BMC Cancer

PRState: Incorporating genetic ancestry in prostate cancer risk scores for men of African ancestry

Meghana S. Pagadala[1,2,3*†] , Joshua A. Linscott[4†], James V. Talwar[5], Tyler M. Seibert[6,7,8,9], Brent Rose[6,7,10], Julie Lynch[11,12], Matthew Panizzon[7,13], Richard Hauger[7,13,14], Moritz H. Hansen[4], Jesse D. Sammon[4], Matthew H. Hayn[4], Karim Kader[10], Hannah Carter[1] and Stephen T. Ryan[4]

Abstract

Background: Prostate cancer (PrCa) is one of the most genetically driven solid cancers with heritability estimates as high as 57%. Men of African ancestry are at an increased risk of PrCa; however, current polygenic risk score (PRS) models are based on European ancestry groups and may not be broadly applicable. The objective of this study was to construct an African ancestry-specific PrCa PRS (PRState) and evaluate its performance.

Methods: African ancestry group of 4,533 individuals in ELLIPSE consortium was used for discovery of African ancestry-specific PrCa SNPs. PRState was constructed as weighted sum of genotypes and effect sizes from genome-wide association study (GWAS) of PrCa in African ancestry group. Performance was evaluated using ROC-AUC analysis.

Results: We identified African ancestry-specific PrCa risk loci on chromosomes 3, 8, and 11 and constructed a polygenic risk score (PRS) from 10 African ancestry-specific PrCa risk SNPs, achieving an AUC of 0.61 [0.60–0.63] and 0.65 [0.64–0.67], when combined with age and family history. Performance dropped significantly when using ancestry-mismatched PRS models but remained comparable when using trans-ancestry models. Importantly, we validated the PRState score in the Million Veteran Program (MVP), demonstrating improved prediction of PrCa and metastatic PrCa in individuals of African ancestry.

Conclusions: African ancestry-specific PRState improves PrCa prediction in African ancestry groups in ELLIPSE consortium and MVP. This study underscores the need for inclusion of individuals of African ancestry in gene variant discovery to optimize PRSs and identifies African ancestry-specific variants for use in future studies.

Keywords: Prostate Cancer, Prostate Cancer Risk, Single nucleotide polymorphism (SNP), Ancestry, African, Polygenic Risk Score (PRS)

†Meghana S. Pagadala and Joshua A. Linscott contributed equally to this work.

*Correspondence: mpagadal@health.ucsd.edu

[1] Department of Medicine, Division of Medical Genetics, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA
Full list of author information is available at the end of the article

Introduction

Prostate cancer (PrCa) remains the most common non-skin malignancy in men, with significant mortality resulting in 1 in 42 men diagnosed with PrCa dying from the disease [1, 2]. Men with an African ancestry have a 1.6- and 2.4-fold increased risk of PrCa diagnosis and age-matched mortality compared to men with a European ancestry [3, 4]. Multiple studies suggest genetic

Pagadala *et al. BMC Cancer* (2022) 22:1289

Page 2 of 11

heritability is high for PrCa [5, 6], with twin studies attributing 57% of PrCa risk to genetic factors [7].

While rare high penetrance genes and missense mutations (e.g., G84E in HOXB13) have been described, they represent an exceedingly small minority of PrCa cases. Single nucleotide polymorphisms (SNPs) in non-coding regions also contribute to PrCa risk, with many falling in the chromosome 8q24 risk region [8, 9]. Genome-wide association studies (GWAS) have identified more than 260 of these SNP susceptibility loci [10, 11]. However, the majority of discovery populations in these GWAS have been of European or Asian ancestry and studies on the role of ancestral genetic background in PrCa risk for other ethnic groups are needed [8, 10, 12, 13].

Incorporating PrCa risk SNPs into a meaningful clinical tool is possible with polygenic risk scores (PRSs) that predict PrCa risk based on the presence of individual inherited SNPs [14–16]. Sun et al. demonstrated that addition of a PRS to family history improved the performance of predicting PrCa in populations of predominantly European ancestry [11, 16–24]. The full utility of such tools for diverse populations or in combination with nomograms is yet to be realized.

Here we use genetic ancestry to separate the ELLIPSE consortium into ancestry groups. We ran association studies within our African ancestry group to identify population-specific SNPs. We then constructed an African ancestry-specific PrCa PRS (PRState) that achieved an AUC of 0.65 [0.64–0.67] when combined with family history and age of diagnosis. Efficacy of PRS was contingent on inclusion of African ancestry group individuals in PRS construction. When only European ancestry group individuals were used to construct PRS, a considerable drop in performance was observed within the African ancestry group. PRS construction from trans-ancestry groups performed comparably to PRState.

Variants in PRState score have been described in previous trans-ancestry analysis; however, we demonstrate these variants contribute to a boost in PrCa prediction performance in both the ELLIPSE and Million Veteran Program African ancestry groups. These findings highlight the importance of ancestry-specific risk SNP identification and will hopefully guide future PRS studies of PrCa in African ancestry groups [25].

Methods
ELLIPSE study subjects and genotype
The Elucidating Loci Involved in Prostate Cancer Susceptibility (ELLIPSE) consortium prostate cancer meta-analysis and genotypes (dbGaP Study Accession: phs001120.v1.p1) was accessed to analyze Affymetrix genotype calls for 91,644 male PrCa case/controls.

Quality assurance
PLINK (RRID:SCR_001757) genotype files consisting of 505,219 calls from the following consent groups were compiled: c1-c3,c6,c8,c10-18,c20,c23,c25,c27-28. Pre-imputation processing of autosomal and X chromosome genotypes followed below steps:

1. Duplicated variants were removed.
2. Heterozygous haploid SNPs were set to missing.
3. SNPs with call rate < 90% were removed.
4. SNPs with minor allele frequency (MAF) < 1% were removed.
5. Individuals with genotype coverage < 90% were removed.
6. Non-ACGT variants were removed.

Strand flips were reversed using snpflip. After preprocessing genotypes, the remaining 410,116 SNPs and 91,644 individuals were input to the secure Michigan Imputation Server (RRID:SCR_017579) [26]. Whole-genome SNPs were imputed with Minimac4 (RRID:SCR_009292) and ancestry-matched reference panel 1000 Genomes Project Phase 3 version 5 (RRID:SCR_008801). Finally, post-imputation duplicated SNPs and SNPs with MAF < 1% were removed.

Ancestry likelihood calculation (FROG-kb)
For ancestry group calculations, we opted for an ancestry group prediction tool that does not require relationships with other individuals, like principal component analysis (PCA). FROG-kb [27] uses Kidd AISNP panel (55 SNPs) to predict likelihood ratios for world geographic regions. Likelihood ratios for 160 populations were calculated and averaged. European ancestry likelihood ratios were determined from populations in "Europe" region and African ancestry likelihood ratios were determined from populations in the "African" region. For the final European ancestry group, we used a European log likelihood > -10, resulting in 5567 individuals. For the final African ancestry group, we used a European log likelihood < -15 and African log likelihood > -15, resulting in 4533 individuals.

