The contribution of rare variation to prostate cancer heritability

Nicholas Mancuso[1*], Nadin Rohland[2,3*], Kristin A. Rand[4,5], Arti Tandon[2,3], Alexander Allen[2,3], Dominique Quinque[2,3], Swapan Mallick[2,3], Heng Li[2,3], Alex Stram[4] , Xin Sheng[4], Zsofia Kote-Jarai[6], Douglas F. Easton[7], Rosalind A. Eeles[6, 8], the PRACTICAL consortium[9], Loic Le Marchand[10], Alex Lubwama[11], Daniel Stram[4,5], Stephen Watya[11,12], David V Conti[4,5], Brian Henderson[4,5], Christopher Haiman[4,5+], Bogdan Pasaniuc[1+], David Reich[2,3+]

1. Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA
2. Department of Genetics, Harvard Medical School, Boston, MA, USA
3. Broad Institute, Cambridge, MA, USA
4. Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
5. Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA
6. The Institute of Cancer Research, London, UK
7. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
8. Royal Marsden National Health Services (NHS) Foundation Trust, London and Sutton, UK
9. Members from the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium are provided in section 4 of this document.
10. Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA
11. School of Public Health, Makerere University College of Health Sciences, Kampala Uganda.
12. Uro Care, Kampala, Uganda

* These authors contributed equally
+ These authors jointly supervised this work;

Supplementary Note

1. Variance components methods performance in simulated phenotypes from real genotype data

To investigate potential biases in heritability estimation using two variance components, we simulated phenotypes starting from the real genotype data with known proportions from rare and common classes of variants. We then employed the variance components model as implemented in GCTA to estimate the contribution of the rare and common classes of variants. We partitioned the sequenced SNPs into *rare* (0.1% < MAF < 1.0%) and *common* (MAF ≥ 1%) classes. Each simulation instance sampled 10,000 SNPs at random from the *rare* and *common* sets of variants with a proportion α selected from the rare component and $1-\alpha$ from the common component. These variants were used as causal variants to simulate phenotypes. Additive effects were sampled from a normal distribution with mean 0 and unit variance for each causal variant. Given the set of causal variants W, continuous phenotypes y were simulated as $y = Wu + e$ where $W_{ij} = \frac{x_{ij}-2p_i}{\sqrt{2p_i(1-p_i)}}$ is the standardized genotype at causal SNPs and $e \sim N(0, \sigma_e^2)$, with $\sigma_e^2 = var(Wu) \cdot \left(\frac{1}{h_g^2} - 1\right)$. Unless otherwise noted we set $h_g^2 = 0.30$.

AI-REML has better performance than EM-REML in simulations. First, we quantified the performance of Average-Information algorithm [1] and Expectation-Maximization [2] in GCTA across 1,000 instances of simulated phenotypes from real African ancestry genotype data. Since the EM algorithm cannot explore negative parameter space in the *complete-data* likelihood we report implicitly constrained results for the EM approach [3]. We initially focused on simulations at $\alpha = 0$ (no contribution from rare variation) and $\alpha = 43\%$ (as we roughly observe in real phenotype data). In the $\alpha = 0$ case, $h^2_{g,rare}$ estimates for EM, constrained AI-REML, and standard AI-REML were 0.032, 0.012, and 0.002 (see Supplementary Table 5, Supplementary Figure 11). The EM-REML and constrained AI-REML on average produced more conservative standard error estimates (SE=0.031 and SE=0.030) compared to the empirical standard deviation (SD=0.016 and SD=0.018). AI-REML estimates showed accurate calibration of the estimated SE with empirical SD=0.029 matching the theoretical SE=0.029 estimated by GCTA. Estimates for $h^2_{g,common}$ were closer to the true value (0.30) for all methods with 0.291, 0.290, and 0.296 for EM-REML, constrained AI-REML, and AI-REML. Importantly, we did not observe any estimate of $h^2_{g,rare} \geq 0.12$ from any method in any simulation with $\alpha = 0$ thus showing that the empirical p-value for $h^2_{g,rare}$ being non-zero in real phenotype data is <0.001. In the case of $\alpha = 43\%$ $\left(h^2_{g,rare} = 0.13\right)$, AI-REML methods performed identically with estimates of $h^2_{g,rare} = 0.126$ (SD=0.036) and mean standard error of 0.035. The EM-REML method estimated $h^2_{g,rare} = 0.131$ (SD=0.034) and mean standard error 0.035 (see Supplementary Table 6, Supplementary Figure 11). Estimates for $h^2_{g,common}$ were equivalent across all methods with $h^2_{g,common} = 0.168$, SD=0.024, and mean SE=0.024. Simulations over a wide range of values for α showed similar results (Supplementary Figures 3-10). We proceeded with standard

and constrained AI-REML method over EM-REML in all subsequent analyses due to better performance with respect to edge-case estimates (i.e. h_g^2 close to 0).

Genotype dosages attain similar results as best-guess imputed calls. Next, we used simulations to assess dosage data versus best-guess imputed data. We see that both types of data give similar performance across all levels of α. At $\alpha = 0$, standard AI-REML estimates for $h^2_{g,rare}$ yielded 0.000 (SD=0.027) with mean standard error 0.027, while dosage estimates attained 0.002 (SD=0.018) with mean SE=0.030. However at $\alpha = 0.50$, the best-guess estimate of $h^2_{g,rare}$ is 0.142 (SD=0.037) and mean SE=0.035 whereas the dosage $h^2_{g,rare}$ estimate yields 0.148 (SD=0.036) and mean SE=0.036. We decided to proceed with dosage data due to the slightly better performance of estimating $h^2_{g,rare}$ at a value of α closer to what is observed in real phenotype.

LD-adjustment yields increased standard errors over no LD adjustment. We explored the role of accounting for linkage disequilibrium (LD) between SNPs using LD corrected-GRMs as estimated by LDAK version 4.2 [4]. Overall, we find that all approaches yield accurate results at all levels of α with LD-adjustment showing a slightly worse result than no LD adjustment. For example at $\alpha = 0$, LD-adjustment yielded $h^2_{g,rare} = 0.020$ (SD=0.048) with asymptotical standard error 0.049 whereas not adjusting for LD yielded a smaller standard deviation $h^2_{g,rare} = 0.000$ (SD=0.027) with mean standard error 0.027. Therefore, for all subsequent results we used AI-REML without LD adjustments.

Varying the number of causal variants yields similar results. Previous simulations employed 10,000 variants as being causal across the rare and common spectrum. To assess the impact of the number of underlying causal variants we repeated simulations above using 1,000 causal SNPs. Overall, results were qualitatively similar across the 1k and 10k causal variant cases. When $\alpha = 0$, we estimated $h^2_{g,rare} = 0.002$ (SD=0.030) with asymptotical GCTA standard error of 0.028 using AI-REML compared with $h^2_{g,rare} = 0.013$ (SD=0.018) with mean standard error 0.030 for constrained AI-REML. While both AI-REML methods performed comparably, negative variance estimates are difficult to interpret biologically hence we report constrained AI-REML results.

2. Variance components methods performance in real phenotype data

Small cumulative genetic effects due to incomplete quality control (QC) could add up to bias heritability estimation. Additionally, cryptic relatedness can confound SNP-heritability estimates due to shared environment or long-range LD leading to biased results. To understand the effect of various QC we estimated $h^2_{g,rare}$ and $h^2_{g,common}$ under differing thresholds and criteria using real phenotype data. While we investigated these effects across all ancestry groups, we focus below on the African ancestry group, which exhibit a significantly non-zero estimate for $h^2_{g,rare}$.

Missing-ness does not substantially impact variance estimation. First we investigated the impact of missing data on heritability estimation in our data. We removed all SNPs that showed an association to the case-control

status prior to imputation at nominal $p \leq 0.01$ under a log-additive model. This reduced the initial set of SNPs ($n_{rare} = 58{,}699$, $n_{common} = 63{,}972$) to $n_{rare} = 54{,}995$ and $n_{common} = 58{,}366$ in the African ancestry group. However, the removal of these SNPs did not significantly affect the h_g^2 estimates for rare and common components. For example, in the African American ancestry group (n=3476, a subset of the African dataset) we estimated $h_{g,rare}^2 = 0.13$ for both the initial set (SE=0.05) of SNPs as well as the reduced set (SE=0.06), while the estimate $h_{g,common}^2 = 0.18$ (SE=0.03) remained similar at 0.17 (SE=0.04) in the reduced set. Similar results were obtained for all other ancestry groups (Supplementary Tables 8, 9 11-18).

Impact of low-coverage sequencing. Errors in sequencing could also bias heritability estimation. Since accuracy is a direct function of coverage, we removed SNPs with lower coverage from GRM computation at various thresholds. Specifically, we removed SNPs that did not obtain a mean coverage of 0x, 2x, 5x, 7x, and 10x. Overall, we found no significant decrease in $h_{g,rare}^2$ until a large fraction of the SNPs were discarded (coverage ≥7x; see Supplementary Table 19).

Relatedness does not significantly impact estimates. We next filtered samples at various levels of relatedness when restricted to the variant set without differentially occurring SNPs. We determined which individuals to remove by utilizing "*--rel-cutoff*" in PLINK1.9 [5]. Additionally, we expect the relatedness variance in fine-mapping GRMs to be much higher as the targeted regions are smaller than genome-wide arrays. Therefore, individuals were assessed on relatedness based on genome-wide array-based GRMs. As relatedness present in the GRM decreases we observe no statistically significant change of $h_{g,rare}^2$ or $h_{g,common}^2$ across ancestry groups (Supplementary Tables 8, 9, 11-18). Indeed, after filtering samples with relatedness greater than 0.025 we estimated $h_{g,rare}^2 = 0.21$ (SE=0.09) in the African American group compared to our original estimate of $h_{g,rare}^2 = 0.13$ (SE=0.05). As these results statistically overlap, we conclude that relatedness is unlikely to explain the non-zero $h_{g,rare}^2$ estimates.

LD-aware calling does not bias heritability estimation. Finally, to see the value added from imputation we estimated heritability using GRMs computed from the original hard-calls of the same set of variants (with the exception of 2 monomorphic SNPs due to missing calls). Again we observe statistically similar results across all ancestry groups (Supplementary Tables 9, 12, 14, 16, 18). In the African Americans for example $n_{rare} = 55{,}963$, $n_{common} = 66{,}706$, we estimated $h_{g,rare}^2 = 0.11$ (SE=0.05) and $h_{g,common}^2 = 0.20$ (SE=0.04). While both estimates for $h_{g,rare}^2$ are significantly non-zero, we see slightly more signal reclaimed from imputation.

Including rest of genome as extra variance component does not impact the result. We repeated the same analyses by incorporating an additional component to account for arrayed SNPs outside the targeted regions (any SNP at least 0.5 Megabases away from intervals, Supplementary Tables 20, 21). This set was further pruned by removing variants within LD with $r^2 > 0.3$. For the African American ancestry group this included $n_{array} = 251{,}919$ SNPs. We estimated $h_{g,rare}^2$ of 0.13 (SE=0.06) with $h_{g,common}^2 = 0.17$ (SE=0.04) and $h_{g,array}^2 = 0.02$ (SE=0.15). As we decreased the relatedness threshold our estimated values for $h_{g,array}^2$ did not

statistically differ from 0 (in part due to the large standard errors). More importantly, we see no depletion in estimated values for $h^2_{g,rare}$ and $h^2_{g,common}$ when estimated jointly with an array component.

Rare-variant risk scores correlate with case/control status in Africans. To see how well genetic risk-scores correlated with the trait we computed BLUP estimates using the software GCTA. We implemented a 10-fold cross-validation strategy for estimating underlying risk for all ancestry groups and correlated the resulting risk-score directly with the dichotomous case/control status for samples. Since there exists a linear relationship between the underlying liability and observed scales, no adjustment to phenotype is necessary to compute correlation values [6]. In the African ancestry group, we observed a correlation of 0.07 (SE=0.02) for the rare component and in 0.15 (SE=0.02) for the common (Supplementary Table 45).

Similar results were attained across all ancestry groups. The above analyses were repeated with other ethnic groups where we observed similar results (Supplementary Tables 11-18, 22-27). Since the Ugandan samples lacked array calls, investigation into the effect of relatedness in inflating h^2_g and robustness under SNPs from outside the targeted loci could not be performed.

