variation is essential for understanding the emergence of somatic phenotypes [22].

Discussion

Increasingly evidence now shows that germline genetic variation strongly shapes the somatic profiles and evolutionary history of prostate cancer. Both rare deleterious variants in DNA damage response (DDR) genes and common polymorphisms have been shown to do so in Caucasian populations [2,22]. Recent sequencing studies have shown that tumors arising in men of Asian, African or African-American ancestry show distinct somatic mutational features [23–26]. This study expands on these observations, looking at a multi-ancestric cohort and provides the first analysis of CNAs in men of Middle Eastern (ME) descent.

Tumors in men of ME descent showed less genomic instability than those of other groups, but harbored amplifications in the glutathione S-transferase family of genes on chromosome 1 (*GSTM1*, *GSTM2*, *GSTM5*), the IQ Motif Containing family of genes on chromosome 3 (*IQCF1*, *IQCF2*, *IQCF13*, *IQCF4*, *IQCF5*, *IQCF6*) and *PARP3*, an enzyme which is required for DNA repair not commonly mutated in tumors arising in men of other ancestries. *GSTM1* amplifications have been reported to be marginally associated with prostate cancer risk in a Caribbean population of African descent [27] warranting further investigation into the function of this gene in non-Caucasian cohorts. All of these variants are putatively somatic, but it is possible our genomic analysis included contaminating germline structural variants, which would be of significant interest. Interestingly, all three samples from men of Middle Eastern ancestry that had a high proportion of African admixture components had an amplification in *GSTM1*.

- Although, we did not observe significant differences between the two ethnicities in the proportion of amplifications and deletions in known genes associated with prostate cancer, we identified significant difference in the *NKX3-1* genomic deletions. This is in line with previous work confirming differences in *ERG*, *PTEN* and *SPINK1* genomic aberrations between the two ancestries (7a). These data confirms that the genomics of prostate cancer in the ME population are different than those in ASN, AFR and EUR populations, likely reflective of factors leading to

the higher incidence, rate of disease progression and patients' cause specific mortality observed in ASN, AFR and EUR compared to ME ethnicity.

In conclusion, ancestry should be considered when investigating and characterizing biomarkers and molecular signature relative to disease progression, prognosis and potentially therapeutic targeting. Despite our very limited sample-size, they suggest germline-somatic interactions may be at play in tumors arising in men of different ancestries. Further multi-omic studies of tumors arising in men of different ancestry are clearly warranted.

Funding

This study was conducted with the support to PCB by the NIH/NCI under award number P30CA016042 and by an operating grant from the National Cancer Institute Early Detection Research Network (1U01CA214194-01). This work was supported by Alberta Precision Labs (RS19-600), Prostate Cancer Canada and Prostate Cancer Foundation USA. The Sheikh Hamdan Medical Award supported partially or completely by grant # MRG/35/2017 (AA, SM, TAB).

Figure Legends

Figure 1 | Population analysis with HGDP-CEPH SNP data

- a) ADMIXTURE cluster plot with $k = 7$. Each sample is represented by a vertical line partitioned into coloured segments whose length is proportional to the ancestry coefficient in up to seven inferred populations. The cohort covariate bar represents which dataset the sample came from where black indicates the current cohort and white indicates HGDP-CEPH.
- b) PCA plot with 63,320 SNPs analyzing the same population as in **Fig. 1A**. Samples are coloured based on their assigned genetic ancestry, with circles indicating samples from the HGDP-CEPH dataset and triangles represent samples in the current cohort.
- c) PCA plot including samples with European genetic ancestry. Samples are coloured based on their geographic region within Europe, or grey if region is unknown. Circles indicate samples from the HGDP-CEPH dataset. Grey triangles represent samples with European ancestry in the current cohort; specific region is unknown.
- d) PCA plot including samples with Middle Eastern ancestry. Samples are coloured based on their geographic region within the Middle East, or grey if region is unknown. Circles indicate samples from the HGDP-CEPH dataset. Grey triangles represent the Middle Eastern samples in the current cohort; the specific region is unknown.