Genome-wide association analyses (GWAS)
PLINK (RRID:SCR_001757) GLM method [28] was used to conduct association analyses with PrCa case/control in European and African ancestry groups. All associations were adjusted for the first 10 principal components (PCA with 55-SNP Kidd panel) and age.

Polygenic risk score calculation
Association analyses within European, African or mixed ancestry training sets were conducted. Significant variants were identified through PLINK (RRID:SCR_001757)

Pagadala *et al. BMC Cancer*   (2022) 22:1289

Page 3 of 11

linkage-based clumping using a p1 threshold of 5e-08, a p2 threshold of 1e-05, an r2 threshold of 0.1 and a kb threshold of 1000 kb. Ten, seven and fourteen significant variants were identified through African, European and trans-ancestry analysis, respectively. For PRS construction, variants were weighted by log (base 10) odds ratio from the training set association statistics, oriented to PrCa risk allele, and combined. ROC-AUC evaluation across folds was conducted using polygenic scores as predictions.

For mismatched ancestry group analysis, European ancestry training sets were used for prediction on African ancestry test sets. For trans-ancestry group analysis, European and African ancestry training sets were combined and tested on African ancestry test sets. All three ancestry group PRSs were evaluated using tenfold cross validation. AUC for each fold and overall are reported. Confidence intervals were calculated using pROC R package.

For contextualization of our results in relation to other genetic risk models, we compared our PRState score (composed of 10 African ancestry-specific variants) to variants published recently in a large meta-analysis of prostate cancer by Conti et al [11]. Half (5) of the variants used in the PRState score were in high linkage disequilibrium (r2 > 0.3) with Conti et al. variants [11]. We excluded these variants and constructed a Conti PRS with reported odds ratio for the African group. To evaluate PRState and Conti polygenic risk scores, we used both scores as features for a logistic regression model with default parameters. Predicted probabilities were used in ROC evaluation. For the Million Veteran Program (MVP), genotype dosages were extracted for 10 PRState variants across participants and weighted by log (base 10) odds ratio from the best-performing fold in ELLIPSE. Performance was evaluated using ROC-AUC analysis in European and African ancestry groups.

Million veteran program study subjects and genotype

Individual ancestry groups in the Million Veteran Program were characterized through Harmonized Ancestry and Race/Ethnicity (HARE) grouping [29]. HARE grouping was specifically developed to categorize MVP individuals based on self-reported ancestry and genetic ancestry. HARE utilizes a support vector machine to output probabilities of an individual's ancestry group using self-identified and genetic ancestry. PrCa, metastatic PrCa and fatal PrCa status was determined through ICD 9/10 diagnosis, procedure code, CPT and HCPCS procedure code, laboratory values, medications and clinical notes from inpatient, outpatient and fee-based care in the VA healthcare system. Family history information was available for only 55,610 of 121,964 African individuals and 322,706

of 461,627 European individuals in MVP. Education and income variables were available for 412,174 individuals and were used to stratify individuals according to socioeconomic status as follows: high socioeconomic status was defined as income > $50,000 and at least a bachelor's degree education level and low socioeconomic status was defined as income < $50,000 and did not obtain bachelor's degree education level). When evaluating genetic information only, the full population was used and the subset of the population with family history information available was used for additional multivariable association. To evaluate the role of socioeconomic factors in prostate cancer prediction, we conducted logistic regression analysis with PRState, family history, age, education level, income levels, and top 10 principal components. To evaluate predictive value, we conducted PRState ROC-AUC analysis separately in European and African ancestry individuals. For socioeconomic analysis, we conducted PRState ROC-AUC analysis separately in African ancestry men of high socioeconomic status and low socioeconomic status.
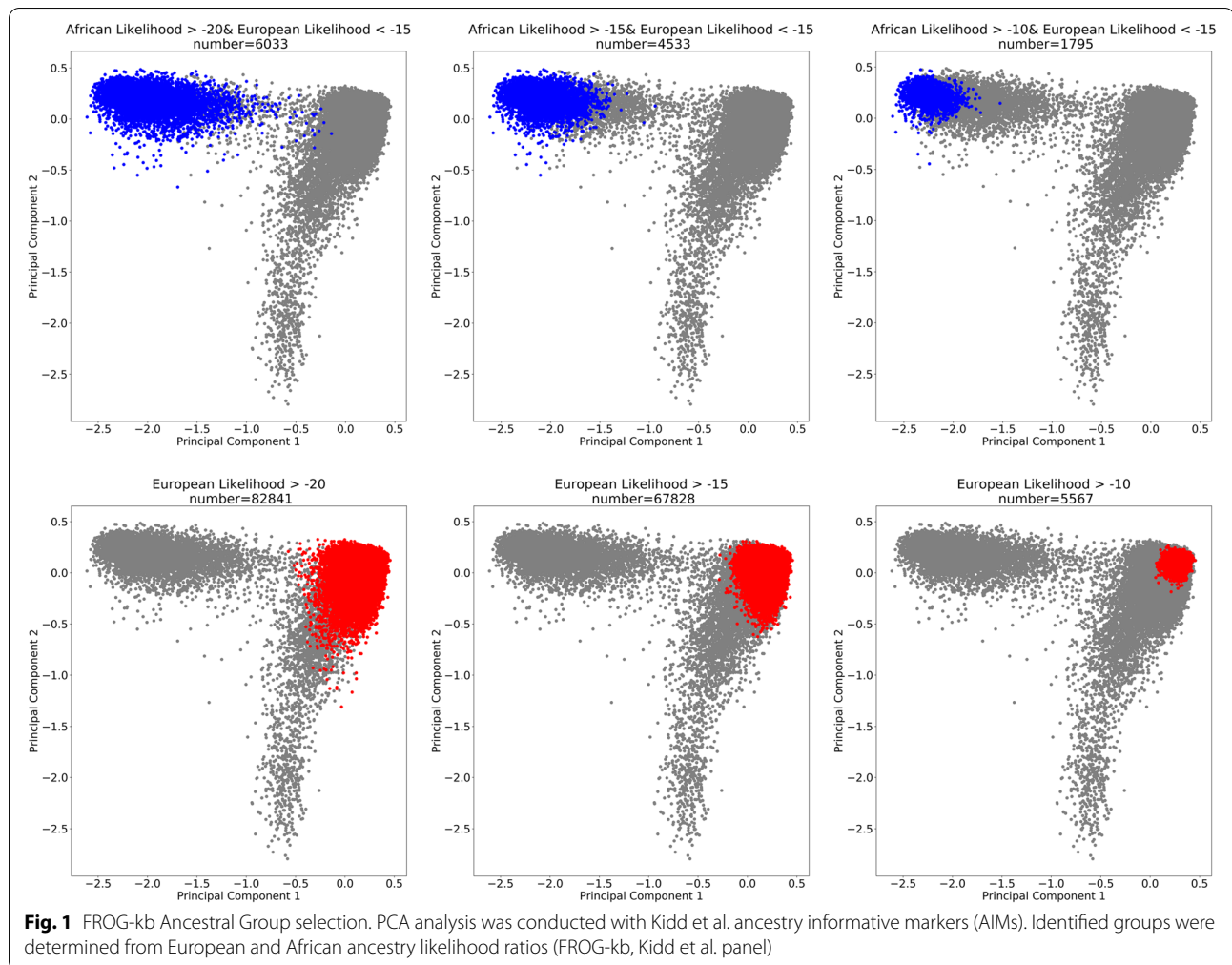
Results