3. Consortium Members

The PRACTICAL Consortium (http://practical.ccge.medschl.cam.ac.uk/):

Rosalind Eeles [1, 2], Doug Easton [3], Zsofia Kote-Jarai [1], Ali Amin Al Olama [3], Sara Benlloch [3], Kenneth Muir [4], Graham G. Giles [5, 6], Fredrik Wiklund [7], Henrik Gronberg [7], Christopher A. Haiman [8], Johanna Schleutker [9, 10], Maren Weischer [11], Ruth C. Travis [12], David Neal [13], Paul Pharoah [14], Kay-Tee Khaw [15], Janet L. Stanford [16, 17], William J. Blot [18], Stephen Thibodeau [19], Christiane Maier [20, 21], Adam S. Kibel [22, 23], Cezary Cybulski [24], Lisa Cannon-Albright [25], Hermann Brenner [26,27], Jong Park [28], Radka Kaneva [29], Jyotsna Batra [30], Manuel R. Teixeira [31], Hardev Pandha[32]

[1] The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey, SM2 5NG, UK, [2] Royal Marsden NHS Foundation Trust, Fulham and Sutton, London and Surrey, UK, [3] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Laboratory, Worts Causeway, Cambridge, UK, [4] University of Warwick, Coventry, UK, [5] Cancer Epidemiology Centre, Cancer Council Victoria, 615 St Kilda Road, Melbourne Victoria, Australia, [6] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria, Australia, [7] Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden, [8] Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA, [9] Department of Medical Biochemistry and Genetics, University of Turku, Turku, Finland, [10] Institute of Biomedical Technology/BioMediTech, University of Tampere and FimLab Laboratories, Tampere, Finland, [11] Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev Ringvej 75, DK-2730 Herlev, Denmark, [12] Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK, [13] Surgical Oncology (Uro-Oncology: S4), University of Cambridge, Box 279, Addenbrooke's Hospital, Hills Road, Cambridge, UK and Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK, [14] Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Strangeways Laboratory, Worts Causeway, Cambridge, UK, [15] Cambridge Institute of Public Health, University of Cambridge, Forvie Site, Robinson Way, Cambridge CB2 0SR, [16] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA, [17] Department of Epidemiology, School of

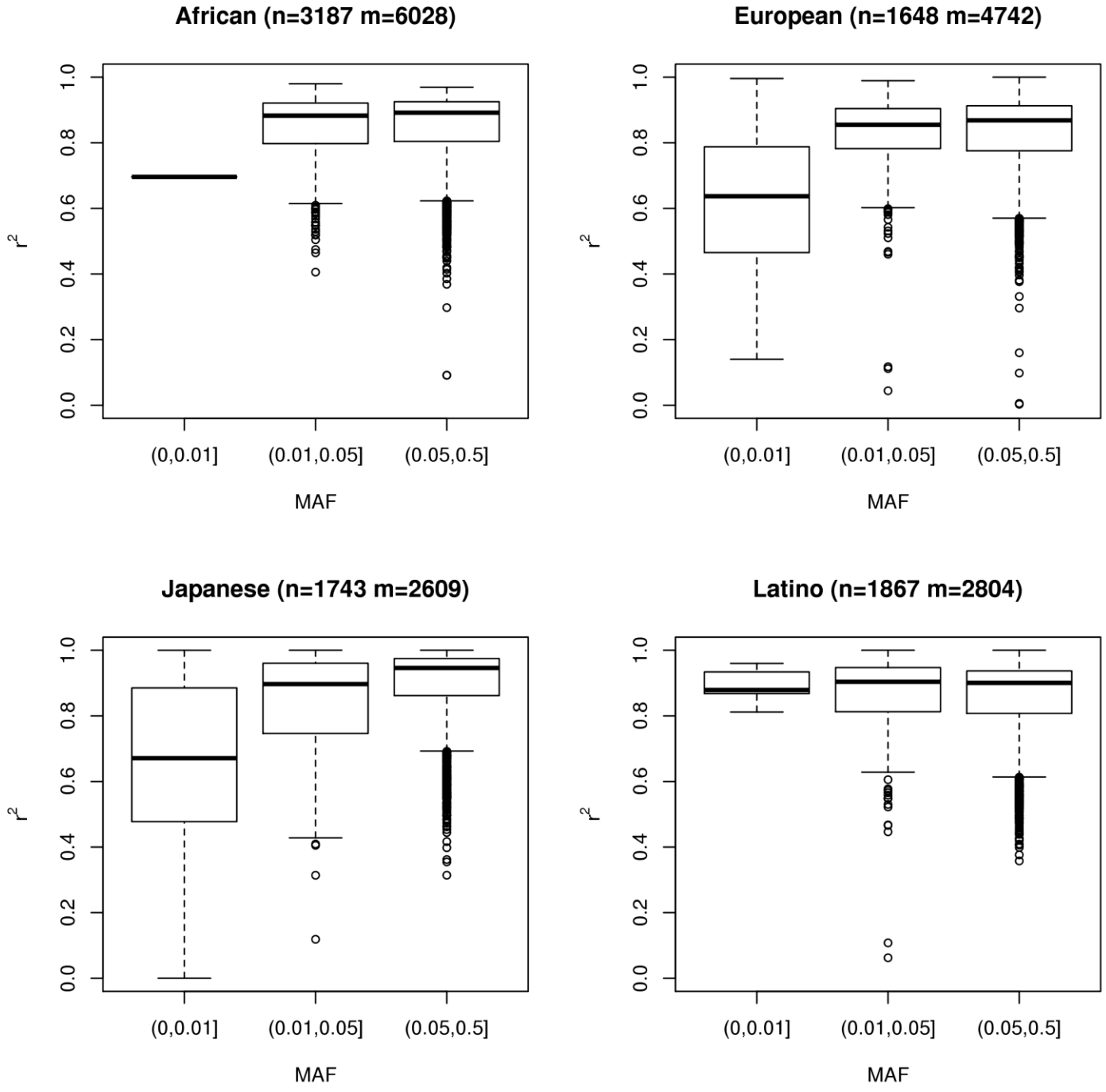
Public Health, University of Washington, Seattle, Washington, USA, [18] International Epidemiology Institute, 1455 Research Blvd., Suite 550, Rockville, MD 20850, [19] Mayo Clinic, Rochester, Minnesota, USA, [20] Department of Urology, University Hospital Ulm, Germany, [21] Institute of Human Genetics University Hospital Ulm, Germany, [22] Brigham and Women's Hospital/Dana-Farber Cancer Institute, 45 Francis Street- ASB II-3, Boston, MA 02115, [23] Washington University, St Louis, Missouri, [24] International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland, [25] Division of Genetic Epidemiology, Department of Medicine, University of Utah School of Medicine[26] Division of Clinical Epidemiology and Aging Research & Division of Preventive Oncology, German Cancer Research Center, Heidelberg Germany, [27]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg Germany, [28]Division of Cancer Prevention and Control, H. Lee Moffitt Cancer Center, 12902 Magnolia Dr., Tampa, Florida, USA, [29] Molecular Medicine Center and Department of Medical Chemistry and Biochemistry, Medical University - Sofia, 2 Zdrave St, 1431, Sofia, Bulgaria, [30] Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation and Schools of Life Science and Public Health, Queensland University of Technology, Brisbane, Australia, [31] Department of Genetics, Portuguese Oncology Institute, Porto, Portugal and Biomedical Sciences Institute (ICBAS), Porto University, Porto, Portugal, [32]The University of Surrey, Guildford, Surrey, GU2 7XH, UK

COGS acknowledgement and funding:

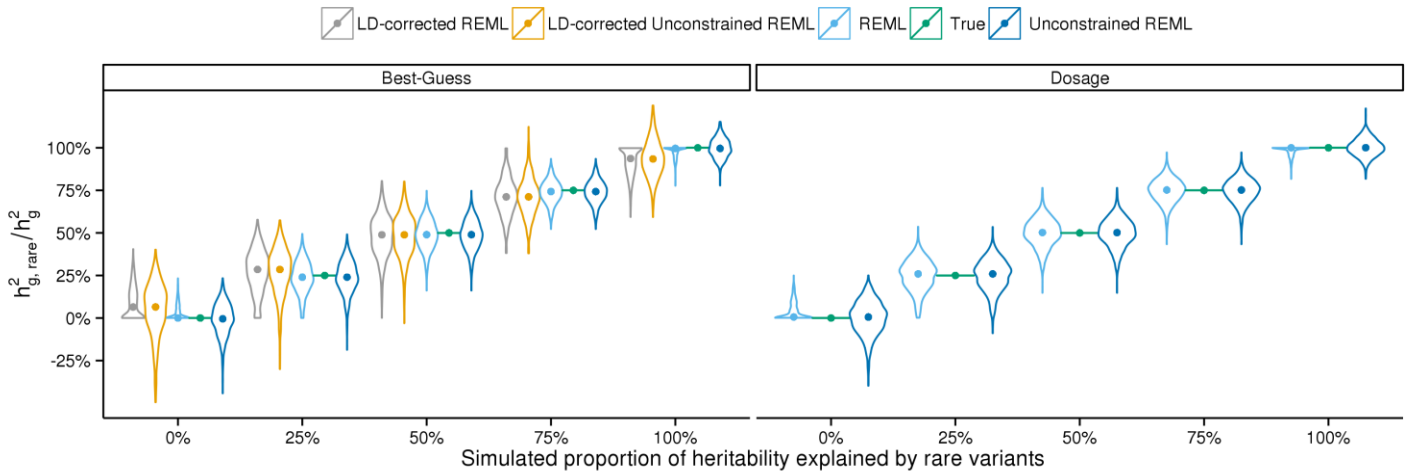
This study would not have been possible without the contributions of the following: Per Hall (COGS); Douglas F. Easton, Paul Pharoah, Kyriaki Michailidou, Manjeet K. Bolla, Qin Wang (BCAC), Andrew Berchuck (OCAC), Rosalind A. Eeles, Douglas F. Easton, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Georgia Chenevix-Trench, Antonis Antoniou, Lesley McGuffog, Fergus Couch and Ken Offit (CIMBA), Joe Dennis, Alison M. Dunning, Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Jacques Simard and Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, Stig E. Bojesen, Sune F. Nielsen, Borge G. Nordestgaard, and the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility

Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund.

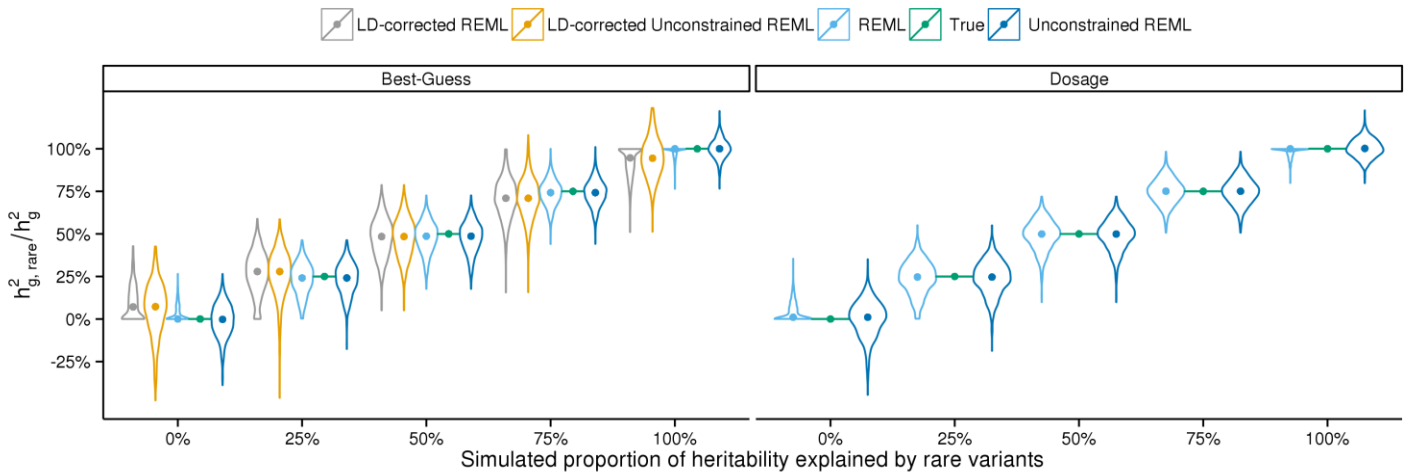
Supplementary Figures



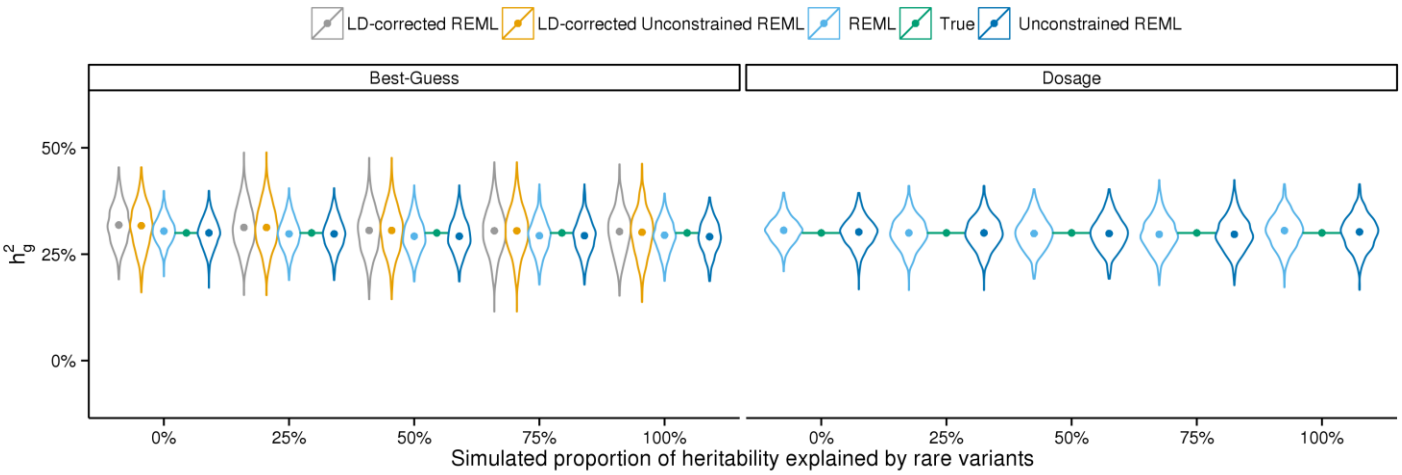
Supplementary Figure 1. Squared Pearson correlation r^2 of the genotype dosages derived from sequencing and hard calls from array data where n is the sample size and m is the amount of SNPs in common. The African ancestry group had 1, 637, and 5390 SNPs for each respective bin; European 28, 275, and 4439; Japanese 97, 226, and 2286; and Latino group 5, 186, 2613 respectively. To estimate accuracy after imputation, we did not replace genotypes for 1,172 African samples with GWAS arrays prior to LD-aware calling; we computed an r^2 of 0.92 across these samples after LD-aware calling as opposed to 0.84 before imputation.



