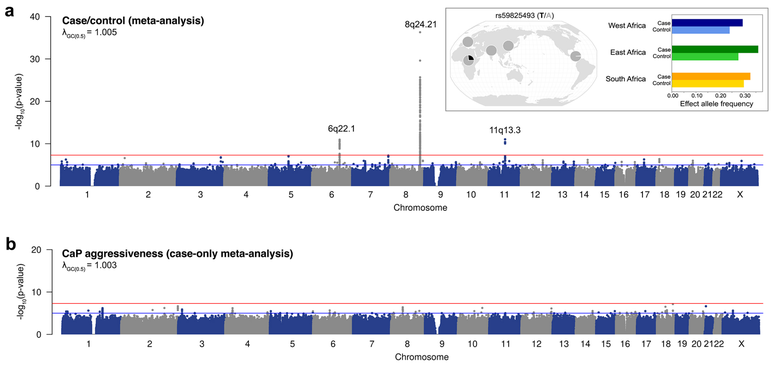
Genetic architecture, evolutionary genomics, and genomic risk of prostate cancer in sub-Saharan Africa

Timothy Rebbeck[1,2], Rohini Janivara[3], Wenlong C. Chen[4], Joseph Lachance[3], Men of African Descent and Carcinoma of the Prostate (MADCaP) Network

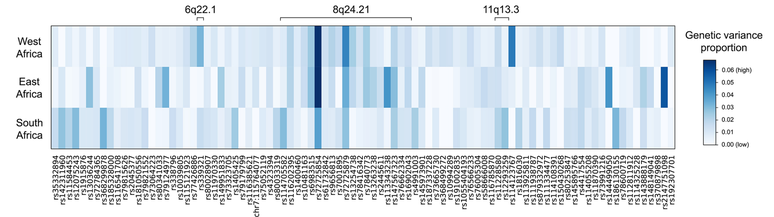
[1]*Dana Farber Cancer Institute*, [2]*Harvard T.H. Chan School of Public Health*, [3]*Georgia Institute of Technology*, [4]*National Cancer Registry (South Africa)*

Abstract

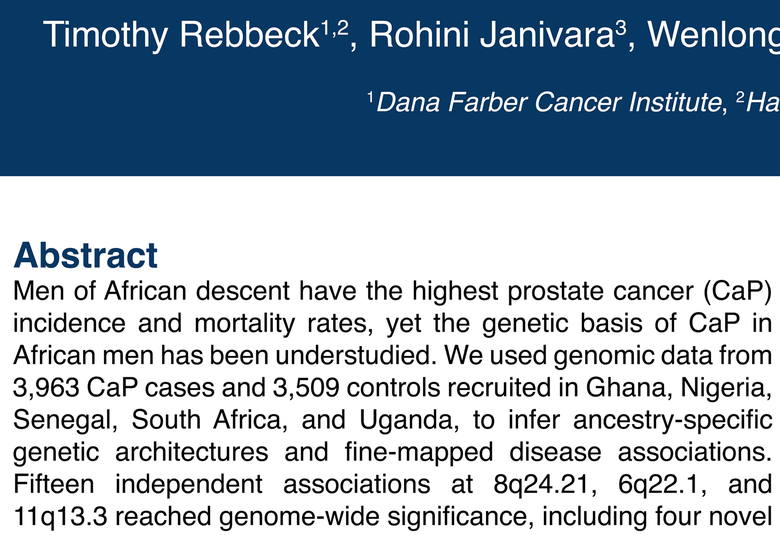
Men of African descent have the highest prostate cancer (CaP) incidence and mortality rates, yet the genetic basis of CaP in African men has been understudied. We used genomic data from 3,963 CaP cases and 3,509 controls recruited in Ghana, Nigeria, Senegal, South Africa, and Uganda, to infer ancestry-specific genetic architectures and fine-mapped disease associations. Fifteen independent associations at 8q24.21, 6q22.1, and 11q13.3 reached genome-wide significance, including four novel associations. Intriguingly, multiple lead SNPs are private alleles, a pattern arising from recent mutations and the out-of-Africa bottleneck. These African-specific alleles contribute to haplotypes with odds ratios above 2.4. We found that the genetic architecture of CaP differs across Africa, with effect size differences contributing more to this heterogeneity than allele frequency differences. Population genetic analyses reveal that African CaP associations are largely governed by neutral evolution. Collectively, our findings emphasize the utility of conducting genetic studies that use diverse populations.