European and African ancestry group identification in ELLIPSE

Principal component analysis (PCA) of 55 ancestry informative markers (AIMs) proposed by Kidd et al. [30] revealed that individuals in the ELLIPSE could be stratified according to ancestral background. European, African and Asian descent individuals formed distinct clusters (Fig. 1). European and African ancestry likelihood thresholds were selected such that population size was maximized while minimizing admixture. The European ancestry group included individuals with European ancestry likelihood ratio > -10, resulting in 5,567 individuals. The African ancestry group included individuals with African ancestry likelihood ratio > -15 and European ancestry Likelihood ratio < -15, resulting in 4,533 individuals (Fig. 1). PCA analysis confirmed thresholds results in tightly clustered European and African ancestry groups. Self-identified ancestry aligned with genetic ancestry defined through AIMs; although some individuals self-identifying as Hispanic were included in European and African ancestry groups (Table 1).

Inclusion of African ancestry group individuals in polygenic risk score construction improves african ancestry group prostate cancer prediction

GWAS was performed in the African ancestry group to identify African ancestry-specific risk loci. Loci on chromosomes 3, 8, and 11 were significantly associated with PrCa risk in men of African ancestry (Fig. 2, Supplementary Table 1). Age, family history, genetic risk

Pagadala *et al. BMC Cancer*   (2022) 22:1289

Page 4 of 11

Fig. 1 FROG-kb Ancestral Group selection. PCA analysis was conducted with Kidd et al. ancestry informative markers (AIMs). Identified groups were determined from European and African ancestry likelihood ratios (FROG-kb, Kidd et al. panel)
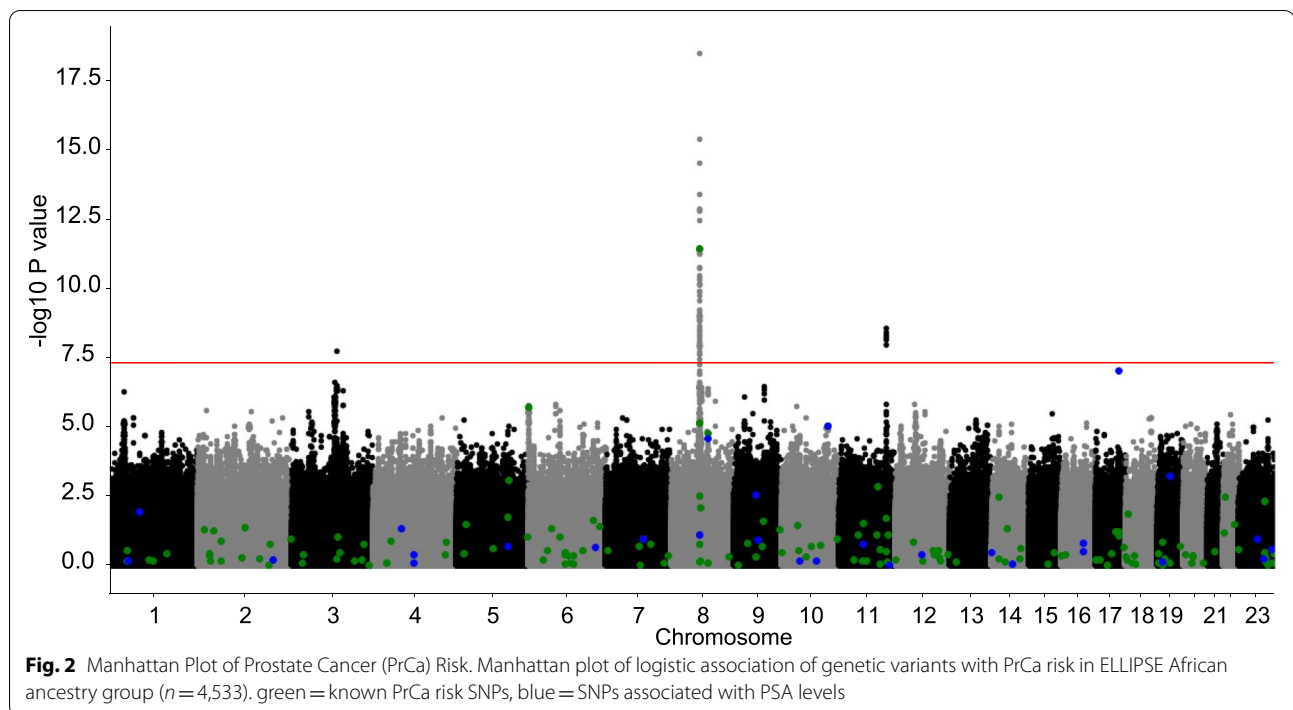
Table 1 Self-identified ancestry of genetically-defined ancestry groups. rows represent ancestry groups based on genetic ancestry. columns represent self-identified ancestry. Values represent the number of individuals in each ancestry group and how they self-identify

	European	African	Latino
European	**5564**	**0**	**3**
African	**0**	**4532**	**1**

score and the combination of these 3 factors for PrCa were evaluated using ROC-AUC analysis. Risk prediction models using genetics resulted in an AUC of 0.61 [0.60–0.63] (Figure S1A) while a combined model (genetics, age and family history) resulted in an AUC of 0.65 [0.64–0.67] (Fig. 3A). To determine the efficacy of using a matched ancestry group model for prediction, we evaluated performance of a European ancestry group model on prediction of PrCa in our African ancestry group

(Supplementary Table 2). Interestingly, model performance was poor with near random performance using a European PRS for prediction in men of African ancestry [AUC: 0.52 (0.50–0.53)] (Figure S1B). When genetics were combined with age and family history, the AUC was 0.59 [0.57–0.60], significantly lower than results from using the matched ancestry group model (Fig. 3B).