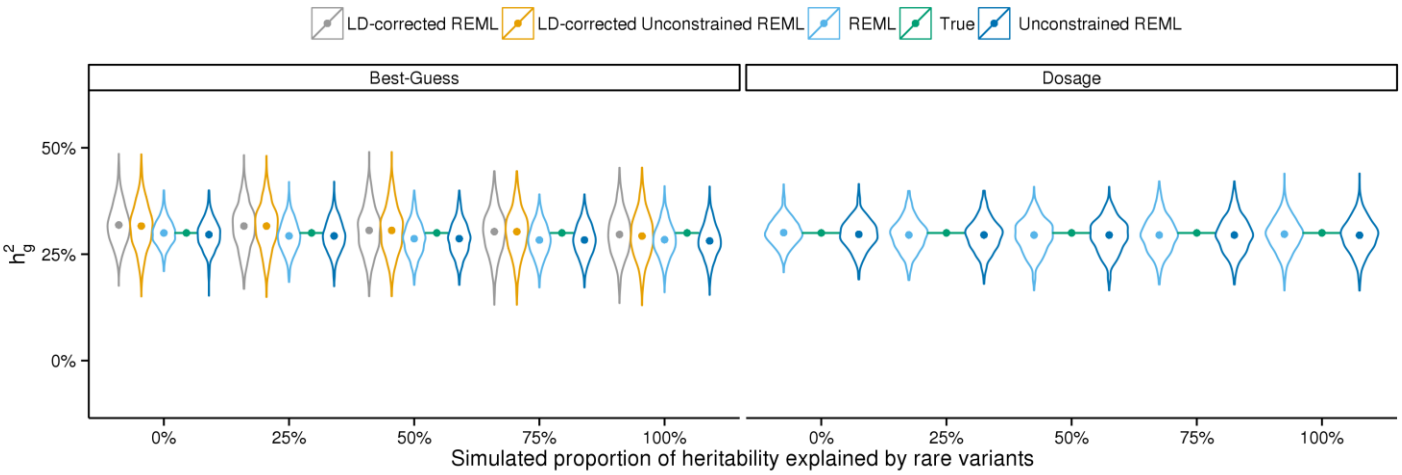
Supplementary Figure 2. Estimated proportion of rare to total SNP heritability ($h^2_{g,rare}/h^2_g$) for continuous phenotypes simulated from African ancestry dosage data (cases=2,054, controls=1,952). 10,000 causal SNPs were selected at random with a proportion sampled from 58,404 rare sequenced SNPs (57,010 best-guess) and 1 – proportion sampled from the 64,267 common sequenced SNPs (62,545 best-guess). This procedure was repeated 1,000 times for each proportion. Allelic effects for each causal set were sampled from the standard-normal distribution such that each SNP explains an equal amount of variance in expectation, given $h^2_g = 0.30$. Different methods of estimating the GRM, such as GRMs estimated with or without LD correction, and constrained versus un-constrained REML estimation are denoted in different colors. GCTA was used to sample effects, construct standard GRMs, and estimate h^2_g. LDAK was used to construct LD-adjusted GRMs.



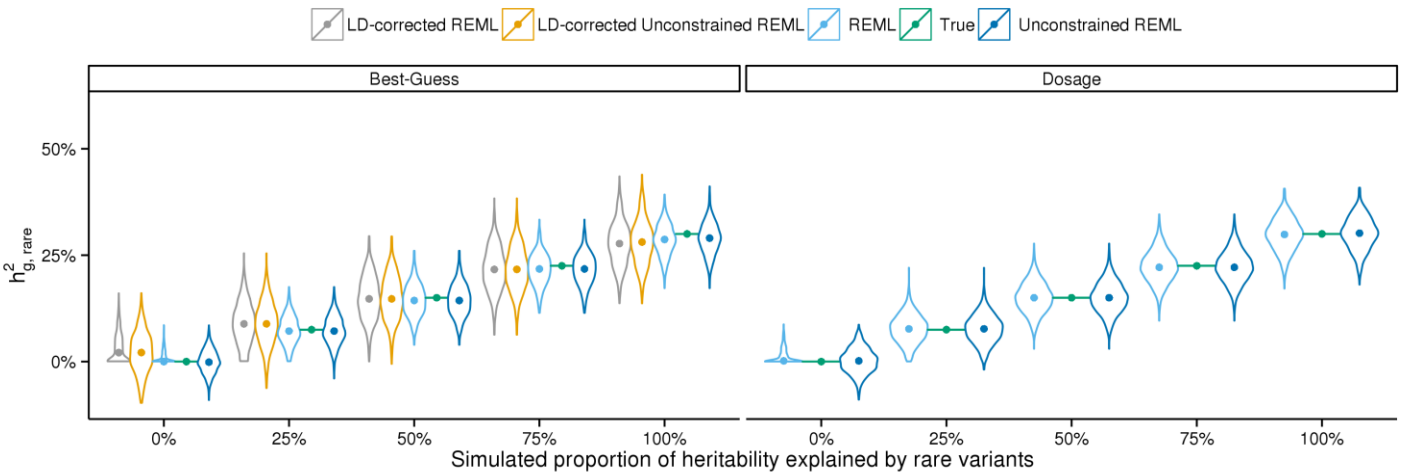
Supplementary Figure 3. Estimated proportion of rare to total SNP heritability ($h^2_{g,rare}/h^2_g$) for phenotypes simulated from African ancestry dosage data and best-guess calls using 1,000 SNPs as causal. See Figure S3 for simulation details.



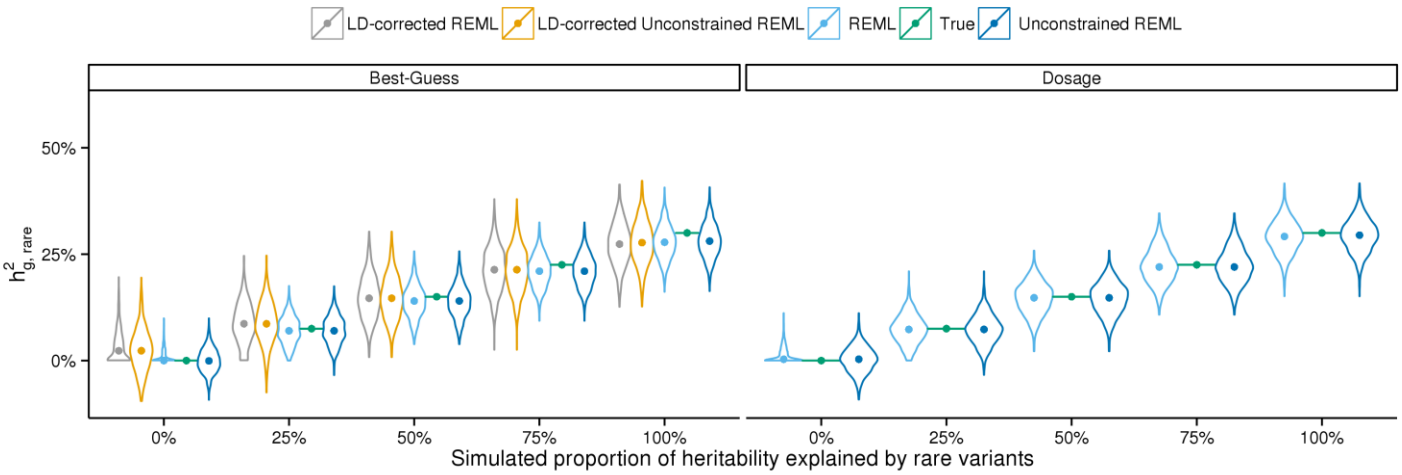
Supplementary Figure 4. Estimated total SNP heritability ($h^2_{g,rare} + h^2_{g,common}$) for phenotypes simulated from African ancestry dosage data and best-guess calls using 10,000 SNPs as causal. See Figure S3 for simulation details.



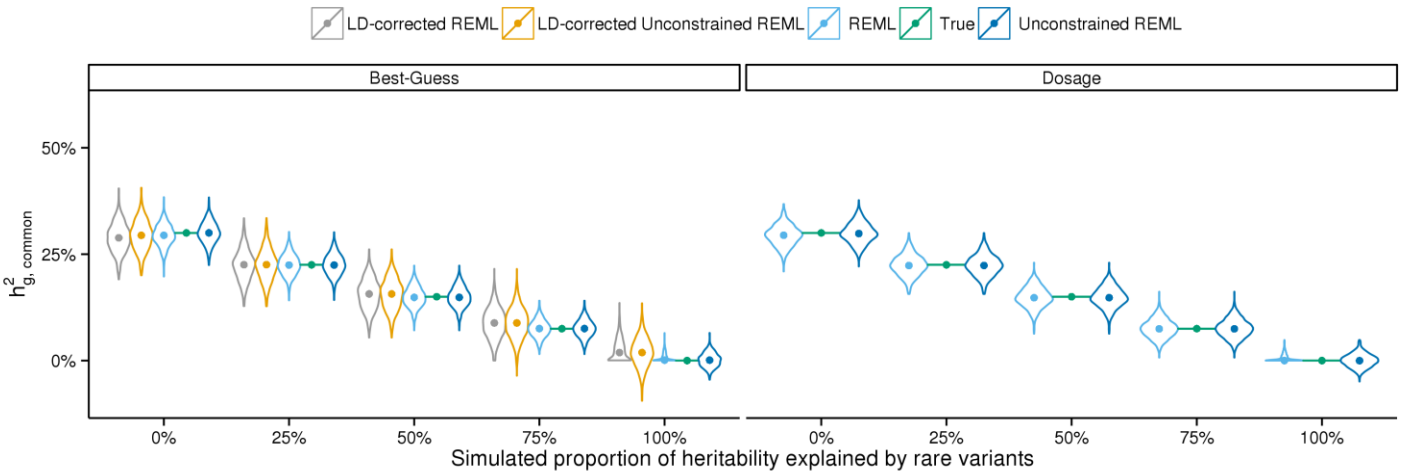
Supplementary Figure 5. Estimated total SNP heritability ($h^2_{g,rare} + h^2_{g,common}$) for phenotypes simulated from African ancestry dosage data and best-guess calls using 1,000 SNPs as causal. See Figure S3 for simulation details.



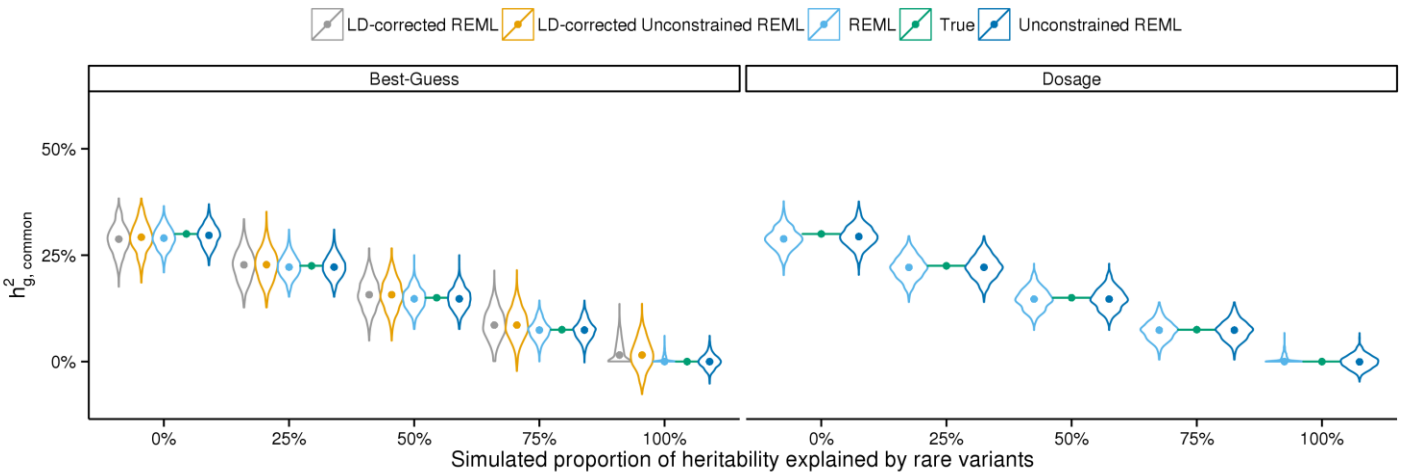
Supplementary Figure 6. Estimated rare SNP heritability ($h^2_{g,rare}$) for phenotypes simulated from African ancestry dosage data and best-guess calls using 10,000 SNPs as causal. See Figure S3 for simulation details.