Figure 2 | Copy number aberration profiles differ based on ancestry

- a) CNA profiles of tumors from men with Middle Eastern, African, East Asian and European ancestry. Within the heatmap, red indicates a copy number amplification while blue indicates a copy number deletion. Each row represents a tumor (ordered by subtype), and each column represents a gene ordered by genomic coordinate. Clinical features per tumor are shown in the covariate bar on the right, including the four previously identified CNA subtypes. IDC; intraductal carcinoma or cribriform architecture.

- b) Percent genome altered (PGA) differed across tumors from men with different ancestries. P value is from a Mann-Whitney test.
- c) The proportion of tumors assigned to the [9] subtypes did not differ between tumor from men with different ancestry. P value is from a chi-squared test. The colour of the stacked bars correspond to the CNA subtypes in **Fig. 1A**.
- d) Specific genes differed in their proportion between tumors from men with Middle Eastern, African, East Asian and European ancestry. Genes are ordered by genomic coordinate per chromosome. Amplifications are represented in red and deletions are represented in blue. Only significant Q values from proportion tests are shown comparing men with Middle Eastern (dark orange), African (red), or East Asian (green) ancestry to men with European ancestry.

Supplementary Figure Legends

Figure S1 | Example image of NKX3-1 deletion identified through FISH

FISH image demonstrating hemizygous *NKX3-1* deletions (green) in prostate cancer sample of ME ancestry. MYC probe is shown as red signal. Magnification x 100.

Figure S2 | Tumor cellularity does not differ across ancestry groups

P value is from a two-way ANOVA.

Figure S3 | ADMIXTURE cluster plots varying inferred ancestries from 2 to 8

Each sample is represented by a vertical line partitioned into coloured segments whose length is proportional to the ancestry coefficient in up to eight inferred populations.

Figure S4 | ADMIXTURE cluster plots for men with Middle Eastern ancestry

Three samples with a larger than average proportion of the Africa primary admixture component are labelled in red.

Figure S5 | CNA profile of prostate cancer associated genes

Each column is a sample and each row is a gene. The color represents whether that sample had a deletion (blue), amplification (red), or is copy number neutral (white). The covariate bar on top indicates the ethnicity of the patient.

Supplementary Tables

Table S1 | Patient clinical information

For each patient, this table provides clinical information including ancestry, age at treatment, pre-treatment PSA, T category, diagnostic Gleason Grade and the presence of intraductal carcinoma (IDC) or cribriform architecture.

Table S2 | Patient by CNA matrix

Gene level CNA data per sample. First five columns contain hg19 gene annotation information (RefGene ID, gene symbol, chromosome, start position, end position). CNA states are as follows: '0' indicates neutral, '1' indicates copy number amplification and '-1' indicates copy number deletion.

Table S3 | CNA analysis statistics

For each gene, this table provides the statistics comparing ancestry groups (African vs. European, East Asian vs. European, Middle Eastern vs. European) using a two-sided proportions test. Amplifications and deletions were compared separately. Tests were performed for each gene in which there were aberrations in at least 5% of samples. NAs indicate that there were not enough samples in that group to perform the statistical test.

References

- 1 Weischenfeldt J, Simon R, Feuerbach L, et al. Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. *Cancer Cell* 2013;23:159–170.
- 2 Taylor RA, Fraser M, Livingstone J, et al. Germline BRCA2 mutations drive prostate cancers with distinct evolutionary trajectories. *Nat Commun* 2017;8:13671.
- 3 SEER Cancer Statistics Review, 1975-2016 [Internet]. SEER. [cited 24 October 2019]. Available from: https://seer.cancer.gov/csr/1975_2016/index.html.
- 3a Washburn JG, Wojno KJ, Dey J, Powell IJ, Macoska JA: **8pter-p23 deletion is associated with racial differences in prostate cancer outcome.** *Clin Cancer Res* 2000, **6**(12):4647-4652.