Lastly, recent studies of trans-ancestral analysis of PrCa risk have been conducted [11], so we wanted to evaluate the performance of using a trans-ancestry group model on our African ancestry group. Models were trained on a set composed of both European and African ancestry groups combined and then tested on the African ancestry group (Supplementary Table 3). We achieved comparable performance to matched ancestry group models. We achieved an AUC of 0.62 [0.61–0.64] using only genetics (Figure S1C) and 0.66 [0.65–0.68] when combined with age and family history (Fig. 3C). Individuals in the top 10th quantile of PRS constructed from matched ancestry model had twofold greater risk of PrCa compared to

Pagadala *et al. BMC Cancer*   (2022) 22:1289

Page 5 of 11

Fig. 2 Manhattan Plot of Prostate Cancer (PrCa) Risk. Manhattan plot of logistic association of genetic variants with PrCa risk in ELLIPSE African ancestry group ($n = 4,533$). green = known PrCa risk SNPs, blue = SNPs associated with PSA levels

the 50th quantile (Fig. 3D). Quantile analysis with odds of PrCa using a European ancestry group model demonstrates no trend between PRS and PrCa risk in the African ELLIPSE Consortium cohort (Fig. 3E).
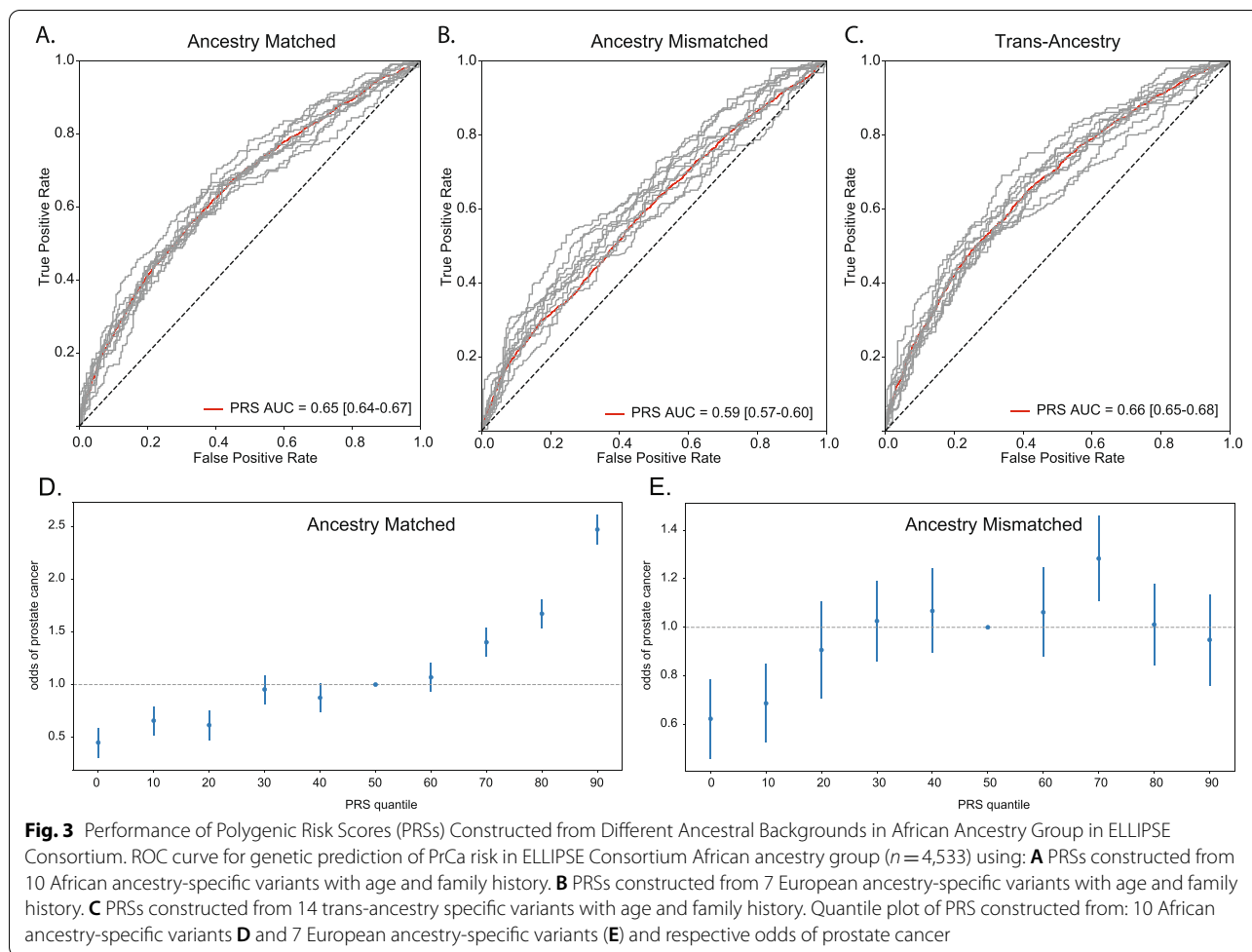
Since the trans-ancestry group model used 5 variants which overlapped with the African ancestry group model, we compared odds ratios of African ancestry-specific risk variants in European, African and trans-ancestry groups to determine if variants exhibited different associations with PrCa (Supplementary Table 4). Odds ratios between trans-ancestry group and African ancestry group association analyses were similar, compared to European ancestry group association analysis (Figure S2). Odds ratios for African ancestry group individuals for 3 of the 8 chromosome 8 variants included in PRS construction were significantly different compared to European ancestry group individuals. These results demonstrate that PrCa variants have different effects based on ancestral background.

African ancestry PrCa Risk variants improve prostate cancer prediction when combined with previous PrCa risk variants

After demonstrating the importance of inclusion of ancestry-matched individuals in polygenic risk score prediction, we sought to evaluate how our PRS constructed with 10 African ancestry-specific variants (PRState) performed in comparison to previous models. We compared performance to a PRS from a previously published large trans-ancestry analysis of PrCa by Conti et al. [11]. Five

of the 10 African ancestry-specific variants we identified were in high linkage disequilibrium with variants previously implicated by Conti et al. and 2 of these 5 variants passed genome-wide significance threshold in the Conti African ancestry group GWAS. We compared our PRState score with the Conti PRS constructed excluding these 5 variants. Addition of PRState score to Conti PRS significantly improves prediction of PrCa by itself (DeLong $P < 0.0003$) (Fig. 4A) and with family history and age (DeLong $P < 0.0003$) (Fig. 4B).