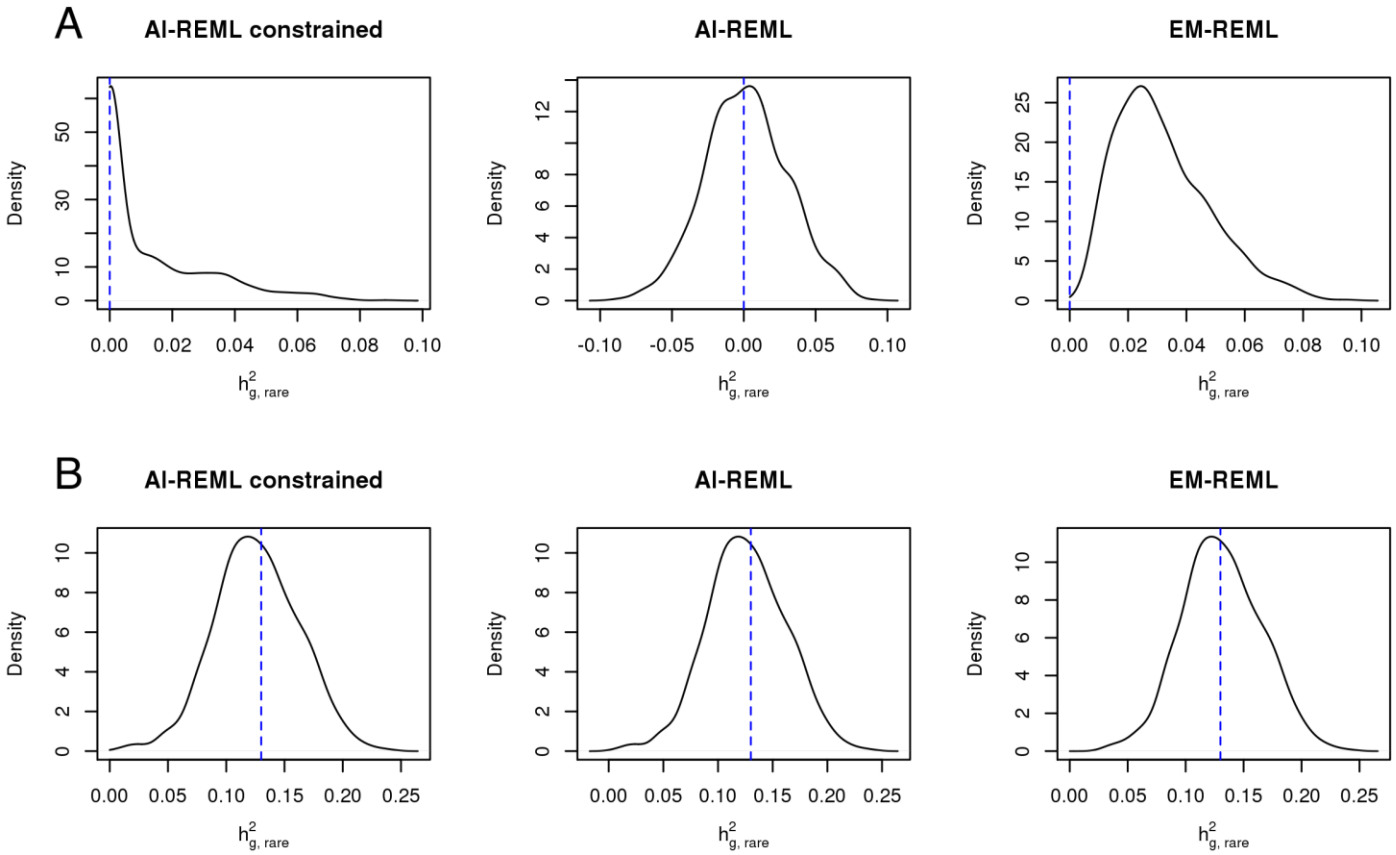
Supplementary Figure 7. Estimated rare SNP heritability ($h^2_{g,rare}$) for phenotypes simulated from African ancestry dosage data and best-guess calls using 1,000 SNPs as causal. See Figure S3 for simulation details.



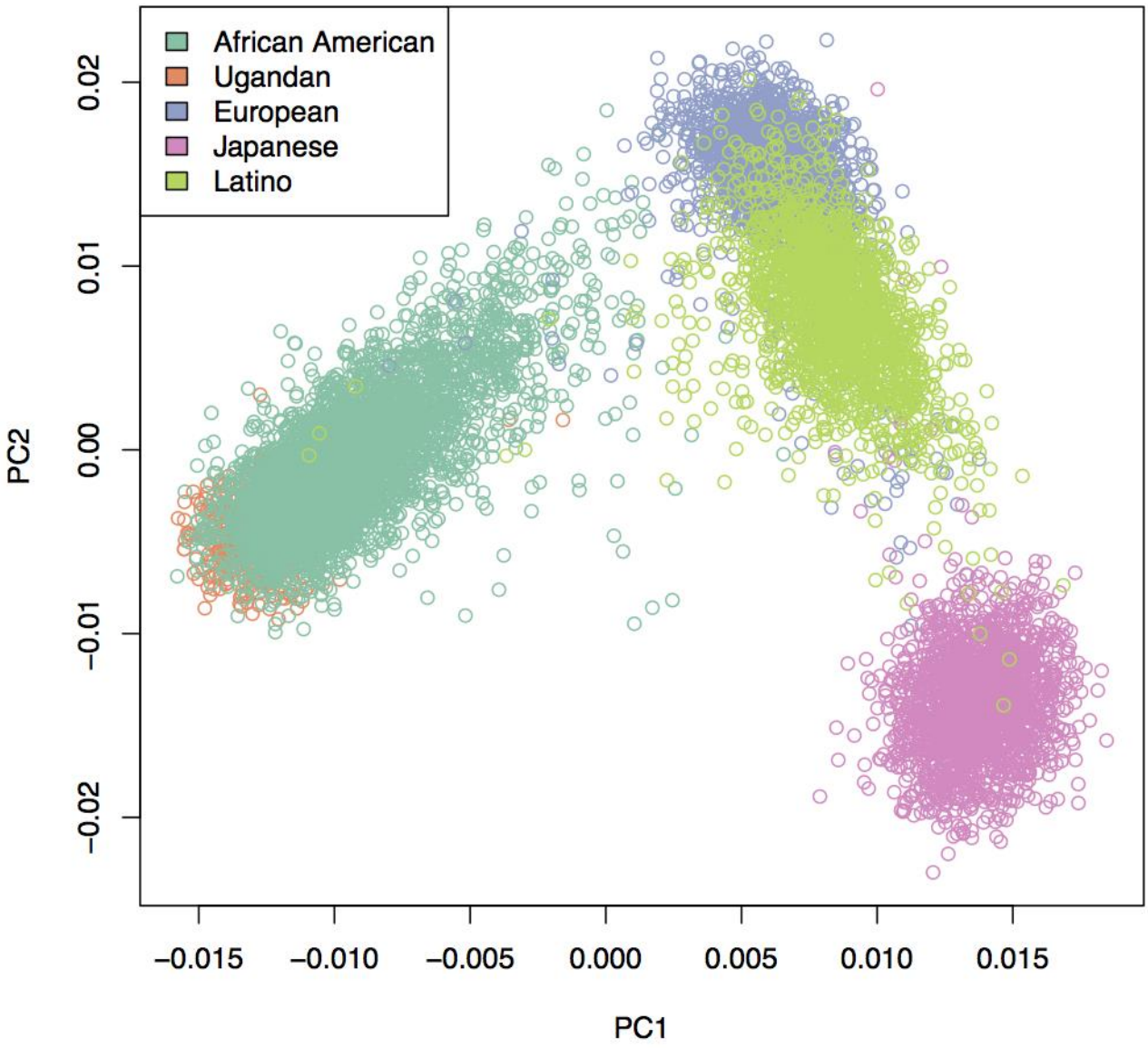
Supplementary Figure 8. Estimated common SNP heritability ($h^2_{g,common}$) for phenotypes simulated from African ancestry dosage data and best-guess calls using 10,000 SNPs as causal. See Figure S3 for simulation details.



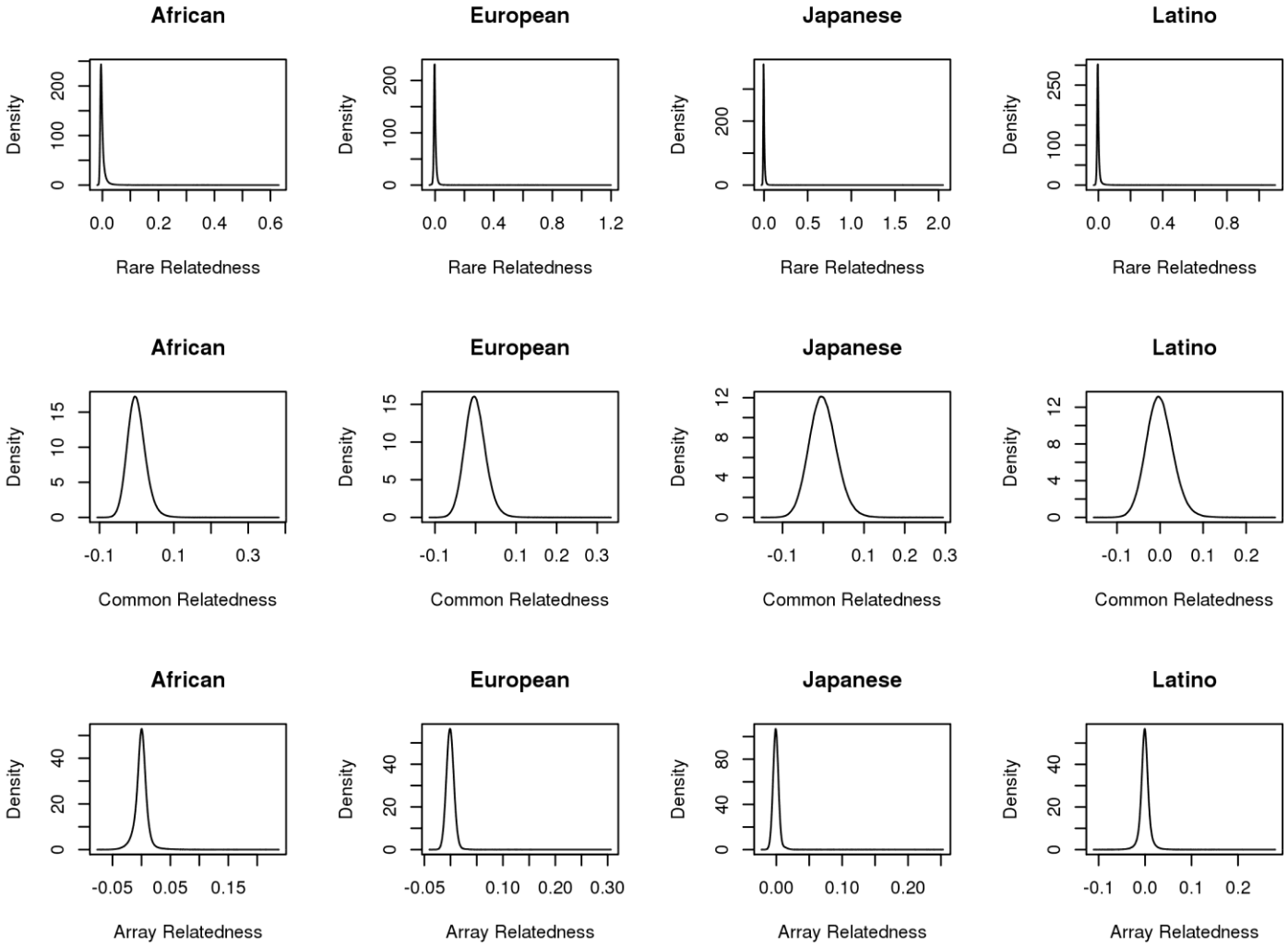
Supplementary Figure 9. Estimated common SNP heritability ($h^2_{g,common}$) for phenotypes simulated from African ancestry dosage data and best-guess calls using 1,000 SNPs as causal. See Figure S3 for simulation details.