- 4 Shan J, Al-Rumaihi K, Rabah D, et al. Genome scan study of prostate cancer in Arabs: identification of three genomic regions with multiple prostate cancer susceptibility loci in Tunisians. *J Transl Med* 2013;11:121.
- 5 Al Olama AA, Kote-Jarai Z, Berndt SI, et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet* 2014;46:1103–1109.
- 6 Kong D-P, Chen R, Zhang C-L, et al. Prevalence and clinical application of TMPRSS2-ERG fusion in Asian prostate cancer patients: a large-sample study in Chinese people and a systematic review. *Asian J Androl* 2019. doi:10.4103/aja.aja_45_19.
- 7 Tonon L, Fromont G, Boyault S, et al. Mutational Profile of Aggressive, Localised Prostate Cancer from African Caribbean Men Versus European Ancestry Men. *Eur Urol* 2019;75:11–15.
- 7a Abdelsalam RA, Khalifeh I, Box A, Kalantarian M, Ghosh S, Abou-Ouf H, Lotfi T, Shahait M, Palanisamy N, Bismar TA: **Molecular characterization of prostate cancer in Middle Eastern population highlights differences with Western populations with prognostic implication.** *J Cancer Res Clin Oncol* 2020, **146**(7):1701-1709
- 8 Espiritu SMG, Liu LY, Rubanova Y, et al. The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. *Cell* 2018;173:1003-1013.e15.
- 9 Lalonde E, Ishkanian AS, Sykes J, et al. Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. *The Lancet Oncology* 2014;15:1521–1532.
- 10 Fraser M, Sabelnykova VY, Yamaguchi TN, et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* 2017;541:359–364.
- 11 Sinha A, Huang V, Livingstone J, et al. The Proteogenomic Landscape of Curable Prostate Cancer. *Cancer Cell* 2019;35:414-427.e6.
- 12 Rebbeck TR. Prostate Cancer Genetics: Variation by Race, Ethnicity, and Geography. *Semin Radiat Oncol* 2017;27:3–10.
- 13 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–842.
- 14 Cann HM, de Toma C, Cazes L, et al. A human genome diversity cell line panel. *Science* 2002;296:261–262.
- 15 Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics* 2011;27:2156–2158.
- 16 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;4:7.
- 17 Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009;19:1655–1664.

- 18P'ng C, Green J, Chong LC, et al. BPG: Seamless, automated and interactive visualization of scientific data. *BMC Bioinformatics* 2019;20:42.
- 19Van Loo P, Nilsen G, Nordgard SH, et al. Analyzing cancer samples with SNP arrays. *Methods Mol Biol* 2012;802:57–72.
- 20Phillips C, Parson W, Lundsberg B, et al. Building a forensic ancestry panel from the ground up: The EUROFORGEN Global AIM-SNP set. *Forensic Sci Int Genet* 2014;11:13–25.
- 21Li JZ, Absher DM, Tang H, et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 2008;319:1100–1104.
- 22Houlahan KE, Shiah Y-J, Gusev A, et al. Genome-wide germline correlates of the epigenetic landscape of prostate cancer. *Nat Med* 2019;25:1615–1626.
- 23Ren S, Wei G-H, Liu D, et al. Whole-genome and Transcriptome Sequencing of Prostate Cancer Identify New Genetic Alterations Driving Disease Progression. *Eur Urol* 2018;73:322–339.
- 24Huang FW, Mosquera JM, Garofalo A, et al. Exome Sequencing of African-American Prostate Cancer Reveals Loss-of-Function ERF Mutations. *Cancer Discov* 2017;7:973–983.
- 25Jaratlerdsiri W, Chan EKF, Gong T, et al. Whole-Genome Sequencing Reveals Elevated Tumor Mutational Burden and Initiating Driver Mutations in African Men with Treatment-Naïve, High-Risk Prostate Cancer. *Cancer Res* 2018;78:6736–6746.
- 26Takata R, Takahashi A, Fujita M, et al. 12 new susceptibility loci for prostate cancer identified by genome-wide association study in Japanese population. *Nat Commun* 2019;10:4422.
- 27Emeville E, Broquère C, Brureau L, et al. Copy number variation of GSTT1 and GSTM1 and the risk of prostate cancer in a Caribbean population of African descent. *PLoS ONE* 2014;9:e107275.