To validate our results, we compared PRState score and Conti PRS performance in the Million Veteran Program. Average AUC for PrCa (Fig. 5A) (DeLong $P < 1e-16$) and metastatic PrCa (Fig. 5B) (DeLong $P < 8e-06$) prediction was significantly higher in individuals of African ancestry when PRState and Conti PRS were combined compared to Conti PRS alone. Combined PRState and Conti score was not associated with significantly higher AUC in predicting death from PrCa (Fig. 5C) (DeLong $P < 0.67$). Interestingly, we find that PRState score performance was associated with significantly better predictive value compared to European individuals for all three defined clinical PrCa endpoints (Fig. 5D, E, F) (PrCa DeLong $P < 1e-16$, metastatic PrCa DeLong $P < 1e-16$, fatal PrCa DeLong $P < 1e-16$). Furthermore, we characterized odds ratios of PRState variants between African and European HARE groups in the Million Veteran Program and noted a significant difference in the odds ratios for certain variants in all
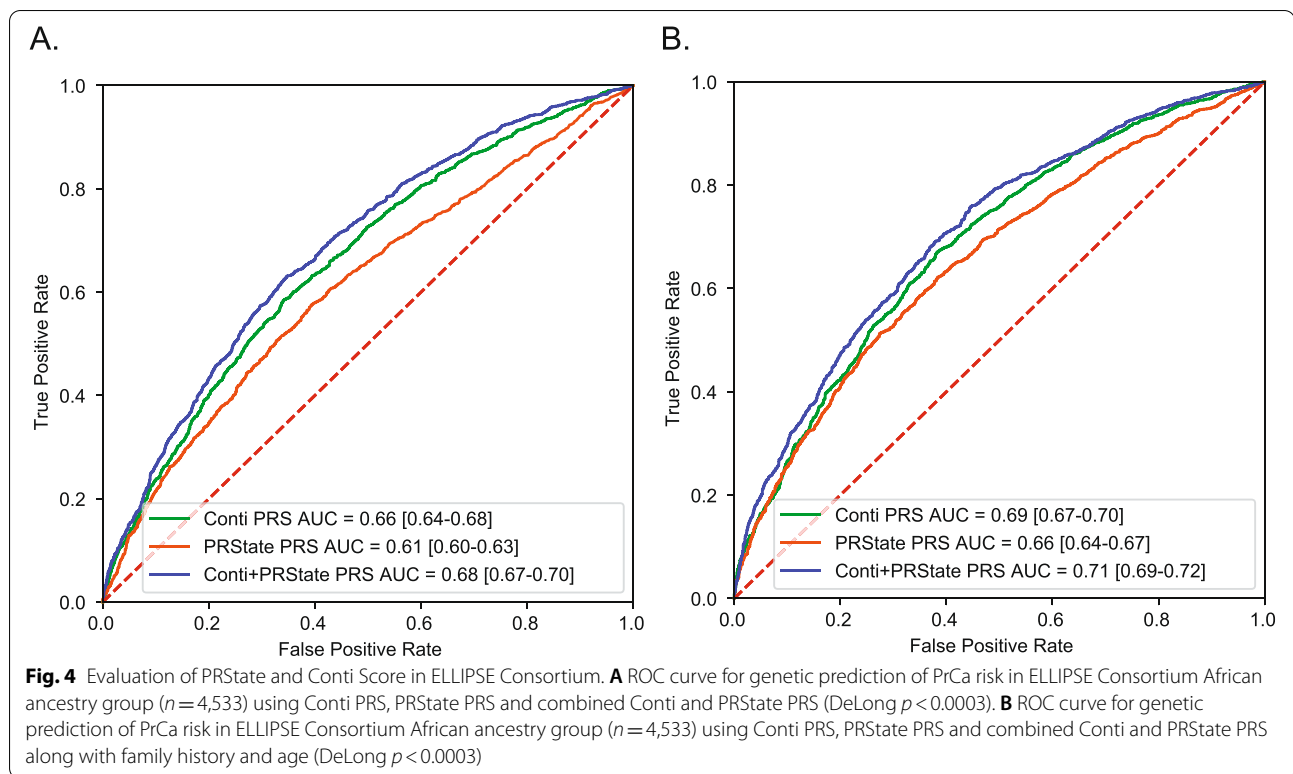
Pagadala *et al. BMC Cancer*   (2022) 22:1289

Page 6 of 11

Fig. 3 Performance of Polygenic Risk Scores (PRSs) Constructed from Different Ancestral Backgrounds in African Ancestry Group in ELLIPSE Consortium. ROC curve for genetic prediction of PrCa risk in ELLIPSE Consortium African ancestry group (*n* = 4,533) using: **A** PRSs constructed from 10 African ancestry-specific variants with age and family history. **B** PRSs constructed from 7 European ancestry-specific variants with age and family history. **C** PRSs constructed from 14 trans-ancestry specific variants with age and family history. Quantile plot of PRS constructed from: 10 African ancestry-specific variants **D** and 7 European ancestry-specific variants (**E**) and respective odds of prostate cancer

three endpoints (PrCa diagnosis, metastasis and death) (Figure S3). The PRState score was significantly associated with all 3 prostate cancer endpoints when we included factors such as income and education level associated with socioeconomic status (Figure S4A). We noted differences in performance of PRState when evaluating risk in African ancestry individuals with high socioeconomic status as compared to low (Figure S4B-D). Specifically, PRState performance in African ancestry men of lower socioeconomic status was lower compared to men of higher socioeconomic status. Furthermore, the region of MVP enrollment had only minor effects on association of PRState with prostate cancer, metastatic prostate cancer and fatal prostate cancer incidence (Figure S5).

These results not only demonstrate the need to include individuals of African ancestry in construction of polygenic risk scores that predict PrCa risk, but also that African ancestry-specific variants are critical for prediction of other PrCa characteristics, such as metastasis.