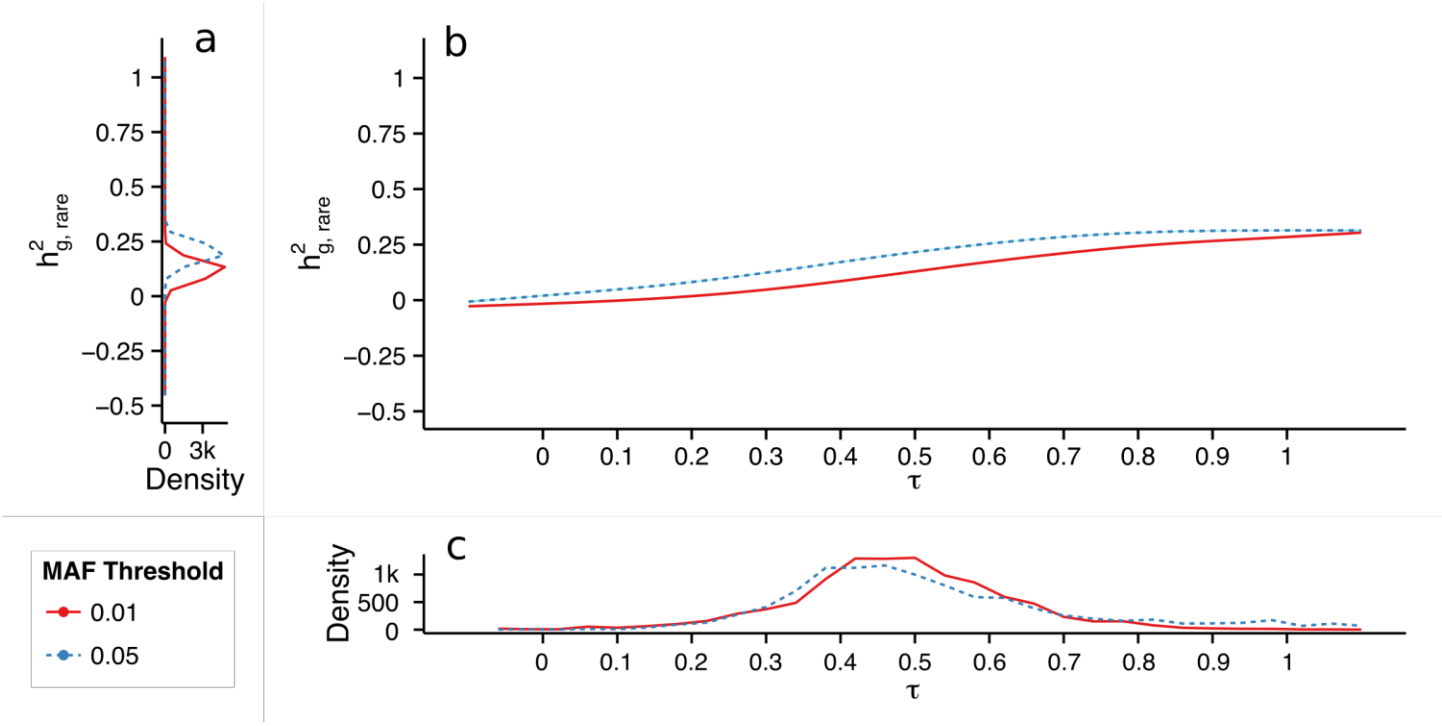
Supplementary Figure 10. Density plots for $h^2_{g,rare}$ estimates reported from constrained AI-REML, standard AI-REML and EM-REML over 1,000 simulated phenotypes. Dashed blue lines represent the simulated $h^2_{g,rare}$ 0 (A) and 0.13 (B).



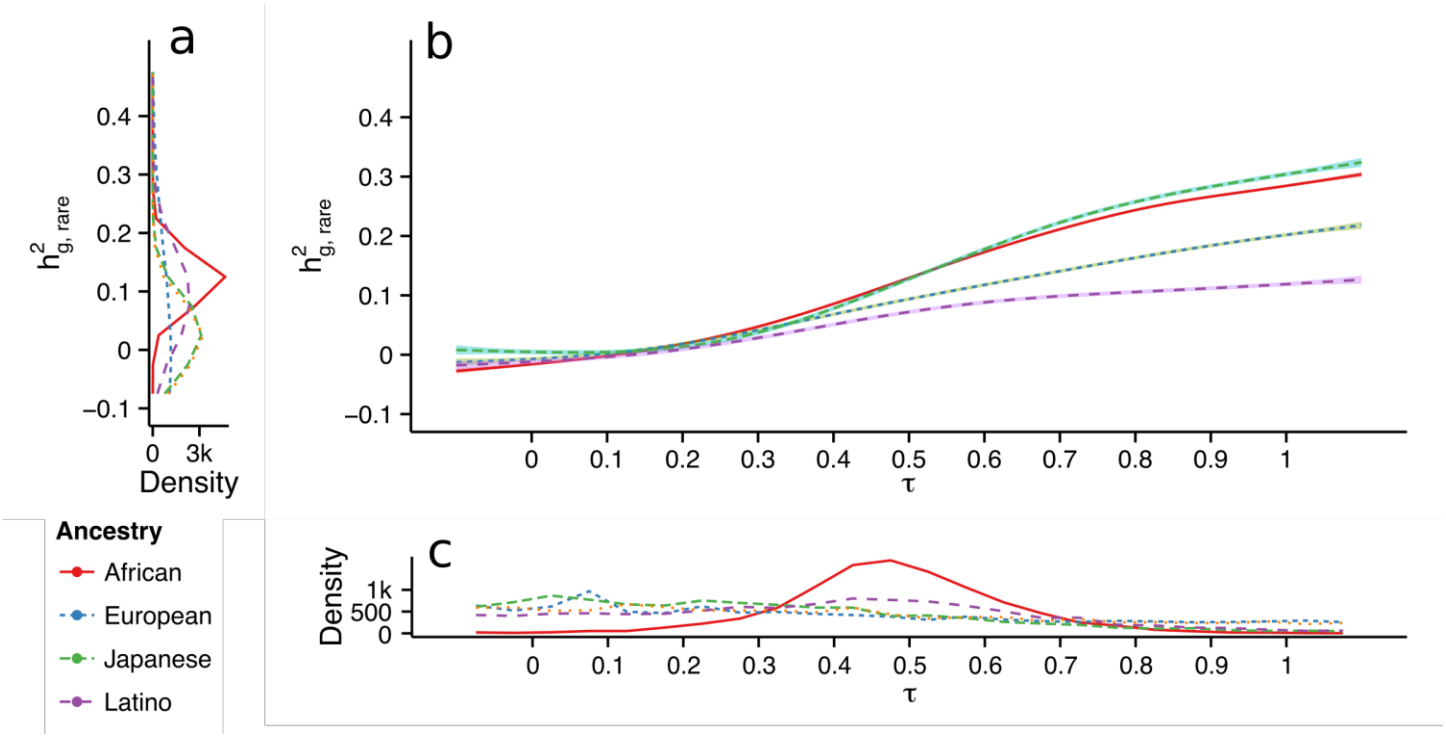
Supplementary Figure 11. Principal components analysis for all ancestry groups based on GRMs estimated from dosage data using all SNPs (rare and common).



Supplementary Figure 12. Density plots for estimates of (off-diagonal) elements in the GRMs. Each row corresponds to relatedness over rare, common, and arrayed variants for each population group.

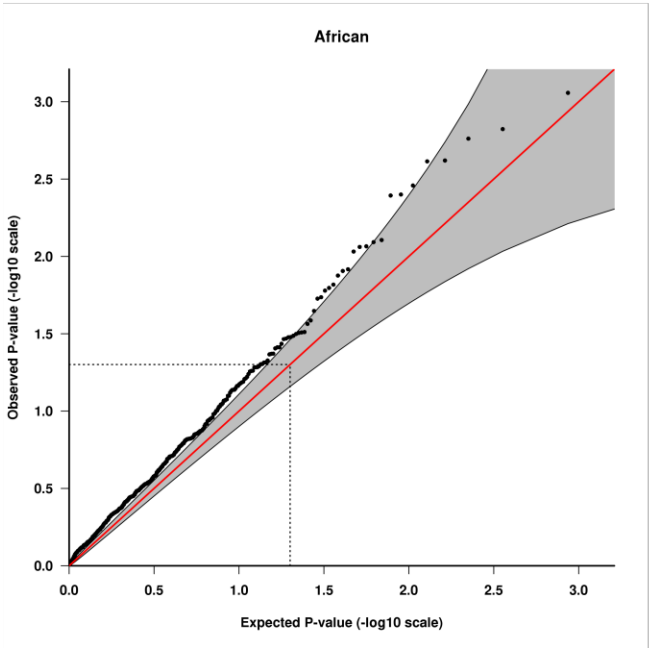


Supplementary Figure 13. Relationship between strength of selection, coupling parameter τ, and allelic effect sizes in PrCa using heritability partitioning for the African ancestry sample using two thresholds on MAF (0.01, and 0.05) to partition variance components. Figure a) density estimate for $h^2_{g,rare}$ obtained from sequence data under normality assumptions resulting from 10,000 sampled points with $\mu = h^2_{g,rare}$ and $\sigma = \widehat{SE}$. Figure b) depicts the influence of τ on $h^2_{g,rare}$. Each point represents an estimate of $h^2_{g,rare}$ given phenotypes simulated from real genotype under the Eyre-Walker model. Figure c) displays the estimated empirical density of τ. Estimates were obtained by matching a sampled value of $h^2_{g,rare}$ from Figure a) with the closest point-estimate from Figure b).

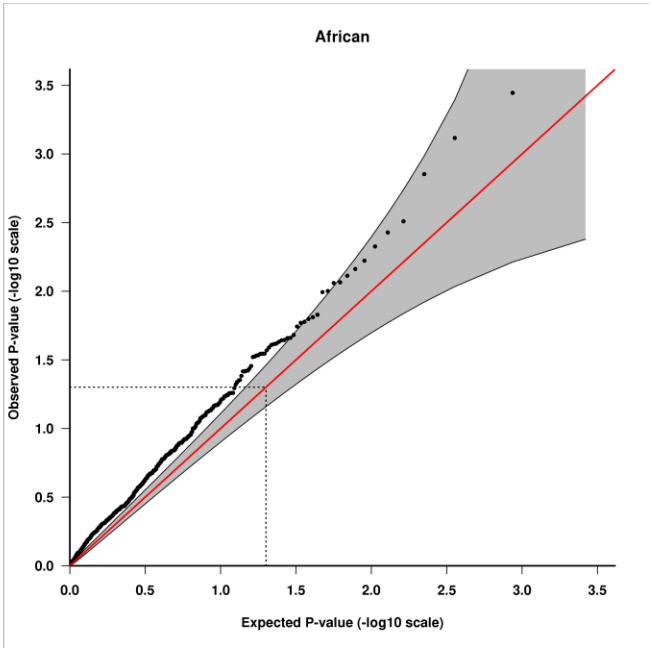


Supplementary Figure 14. Relationship between strength of selection, coupling parameter τ, and allelic effect sizes in PrCa using heritability partitioning for all ancestry groups. Figure a) shows the density estimate for $h^2_{g,rare}$ obtained from real data. Figure b) depicts the mean-value of $h^2_{g,rare}$ as function of τ. Each point represents an estimate of $h^2_{g,rare}$ given phenotypes simulated from real genotype under the Eyre-Walker model. Figure c) displays the estimated empirical density of τ. Estimates were obtained by matching a sampled value of $h^2_{g,rare}$ from Figure a) with the closest point-estimate from Figure b).