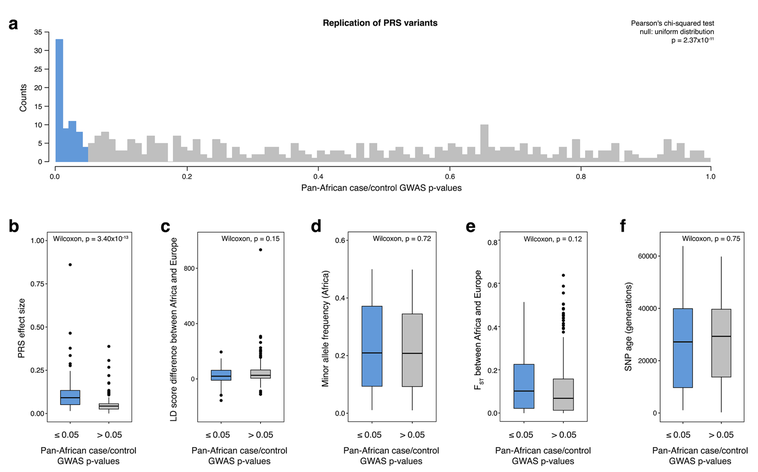
Pan-African GWAS results. Red lines in Manhattan plots mark the genome-wide significant p-value threshold of 5×10-8 and blue lines mark the marginally significant p-value threshold of 1×10-5. a, Case/control pooled meta-analysis of 3,963 African cases and 3,509 African controls. Inset: global and African allele frequencies for rs59825493. b, Case-only GWAS of CaP aggressiveness. Sample size: 3,501 African cases with Grade Group information.



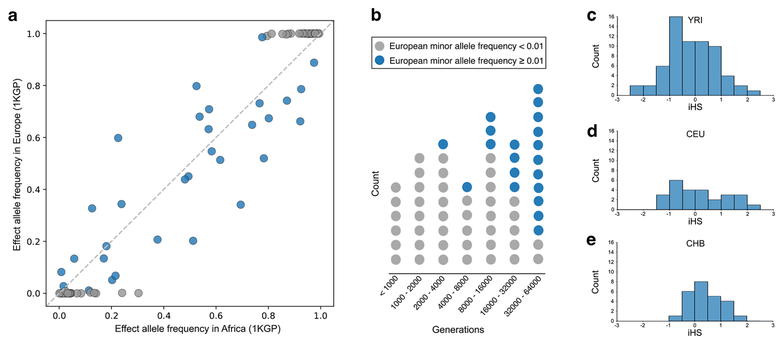
Regional heterogeneity in the genetic architecture of CaP. This plot focuses on 90 independent marginally associated variants (p-value threshold: 1×10-5). Note that LD pruning (r2 < 0.2) yields a slightly different set of variants at 8q24.21 than fine-mapping. a, Heatmap quantifying the relative contributions of 90 CaP-associated variants to the genetic variance of case/control status.



Population genetic structure of African men in this study. a, Principal component analysis (PCA) reveals three broad clusters, corresponding to West, East, and South Africa. Colors of points indicate study site and genotype technology. b, ADMIXTURE plot showing the shared genetic ancestries of samples (K = 5).



Replication of PRS variants in sub-Saharan Africa. Here, we focus on how well PRS variants from Wang et al.49 replicate in our pan-African meta-analysis of CaP cases and controls. a, Previously identified PRS variants are enriched for low p-values in our case/control GWAS (under a null hypothesis of no replication p-values would be uniformly distributed). Bins are colored by pan-African case/control p-value (≤ 0.05 in blue and > 0.05 in gray). Panels b-f examine whether different features of PRS variants from Wang et al.49 PRS are associated with lower p-values in our study (Wilcoxon rank-sum tests): b, Effect sizes (PRS weights); c, LD score differences between Europe and Africa (LD scoreEUR – LD scoreAFR); d, Minor allele frequencies in Africa; e, FST between Europe and Africa (higher values are indicate of larger allele frequency differences between EUR and AFR in the 1KGP); f, SNP age, in generations.



Population and evolutionary genetics of CaP-associated variants. These plots focus on independent marginally associated variants (p-value threshold: 1×10-5) that overlap with 1KGP SNPS. a, Allele frequencies of CaP-associated variants in Europe and Africa (1KGP data, 87 SNPs). Gray circles indicate variants that are not polymorphic in Europe (minor allele frequency < 0.01). b, Ages of 63 CaP-associated variants estimated using GEVA in African populations from the 1KGP data. Gray circles indicate SNPs that are not polymorphic in Europe. c-e, Tests of recent natural selection acting on CaP-associated variants in Africa, Europe, and East Asia (c, YRI: Yoruba in Ibadan, Nigeria; d, CEU: Utah residents with Northern and Western European Ancestry; e, CHB: Han Chinese in Beijing). Because iHS statistics require minor allele frequencies > 1% in the target population, our analysis was restricted to 65 CaP-associated variants in Africa, 26 variants in and 25 variants in East Asia. iHS scores quantify signatures of natural selection, and they follow a standard normal distribution under a null hypothesis of neutral evolution. CaP-associated variants are not enriched for extreme iHS scores (i.e., z-scores ≤ -2 or ≥ 2).

Conclusions

- Multiple genetic associations with CaP are due to private, continent-specific, alleles
- The genetic architecture of CaP varies across Africa
- Most CaP-associated loci are governed by neutral evolution
- There are benefits to conducting genetic studies of diverse understudied populations

Funding: U01CA184374, R01CA257328, and R35GM133727