Discussion

A critical limitation in the majority of genetic studies in PrCa has been the overrepresentation of men with non-Hispanic European ancestry. Considering both the higher incidence and mortality of PrCa in men of African ancestry, this problem prevents the discovery of gene variants conferring PrCa risk in African and other ancestries. Using PrCa risk SNPs identified to be specific for men with African ancestry in the ELLIPSE consortium from chromosomes 3, 8, and 11 we constructed an African ancestry-specific polygenic risk score (PRState) that achieved AUCs of 0.61 [0.60–0.63] alone and 0.65 [0.64–0.67] when family history and age were added. To demonstrate the utility of an ancestry specific PRState, we then compared performance to a mixed European and African ancestry (trans-ancestry) model and a mismatched model from a primarily European ancestral group. We achieved comparable performance using a trans-ancestry group model, but there was a significant drop in performance with the European ancestry group model in ELLIPSE. Additionally, the PRState score improved PrCa
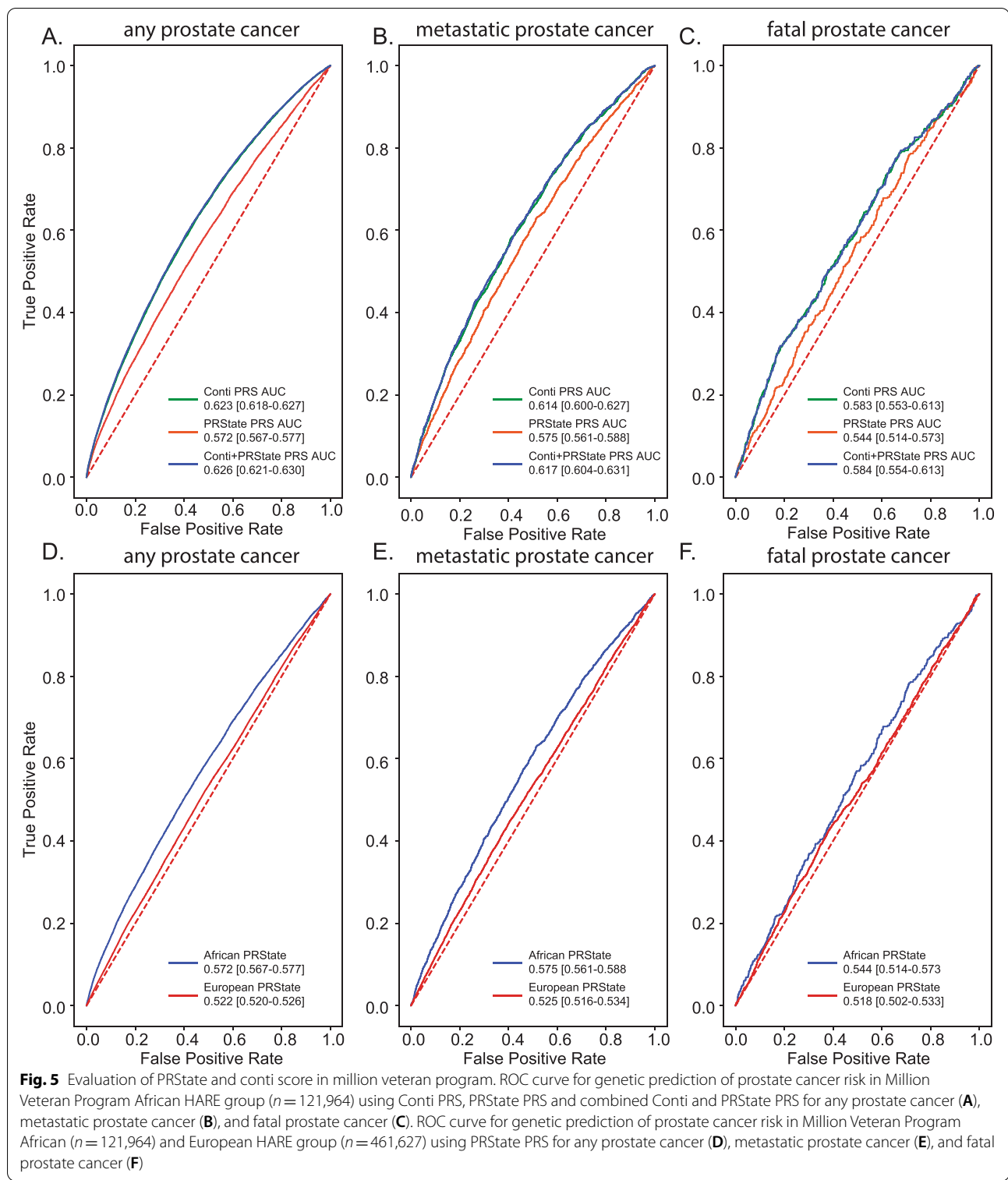
Pagadala *et al. BMC Cancer*   (2022) 22:1289

Page 7 of 11

Fig. 4 Evaluation of PRState and Conti Score in ELLIPSE Consortium. **A** ROC curve for genetic prediction of PrCa risk in ELLIPSE Consortium African ancestry group (*n*=4,533) using Conti PRS, PRState PRS and combined Conti and PRState PRS (DeLong *p* < 0.0003). **B** ROC curve for genetic prediction of PrCa risk in ELLIPSE Consortium African ancestry group (*n*=4,533) using Conti PRS, PRState PRS and combined Conti and PRState PRS along with family history and age (DeLong *p* < 0.0003)

prediction performance when combined with a PRS constructed from a previous larger PrCa meta-analysis conducted by Conti et al. [11] Although half of the PRState variants were in high linkage disequilibrium with Conti et al. variants, we demonstrate these African ancestry-specific variants significantly improve PrCa prediction in our discovery cohort (ELLIPSE) and external validation cohort (Million Veteran Program) of African ancestry group individuals. The results of our PRState study underscore the importance of including men with African ancestry when building genetic risk models and highlight African ancestry-specific PrCa variants that warrant further investigation.

PrCa is one of the most heritable cancer types and ancestry is an important determinant of PrCa risk. Using only a 55 SNP panel and likelihood estimates from forensic genetic tool FROG-Kb, we were able to estimate genetic ancestry that aligned with self-identified ancestry in the ELLIPSE consortium. Although we had self-identified ancestry information available, our approach could be tested in cohorts where self-identified ancestry was not acquired. FROG-kb returns ancestry likelihood estimates for any panel of populations and individuals could have high ancestry likelihood estimates for several groups. For our study, we used FROB-kb to define categorical ancestry groups for PrCa risk variant discovery in the African ancestry group. Specifically, for defining the

African ancestry group, we used a low European likelihood threshold combined with a high African likelihood threshold to define groups with little overlap in principal component analysis (Fig. 1). However, FROB-kb estimates do not have to be used categorically and can also be incorporated in the model as continuous measures. We note that the 55 SNP panel was designed to broadly distinguish populations and was applied to the ELLIPSE cohort that comprises predominantly African ancestry men in the US and UK. This biases our detection of risk variants and the PRState score to be specific to this group. Importantly, there are differences in prostate cancer incidence in African-born and US-born African ancestry men, with Western African-born men having the highest incidence of prostate cancer worldwide [31], and Africa encompasses many genetically diverse groups [32]. Our approach also excluded individuals of admixed ancestry.