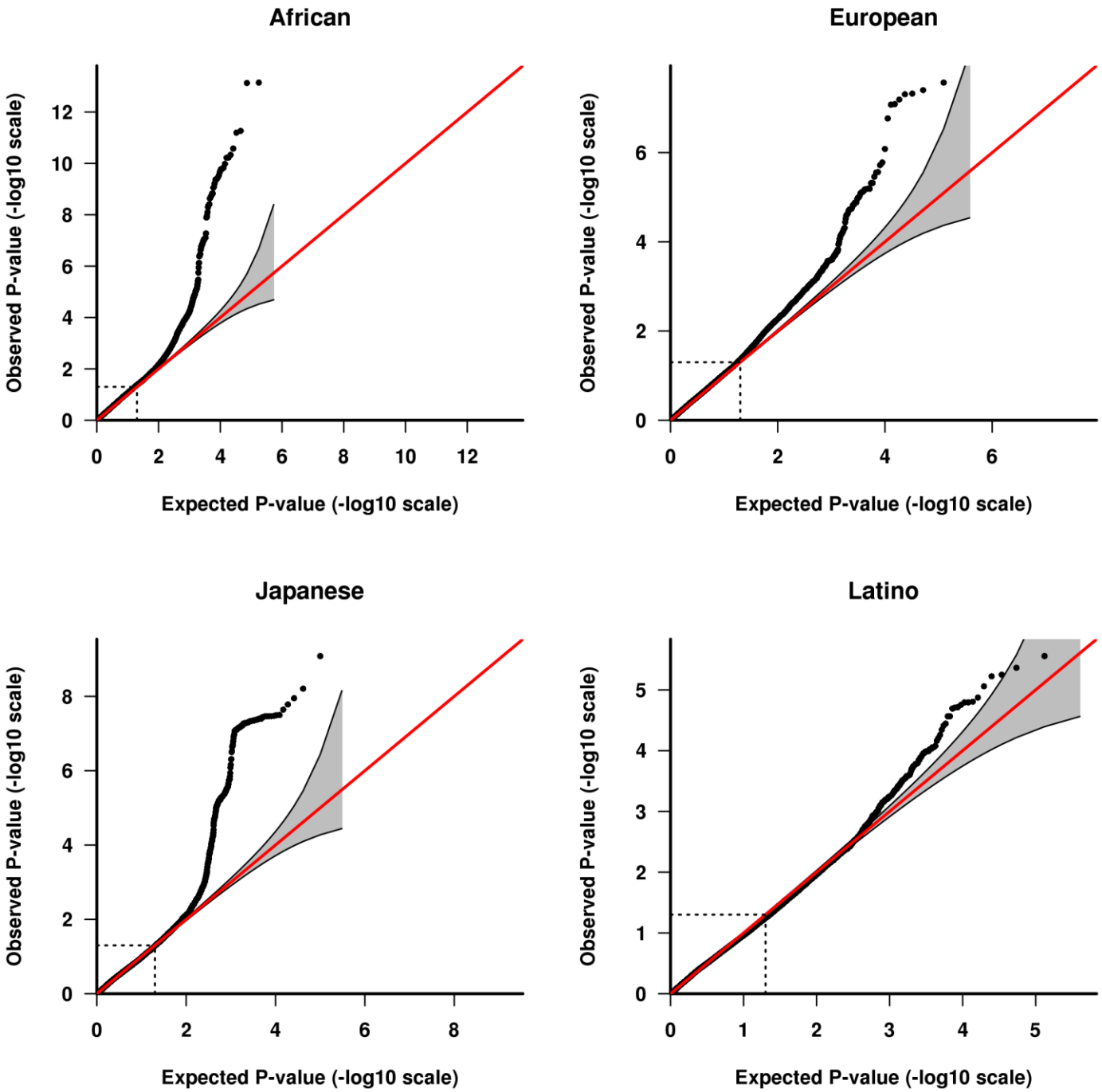
A



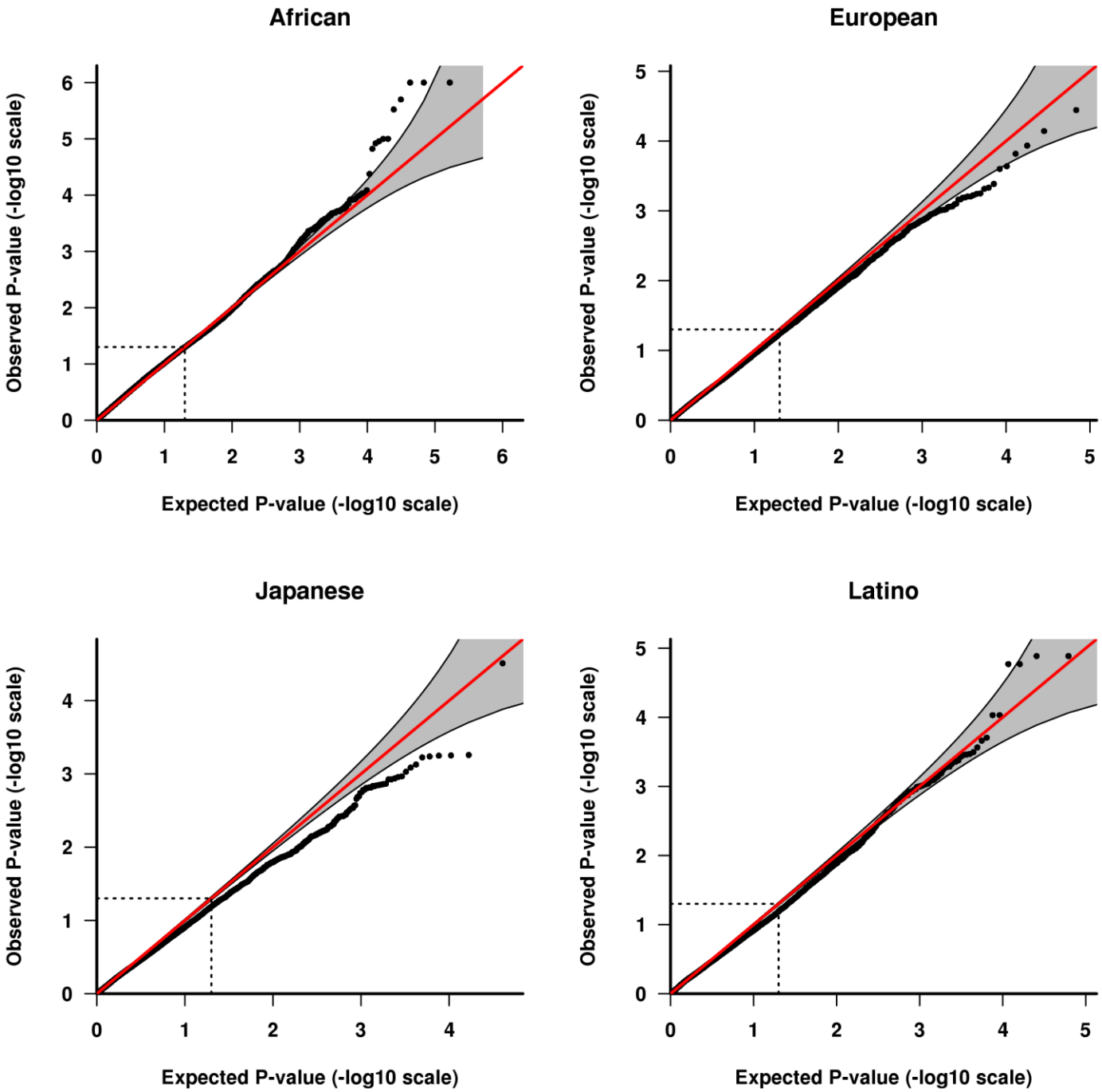
B



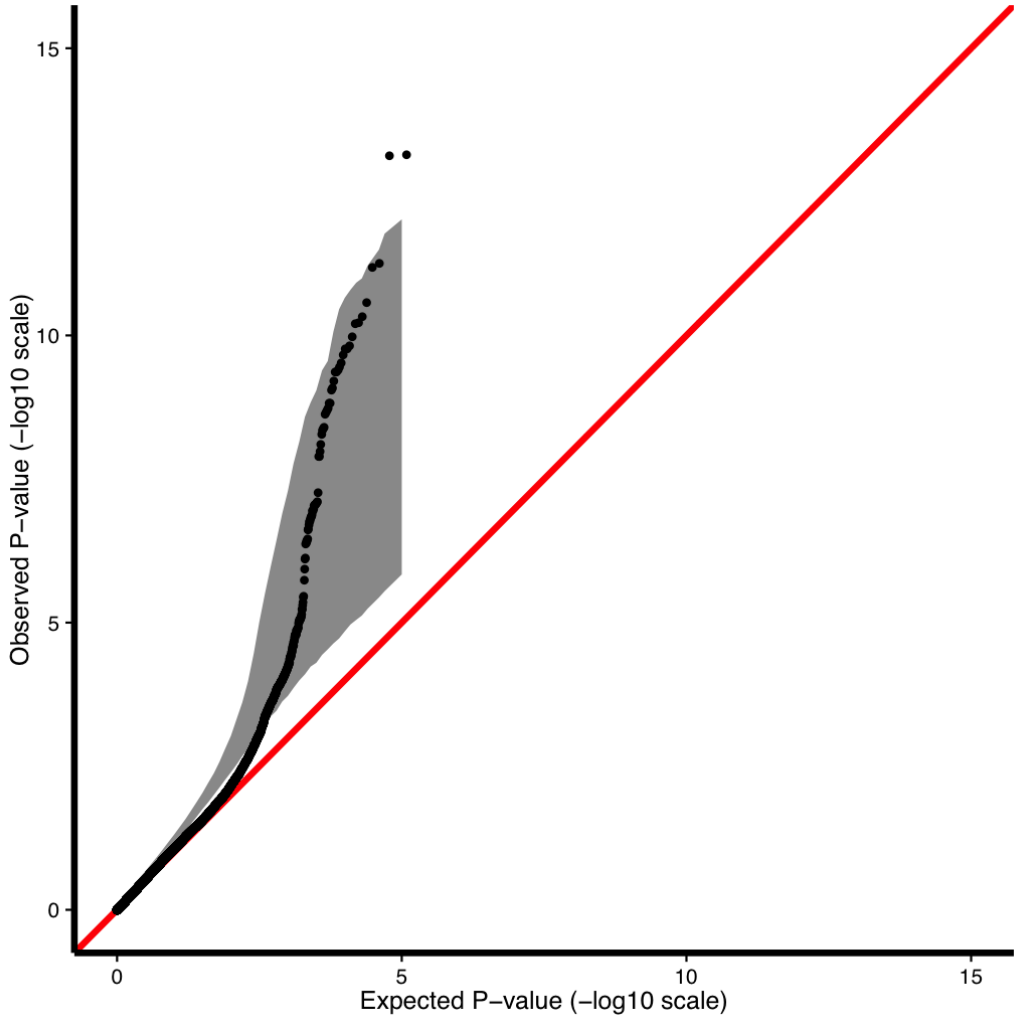
Supplementary Figure 15. QQ-Plot of association statistics from SKAT-O tests over non-overlapping sliding windows across the targeted regions for the African ancestry group when tested marginally (A) and when conditioned on the known risk variants (B). We grouped at most 100 SNPs within each of the 63 regions resulting in a total of 601 sets to test for association. Tests were run with PLINKSEQ software version 0.10 using the best-guess calls.



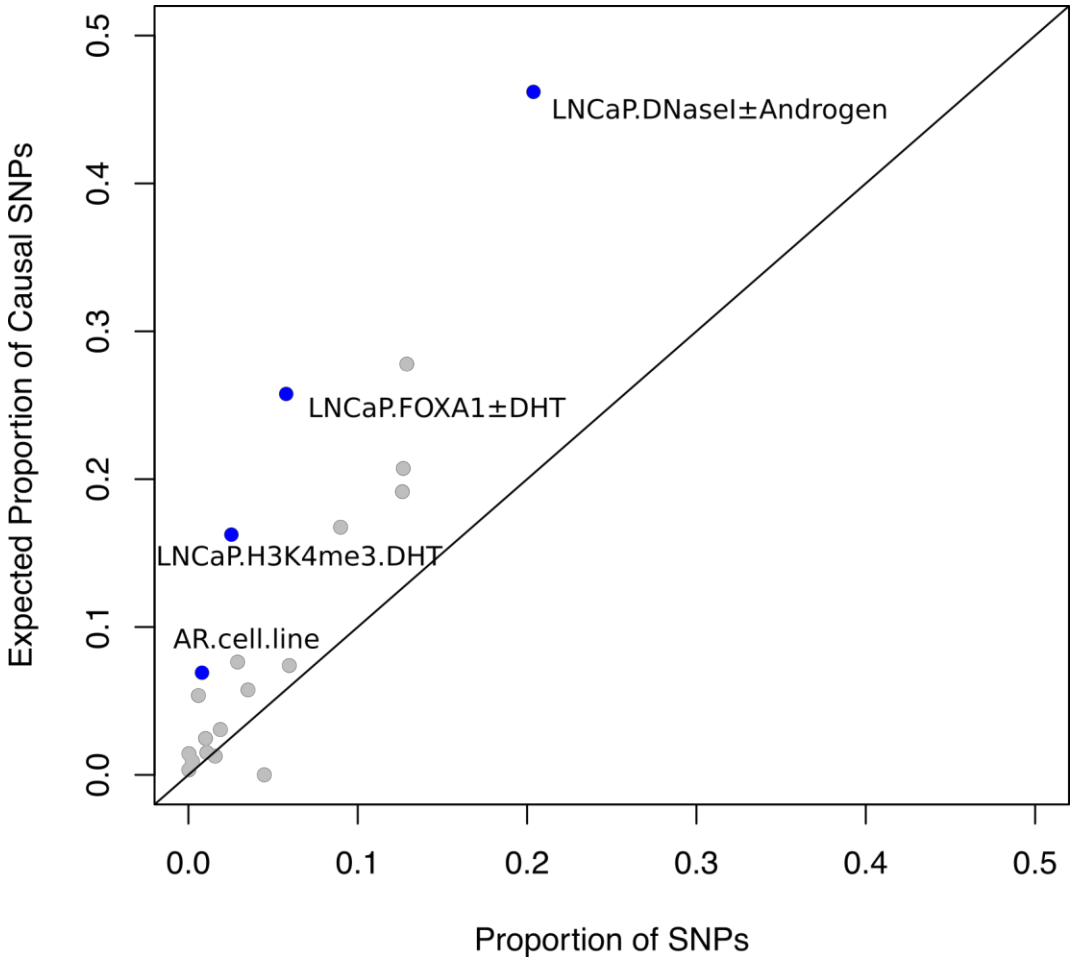
Supplementary Figure 16. QQ-plot for association to prostate cancer –log P-values for each ancestry group.



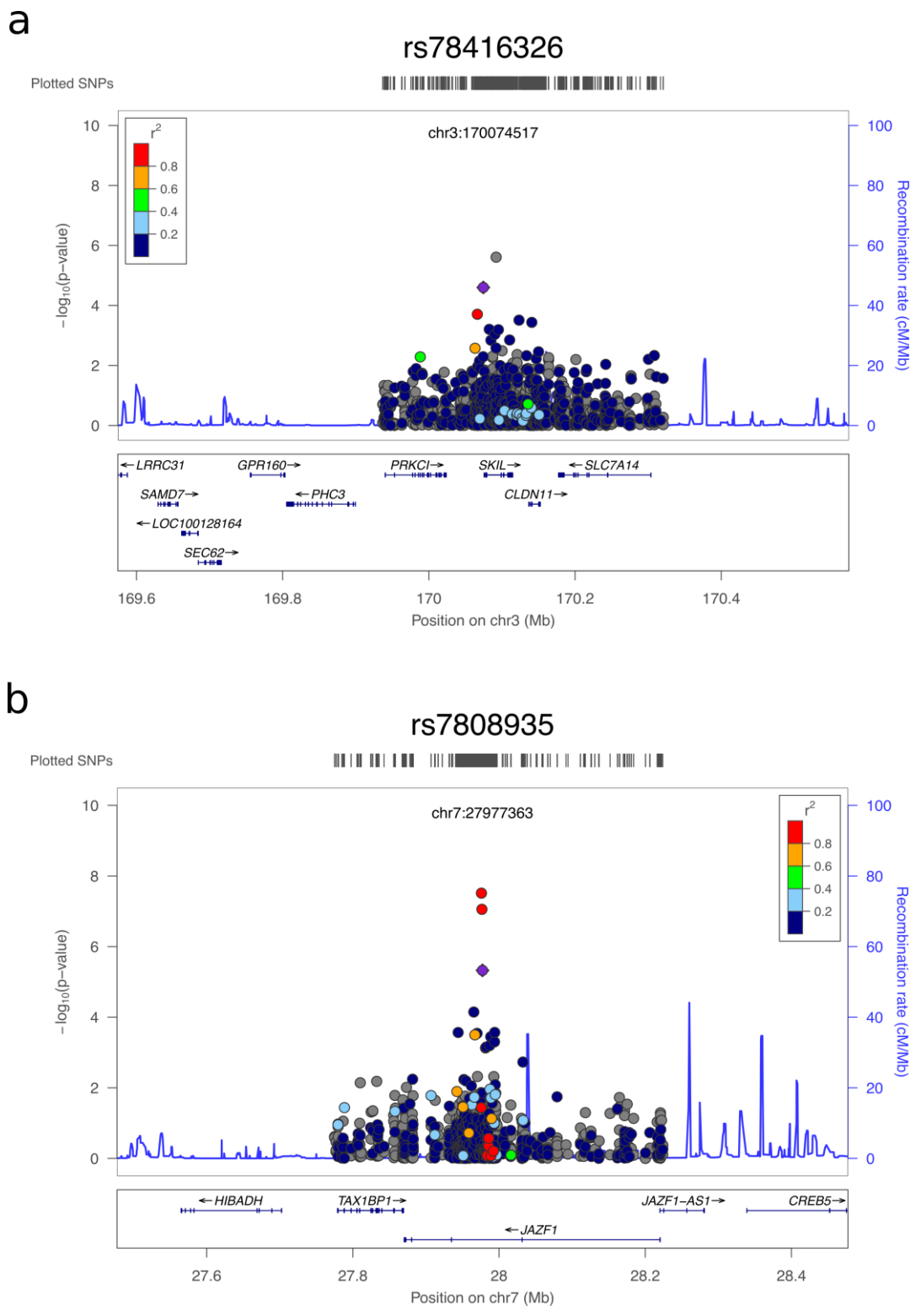
Supplementary Figure 17. QQ-plot for –log P-values for association to prostate cancer when conditioned on the index SNPs for each ancestry Group.