Despite these caveats and a relatively small discovery cohort of 4,533 individuals, we were able to identify 10 African ancestry-specific variants (of which 5 were novel), that predicted PrCa in ELLIPSE similar to the PRS constructed from Conti et al. with over 250 variants (ELLIPSE PRState AUC=0.66 [0.64–0.67] versus Conti PRS AUC=0.69 [0.67–0.70]). Half of the PRState variants overlapped with variants identified by Conti et al., which included over 5 × more individuals of African

Fig. 5 Evaluation of PRState and conti score in million veteran program. ROC curve for genetic prediction of prostate cancer risk in Million Veteran Program African HARE group (*n* = 121,964) using Conti PRS, PRState PRS and combined Conti and PRState PRS for any prostate cancer (**A**), metastatic prostate cancer (**B**), and fatal prostate cancer (**C**). ROC curve for genetic prediction of prostate cancer risk in Million Veteran Program African (*n* = 121,964) and European HARE group (*n* = 461,627) using PRState PRS for any prostate cancer (**D**), metastatic prostate cancer (**E**), and fatal prostate cancer (**F**)

ancestry than our discovery cohort. It is worth noting that half of the PRState variants had been reported by Conti et al., whereas the novel variants included in PRState were identified using our methods. We applied a

rigorous selection process for an African ancestry group that allowed identification of PrCa variants with African ancestry-specific effects. In order to test if our small sample was overfitting a prediction model, we applied

Pagadala *et al. BMC Cancer*   (2022) 22:1289

Page 9 of 11

our findings to an external cohort (the Million Veteran Program). PRState improved prediction of PrCa detection and demonstrated these variants were critical for prediction of high-risk PrCa that leads to metastasis in an African ancestry group within the Million Veteran Program. PRState was also associated with significantly higher AUC in the African ancestry group compared to European ancestry group for any PrCa, metastatic PrCa and fatal PrCa. These results suggest PRState improves prediction of metastatic prostate cancer, which may help to safely avoid screening in men with low genetic risk of metastatic disease [21].

In the above analysis, we found PrCa risk SNPs in men of African descent are located at distinct loci that differ from PrCa risk loci identified in men of European ancestry. Variants on chromosome 8 were identified in both European and African ancestry group PrCa GWAS, however, the African ancestry-specific chromosome 8 variants (rs113343238, rs16902008, rs943270004, rs116845582, rs59825493) were not significantly associated with PrCa risk in the European ancestry group in the ELLIPSE Consortium [33, 34]. Interestingly, in the Million Veteran Program, certain variants were protective in the European ancestry group but associated with higher PrCa diagnosis in the African ancestry group (rs113343238, rs943270004). These results suggest that even within well-known PrCa risk loci, defining ancestry group differences will likely improve genetic risk models.

Our PRState analyses identified 10 African ancestry-specific variants in our discovery cohort of 4,533 individuals and demonstrated its improved predictive power. With a larger cohort, we could apply our approach to identify potential novel African ancestry-specific PrCa risk variants. While this work has demonstrated the feasibility of using a small number of SNPs to define ancestry backgrounds and then predict genetic risk of PrCa, there are several limitations. The ELLIPSE data set is the largest complete PrCa cohort and we focused on 4,533 patients in an African ancestry group representing < 5% of the total ELLIPSE cohort. Thus the study was not powered to identify potentially meaningful SNPs, as indicated by the suggestive peaks in the African ancestry group GWAS on chromosomes 9, and 12, which did not reach statistical significance. Additionally, we recognize African ancestry encompasses a wide breadth of genetic diversity that can not be wholly defined by a small sample. Nevertheless, we believe our PRState study indicates the inclusion of ancestral inherited risk is an important variable for analyzing PrCa risk similar to family history. Further investigation of a larger African ancestry sample will likely improve the signal and determine the magnitude of PrCa risk

conferred by the African ancestry-specific SNPs identified in our study. Future studies will expand these methods to other under-represented ancestry groups with a goal of developing PrCa risk stratifying tools based on an individual's ancestral background. Increasing the number of non-white patients in databases such as the ELLIPSE consortium is a key element to furthering research in these groups.

Conclusions

Prostate cancer demonstrates the highest genetic risk of all cancers, and sorting patients by genetic ancestry compared to self-identified ancestry allows discovery of heritable risk SNPs in men with different ancestries. We have shown that a 55 SNP panel can be used to separate European and African ancestral groups and to support identification of risk SNPs unique to African ancestry, which importantly improves prediction of PrCa risk in men of African descent, especially when combined with age and family history.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-022-10258-3.

Additional file 1. Supplemental Table 1: ELLIPSE African Ancestry GWAS Statistics (*n*=4553). **Table 2** ELLIPSE European Ancestry GWAS Statistics (*n*=5567). **Table 3** ELLIPSE European and African Ancestry GWAS Statistics (*n*=10100). **Table 4** Significant Variants Identified in ELLIPSE European, African, and Trans-Ancestry Groups.

Additional file 2. Supplemental Figure 1: Performance of Polygenic Risk Scores Constructed from Different Ancestral Backgrounds in African Ancestry Group in ELLIPSE Consortium. ROC curve for genetic prediction of prostate cancer risk in ELLIPSE Consortium African ancestry group (*n*=4,533) using: (A) Polygenic risk scores constructed from 10 African ancestry-specific variants only. (B) Polygenic risk scores constructed from 7 European ancestry-specific variants only. (C) Polygenic risk scores constructed from 14 trans-ancestry specific variants only. **Figure 2:** Comparison of African Prostate Cancer Variant Odds Ratios in European, African, and Trans-Ancestry Groups in ELLIPSE Consortium. Odds ratios for 10 African ancestry-specific PrCa variants for prostate cancer in the ELLIPSE Consortium for European (*n*=5,667), African (*n*=4,553) and Trans-Ancestry (*n*=10,100) ancestry groups. **Figure 3:** Comparison of African Prostate Cancer Variant Odds Ratios in European and African HAREGroups in Million Veteran Program. Odds ratios for 10 African ancestry-specific PrCa variants for any prostate cancer (A), fatal prostate cancer (B), and metastatic prostate cancer (C) in the African (*n*=121,964) and European (*n*=461,627) HARE Million Veteran Program groups. **Figure 4:** Evaluation of PRState Score in Million Veteran Program Including Socioeconomic Information. (A) Odds ratio of PRState with prostate cancer clinical endpoints when socioeconomic variables were included (green) versus not (orange). Socioeconomic variable include income (>$50,000) and education level (at least bachelor's degree). ROC-AUCanalysis for any (B), metastatic (C) and fatal (D) prostate cancer divided by socioeconomic group. High socioeconomic group are African ancestry individuals with income >$50,000 and at least bachelor's degree. Low socioeconomic group are African ancestry individuals with income <$50,000 and education level less than bachelor's degree. **Figure 5:** Evaluation of PRState Score in Million Veteran Program Based On Geographic Region. Odds ratio of PRState with prostate cancer clinical endpoints divided by VA enrollment site geographical region.