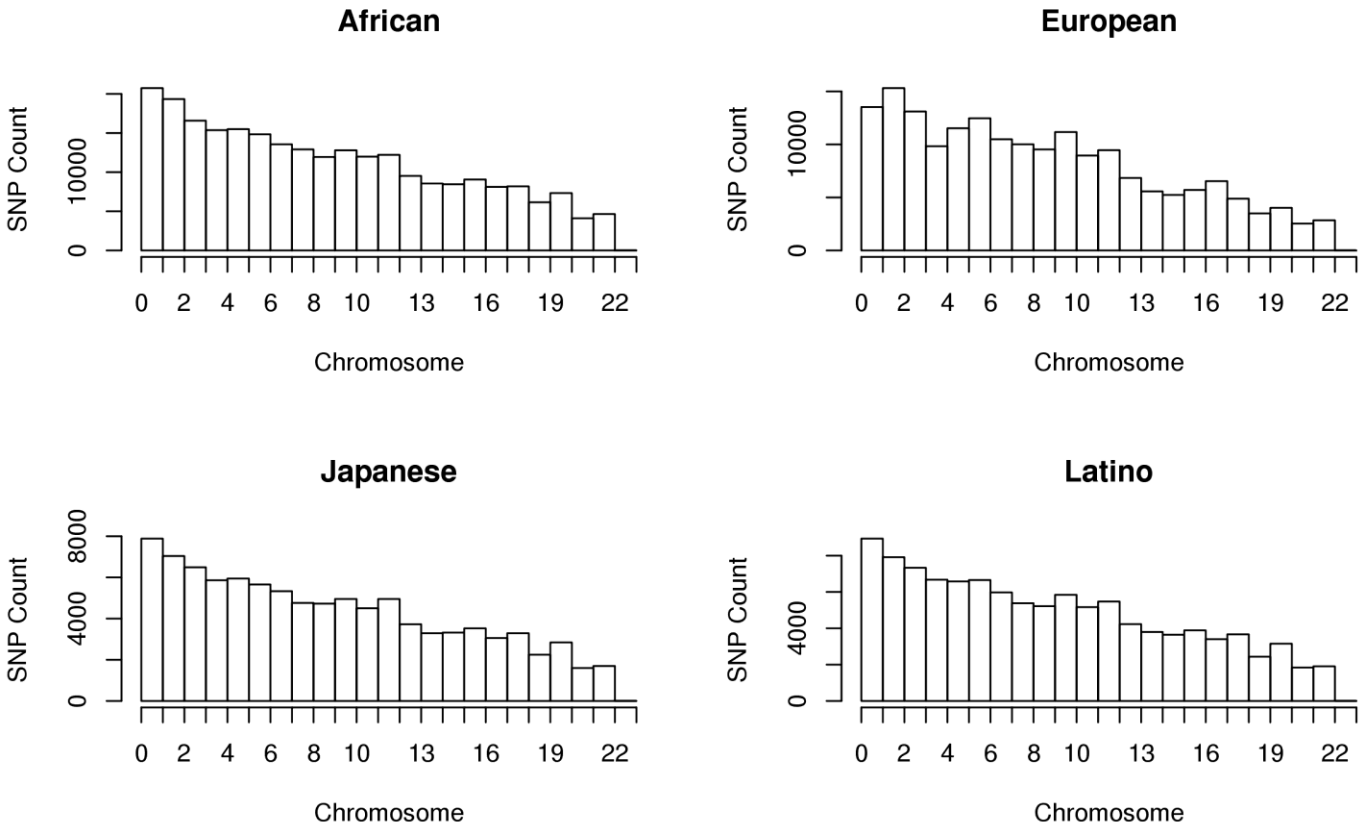
Supplementary Figure 18. QQ-plot of the association statistics for the African ancestry group with 95% confidence intervals estimated from simulations. We simulated phenotypes starting from the empirical genotypes assuming the same contribution of rare and common variants as in the real phenotype data (i.e. $h^2_{g,rare} = 0.12$, $h^2_{g,common} = 0.17$) over 1,000 causal variants. For each simulation, we binned SNPs based on their expected –log10 p-values and estimated the empirical 95% interval for that bin using the [2.5%, 97.5%] percentiles across 1,000 simulations.



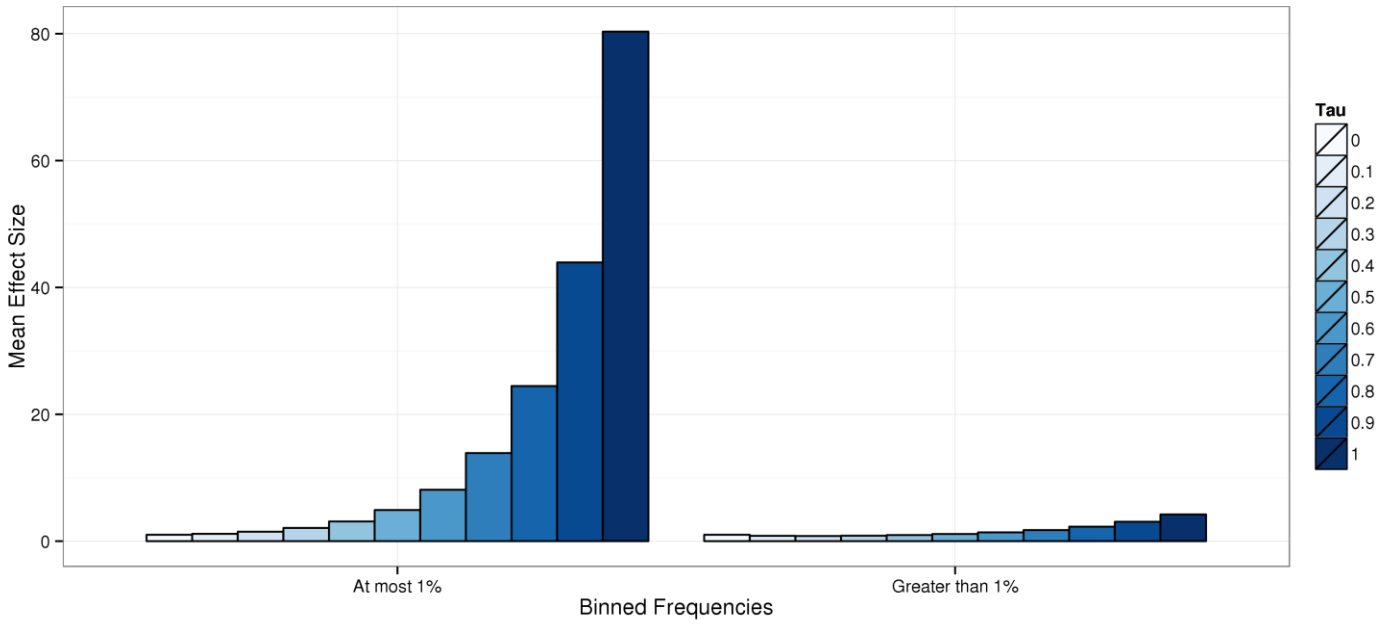
Supplementary Figure 19. Functional enrichment for 20 functional annotations previously involved in PrCa. The Each point corresponds to a functional category, the x-axis shows the proportion of SNPs within the category and the y-axis denotes the proportion of causal SNPs within that annotation as estimated by PAINTOR. The color indicates marginally significant enrichment ($p<0.05$).



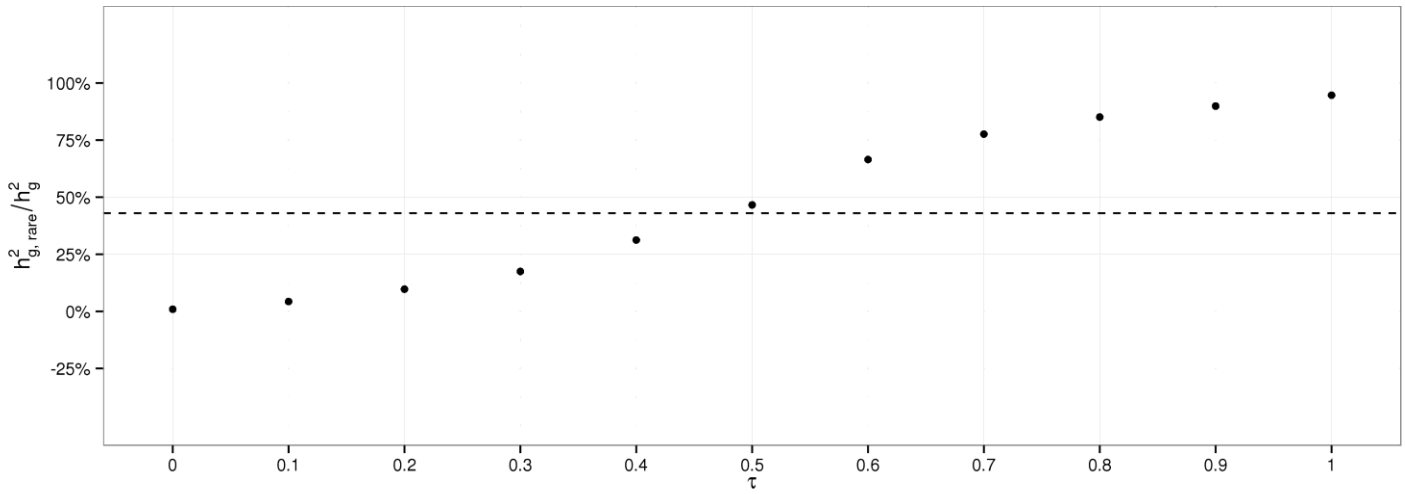
Supplementary Figure 20. Example fine-mapping results using PAINTOR. Candidate sites a) rs78416326 (chr3:170074517, p-value=8.18E-06) and b) rs10486567(chr7:27976563, p-value=2.87E-07) both exhibited posterior probabilities > 0.99 to be causal.



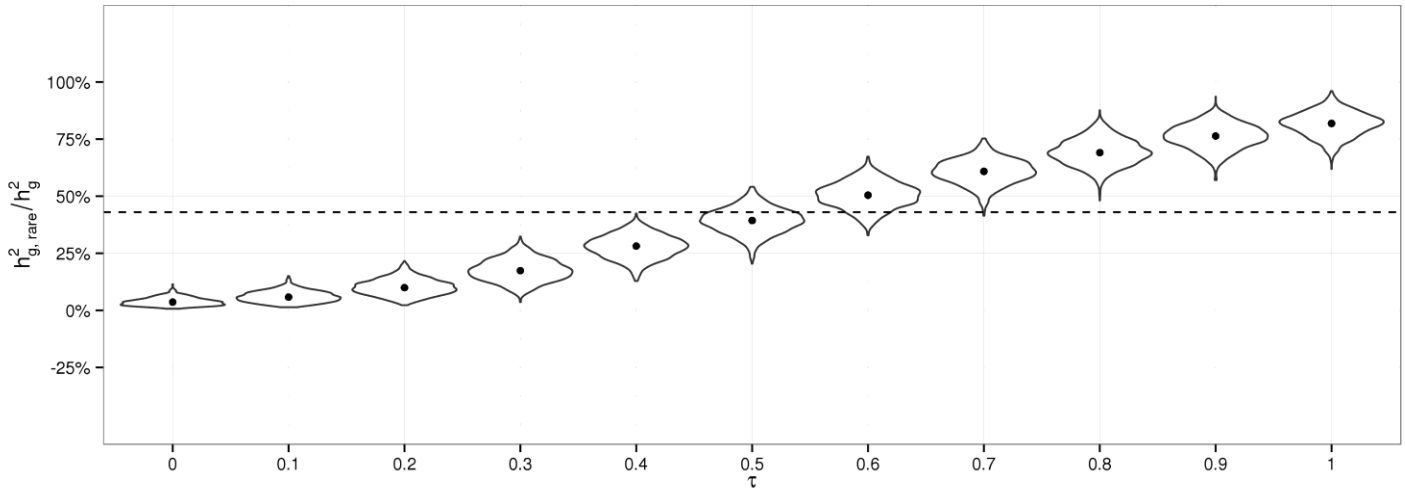
Supplementary Figure 21. Histogram of the arrayed SNPs across chromosomes for each ancestry group.