Pagadala *et al. BMC Cancer*   (2022) 22:1289

Page 10 of 11

Availability of data and materials
Summary statistics are within the manuscript and Supporting Information files. The data analyzed in this study were obtained from dbgap (accession id: phs001120.v1.p1) and Million Veteran Program, Office of Research and Development, Veterans Health Administration. Access to raw genotype and phenotype data can be obtained through application. Code used for this manuscript are at: https://github.com/meghanasp21/PRState

Declarations

Ethics approval and consent to participate
The Central Veterans Affairs Institutional Review Board (IRB) and site-specific Research and Development Committees approved the Million Veteran Program study. Analysis of ELLIPSE consortium analysis was approved through dbGaP. All relevant ethical regulations were followed. This publication does not represent the views of the Department of Veterans Affairs or the United States Government.

Consent for publication
Not applicable.

Competing interests
T.M.S reports honoraria from Varian Medical Systems and WebMD; he has an equity interest in CorTechs Labs, Inc. and serves on its Scientific Advisory Board; he has received in-kind research support from GE Healthcare via a research agreement with the University of California San Diego. These companies might potentially benefit from the research results. The terms of this arrangement have been reviewed and approved by the University of California San Diego in accordance with its conflict-of-interest policies.

Author details
[1]Department of Medicine, Division of Medical Genetics, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA. [2]Medical Scientist Training Program, University of California San Diego, La Jolla, CA, USA. [3]Biomedical Science Program, University of California, San Diego, La Jolla, CA, USA. [4]Division of Urology, Maine Medical Center, Portland, ME, USA. [5]Bioinformatics and Systems Biology Program, University of California, San Diego, La Jolla, CA, USA. [6]Department of Radiation Medicine and Applied Sciences, University of California, San Diego, La Jolla, CA, USA. [7]VA San Diego Healthcare System, La Jolla, CA, USA. [8]Department of Radiology, University of California San Diego, La Jolla, CA, USA. [9]Department of Bioengineering, University of California San Diego, La Jolla, CA, USA. [10]Department of Urology, University of California San Diego, La Jolla, CA, USA. [11]VA Salt Lake City Healthcare System, Salt Lake City, UT, USA. [12]School of Medicine, University of Utah, Salt Lake City, UT, USA. [13]Center for Behavioral Genetics of Aging, University of California San Diego, La Jolla, CA, USA. [14]Center of Excellence for Stress and Mental Health (CESAMH), San Diego Healthcare System, San Diego, CA, USA.

References
1. Henley SJ, et al. Annual report to the nation on the status of cancer, part I: National cancer statistics. Cancer. 2020;126:2225–49.
2. Brawley OW. Prostate cancer epidemiology in the United States. World J Urol. 2012;30:195–200.
3. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA Cancer J Clin. 2014;64:9–29.
4. Kolonel LN, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol. 2000;151:346–57.
5. Hjelmborg JB, et al. The heritability of prostate cancer in the Nordic Twin Study of Cancer. Cancer Epidemiol Biomarkers Prev. 2014;23:2303–10.
6. Lichtenstein P, et al. Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000;343:78–85.
7. Mucci LA, et al. Familial risk and heritability of cancer among twins in nordic countries. JAMA. 2016;315:68–76.
8. Benafif S, Kote-Jarai Z, Eeles RA. A review of prostate cancer genome-wide association studies (GWAS). Cancer Epidemiol Biomark Prev. 2018;27:845–57.
9. Eeles R, et al. The genetic epidemiology of prostate cancer and its clinical implications. Nat Rev Urol. 2014;11:18–31.
10. Schumacher FR, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet. 2018;50:928–36.
11. Conti DV, et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. Nat Genet. 2021;53:65–75.
12. Du Z, et al. A genome-wide association study of prostate cancer in Latinos. Int J Cancer. 2020;146:1819–26.
13. Conti DV, et al. Two Novel Susceptibility Loci for Prostate Cancer in Men of African Ancestry. J Natl Cancer Inst. 2017;109:djx084.
14. Helfand BT, Kearns J, Conran C, Xu J. Clinical validity and utility of genetic risk scores in prostate cancer. Asian J Androl. 2016;18:509–14.
15. Liss MA, Xu J, Chen H, Kader AK. Prostate genetic score (PGS-33) is independently associated with risk of prostate cancer in the PLCO trial. Prostate. 2015;75:1322–8.
16. Sun J, et al. Genetic score is an objective and better measurement of inherited risk of prostate cancer than family history. Eur Urol. 2013;63:585–7.
17. Karunamuni RA, et al. Additional SNPs improve risk stratification of a polygenic hazard score for prostate cancer. Prostate Cancer Prostatic Dis. 2021;24:532–41.
18. Huynh-Le M-P, et al. Common genetic and clinical risk factors: association with fatal prostate cancer in the Cohort of Swedish Men. Prostate Cancer Prostatic Dis. 2021. https://doi.org/10.1038/s41391-021-00341-4.
19. Huynh-Le M-P, et al. Polygenic hazard score is associated with prostate cancer in multi-ethnic populations. Nat Commun. 2021;12:1236.
20. Seibert TM, et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. BMJ. 2018;360: j5757.
21. Pagadala MS, et al. Polygenic risk of any, metastatic, and fatal prostate cancer in the Million Veteran Program. J Natl Cancer Inst. 2022;djac199.
22. Huynh-Le M-P, et al. Prostate cancer risk stratification improvement across multiple ancestries with new polygenic hazard score. Prostate Cancer Prostatic Dis. 2022. https://doi.org/10.1038/s41391-022-00497-7.
23. Huynh-Le M-P, et al. A genetic risk score to personalize Prostate Cancer screening, applied to population data. Cancer Epidemiol Biomarkers Prev. 2020;29:1731–8.
24. Karunamuni RA, et al. African-specific improvement of a polygenic hazard score for age at diagnosis of prostate cancer. Int J Cancer. 2021;148:99–105.
25. Kim MS, et al. Testing the generalizability of ancestry-specific polygenic risk scores to predict prostate cancer in sub-Saharan Africa. Genome Biol. 2022;23:1–16.
26. Das S, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48:1284–7.

Pagadala *et al. BMC Cancer*   (2022) 22:1289

Page 11 of 11

27. Kidd KK, et al. The redesigned Forensic Research/Reference on Genetics-knowledge base. FROG-kb Forensic Sci Int Genet. 2018;33:33–7.
28. Purcell S, et al. PLINK: A Tool Set for Whole-Genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
29. Fang H, et al. Harmonizing genetic ancestry and self-identified race/ethnicity in genome-wide association studies. Am J Hum Genet. 2019;105:763–72.
30. Kidd KK, et al. Progress toward an efficient panel of SNPs for ancestry inference. Forensic Sci Int Genet. 2014;10:23–32.
31. Center MM, et al. International variation in prostate cancer incidence and mortality rates. Eur Urol. 2012;61:1079–92.
32. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet. 2008;9:403–33.
33. Haiman CA, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet. 2007;39:638–44.
34. Cheng I, et al. 8q24 and prostate cancer: association with advanced disease and meta-analysis. Eur J Hum Genet. 2008;16:496–505.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.