Supplementary Figure 22. Influence of τ on inflating mean allelic effect size as a function of MAF for the simulated African phenotypes. Results were partitioned by different values of τ. Effects were averaged over 1,000 simulations.



Supplementary Figure 23. Influence of τ on proportion of rare to total narrow-sense heritability ($h^2_{g,rare}$ / h^2_g) for simulated heritability estimates for African dosage data assuming all 122,671 SNPs are causal. Effects were simulated using the Eyre-Walker model linking selection with strength of the effect. As in other simulations, initial h^2_g values were set to 30% when simulating effects. The dashed line represents the observed proportion of rare h^2_g in the empirical African data (~43%).



Supplementary Figure 24. Influence of τ on proportion of rare to total narrow-sense heritability ($h^2_{g,rare}$ / h^2_g) for simulated h^2_g in African dosage data. In order to simulate effects, 10k causal SNPs were selected at random. Effects were simulated using the Eyre-Walker model coupling selection with strength of the effect. Initial h^2_g values were set to 30% when simulating effects. The dashed line represents the observed proportion of rare h^2_g in the empirical African data (~43%).

Supplementary Tables

Locus	Targeted Region (hg19)	Region Length	Total Targeted Bases	Covered Bases	Proportion Covered	Sites Discovered	Mean Sample Coverage (std)	
1q32	204398938-204632481	233544	257048	205586	0.80	3177	10.19	(2.96)
2p24	20877537-20893246	15710	27403	30183	1.10	257	12.89	(3.79)
2p21	43448658-43630459	181802	194739	157540	0.81	2480	9.74	(2.80)
2p15	62802407-63428571	626165	626166	454560	0.73	5998	9.54	(2.86)
2p11	85731890-85884384	152495	261035	192380	0.74	1729	7.89	(2.29)
2q31	173296624-173404804	108181	119562	96325	0.81	1584	9.15	(2.63)
2q37	238353513-238454154	100642	137516	117997	0.86	1822	8.17	(2.34)
2q37	242072015-242153671	81657	127662	106745	0.84	1168	6.41	(1.83)
3p12	87087944-87173324	85381	101936	83902	0.82	1179	8.27	(2.46)
3p11	87408634-87587508	178875	268894	201074	0.75	2720	10.84	(3.27)
3q13	106926122-106982317	56196	158498	119766	0.76	750	13.34	(4.00)
3q13	112948559-113318460	369902	467873	357035	0.76	4803	10.39	(3.08)
3q21	127715196-128138129	422934	464069	401248	0.86	5189	9.58	(2.76)
3q23	141030543-141338641	308099	333558	279486	0.84	3960	11.96	(3.55)
3q26	170042811-170160493	117683	160749	106689	0.66	1344	8.59	(2.57)
4q13	73856466-74548474	692009	692046	525840	0.76	8216	10.08	(3.02)
4q22	95389193-95597814	208622	216527	178147	0.82	2596	9.13	(2.68)
4q24	106042948-106212562	169615	171944	139028	0.81	1942	9.88	(2.91)
5p15	1236690-1359938	123249	174159	150204	0.86	2966	14.31	(4.29)
5p15	1887982-1900829	12848	19294	18960	0.98	245	10.30	(3.04)
5q12	44130776-44396015	265240	356140	281819	0.79	3464	10.77	(3.24)
6p21	30932068-31312326	380259	388800	7854	0.02	18	0.04	(0.01)
6p21	32030284-32339076	308793	351837	37962	0.11	3	0.11	(0.03)
6p21	41504126-41573871	69746	94113	92582	0.98	1029	11.01	(3.32)
6q21	109276321-109500213	223893	309560	236999	0.77	2736	11.00	(3.18)
6q22	117081588-117243260	161673	174271	135939	0.78	1871	8.78	(2.59)
6q25	153363869-153504165	140297	212793	159582	0.75	1815	11.10	(3.30)
6q25	160705167-160980330	275164	283734	216170	0.76	4010	11.66	(3.39)
7p15	20968341-21051042	82702	244859	203996	0.83	1342	13.29	(3.99)
7p15	27940608-27996570	55963	79584	78613	0.99	831	12.91	(3.78)
7q21	97695363-97853184	157822	169861	132700	0.78	2047	8.66	(2.50)
8p21	23417091-23478809	61719	67336	57772	0.86	1081	11.41	(3.29)
8p21	23482464-23549423	66960	80222	65148	0.81	1081	9.26	(2.68)
8p21	25875310-25935389	60080	214157	180141	0.84	1008	9.93	(2.89)
8q24	127730818-128540818	810001	827112	654606	0.79	12403	11.25	(3.27)
10q11	51457562-51561799	104238	111577	67363	0.60	962	5.74	(1.63)
10q24	104239100-104537034	297935	336612	236926	0.70	2883	7.93	(2.29)
10q26	122805547-122888555	83009	173390	143988	0.83	1290	12.20	(3.56)
10q26	126681688-126734419	52732	75374	76260	1.01	1042	9.12	(2.63)
11p15	2213166-2238574	25409	46917	47327	1.01	730	5.14	(1.48)
11q12	58667154-59037217	370064	406674	204823	0.50	3117	7.88	(2.31)
11q13	68810837-69041847	231011	253782	216792	0.85	4023	7.90	(2.24)
11q22	102387691-102418118	30428	169378	123639	0.73	423	13.05	(3.90)

12q13	49366890-49937813	570924	570925	369720	0.65	5421	7.43	(2.15)
12q13	53228880-53364233	135354	317634	241121	0.76	1749	6.34	(1.81)
12q21	82011500-82337901	326402	473879	365640	0.77	4016	10.55	(3.24)
12q24	114608145-114692626	84482	175880	161397	0.92	1622	15.57	(4.58)
13q22	73706017-73749745	43729	65341	55276	0.85	661	9.80	(2.84)
14q22	53289808-53420350	130543	351515	251002	0.71	1824	10.95	(3.27)
14q24	69074684-69136881	62198	180971	157972	0.87	959	10.49	(3.01)
16p13	4285078-4353057	67980	198984	139153	0.70	686	4.97	(1.42)
17p13	464832-700020	235189	341885	228010	0.67	3461	7.31	(2.09)
17q12	36068728-36084261	15534	45129	49284	1.09	238	13.51	(3.98)
17q12	36086689-36122227	35539	38118	31646	0.83	480	7.97	(2.29)
17q21	47318157-47477146	158990	302166	207698	0.69	1779	7.99	(2.35)
17q24	69077078-69242932	165855	177942	146251	0.82	2676	9.19	(2.69)
18q23	76756673-76808210	51538	174489	139537	0.80	722	7.94	(2.35)
19q13	38708970-38903099	194130	219780	152746	0.69	2239	6.24	(1.78)
19q13	41949474-42063675	114202	317526	204618	0.64	1510	5.02	(1.44)
19q13	51353402-51379893	26492	75209	71831	0.96	484	12.37	(3.47)
20q13	62225094-62401143	176050	311452	229944	0.74	2267	4.45	(1.25)
22q13	40389007-40527858	138852	368560	292566	0.79	1718	9.61	(2.78)
22q13	43494731-43505695	10965	37229	37812	1.02	104	5.69	(1.68)
Summary		11305695	15153375	11114920	0.78	2189.68	9.27	(2.72)

Supplementary Table 2. Coverage statistics for each of the 63 targeted regions.

Ancestry	Number of SNPs	Proportion not in 1KG
African	122,743	0.28
European	86,828	0.39
Japanese	69,915	0.41
Latino	92,373	0.27
Meta	196,786	0.44

Supplementary Table 3. SNPs identified through sequencing. A SNP was considered to be "not found" within the 1KG dataset if either the position was not called, or if the alternative allele frequency was estimated to be 0.0 for each respective ancestry group.

Ancestry	Sample Size	h_g^2 (SE)		$h_{g,rare}^2$ (SE)		$h_{g,common}^2$ (SE)	
African	4006	0.31	(0.04)	0.12	(0.05)	0.17	(0.03)
European	1753	0.34	(0.07)	0.00	(0.06)	0.27	(0.06)
Japanese	1770	0.23	(0.06)	0.05	(0.07)	0.13	(0.04)
Latino	1708	0.18	(0.06)	0.00	(0.06)	0.14	(0.05)

Supplementary Table 4. Constrained AI-REML estimates of h_g^2 (standard errors) for a single component over all SNPs compared to the estimates partitioned by MAF. The point-estimates are slightly biased upward for the single-component results; however, they fall within the standard error levels based on the sum of rare and common components. GRMs were estimated using dosage data in GCTA.

	# Samples	Rare ($\alpha = 0.0$)			Common ($1 - \alpha = 1.0$)		
		Mean $h^2_{g,rare}$	SD $h^2_{g,rare}$	GCTA SE	Mean $h^2_{g,common}$	SD $h^2_{g,common}$	GCTA SE
EM-REML	4006	0.032	0.016	0.031	0.291	0.023	0.027
AI-REML constrained	4006	0.012	0.018	0.030	0.290	0.024	0.027
AI-REML	4006	0.002	0.029	0.029	0.296	0.024	0.027

Supplementary Table 5. Summarized heritability estimates on 1,000 simulated phenotypes from African ancestry genotype data with $\alpha = 0.0\%$ of phenotypic variation due to rare alleles. Joint REML estimation was performed with EM, constrained AI, and standard AI algorithms for rare and common components given each simulated phenotype. We report the mean SNP-heritability attributable to each component along with the sample standard deviation and mean estimated standard error (GCTA SE).

	# Samples	Rare ($\alpha = 0.43$)			Common ($1 - \alpha = 0.57$)		
		Mean $h^2_{g,rare}$	SD $h^2_{g,rare}$	GCTA SE	Mean $h^2_{g,common}$	SD $h^2_{g,common}$	GCTA SE
EM-REML	4006	0.131	0.034	0.035	0.168	0.024	0.024
AI-REML constrained	4006	0.126	0.036	0.035	0.168	0.024	0.024
AI-REML	4006	0.126	0.036	0.035	0.168	0.024	0.024

Supplementary Table 6. Summarized heritability estimates on 1,000 simulated phenotypes from African ancestry genotype data with $\alpha = 43\%$ of phenotypic variation due to rare alleles ($h^2_{g,rare} = 0.13$). Joint REML estimation was performed with EM, constrained AI, and standard AI algorithms for rare and common components given each simulated phenotype. We report the mean SNP-heritability attributable to each component along with the sample standard deviation and mean estimated standard error (GCTA SE).

Ancestry	Sample Size	$h^2_{g,rare}$ (SE)		$h^2_{g,common}$ (SE)	
African	4006	0.14	(0.06)	0.38	(0.04)
European	1753	0.08	(0.09)	0.17	(0.09)
Japanese	1770	0.01	(0.08)	0.03	(0.07)
Latino	1708	0.06	(0.09)	0.27	(0.09)

Supplementary Table 7. Constrained AI-REML estimates of rare and common SNP-heritability for each ancestry group using GRMs adjusted for linkage disequilibrium. LD adjustment was computed using LDAK using best-guess genotype calls from sequencing data.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	3476	58699	63972	0.13	0.05	0.18	0.03
W/o Diff Miss	0.5	3073	54995	58366	0.13	0.06	0.17	0.04
W/o Diff Miss	0.25	3066	54995	58366	0.12	0.06	0.17	0.04
W/o Diff Miss	0.1	2982	54995	58366	0.11	0.06	0.16	0.04
W/o Diff Miss	0.075	2909	54995	58366	0.11	0.06	0.16	0.04
W/o Diff Miss	0.05	2762	54995	58366	0.13	0.06	0.16	0.04
W/o Diff Miss	0.025	2073	54995	58366	0.21	0.09	0.22	0.05

Supplementary Table 8. Constrained AI-REML SNP-heritability estimates at various quality control criteria from genotype dosage calls for the African American ancestry group. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. The large drop in sample size from initial set relatedness threshold is the result of only having n=3078 samples with array data. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 in the hard calls before imputation ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	3476	55963	66706	0.11	0.05	0.20	0.04
W/o Diff Miss	0.5	3073	52510	60849	0.12	0.06	0.19	0.04
W/o Diff Miss	0.25	3066	52510	60849	0.12	0.06	0.19	0.04
W/o Diff Miss	0.1	2982	52510	60849	0.10	0.06	0.19	0.04
W/o Diff Miss	0.075	2909	52510	60849	0.09	0.06	0.19	0.04
W/o Diff Miss	0.05	2762	52510	60849	0.11	0.06	0.19	0.04
W/o Diff Miss	0.025	2073	52510	60849	0.19	0.09	0.25	0.06

Supplementary Table 9. Constrained AI-REML SNP-heritability estimates at various quality control criteria from genotype hard calls prior to imputation for the African American ancestry group. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. The large drop in sample size from initial set relatedness threshold is the result of only having n=3078 samples with array data. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Rare Relatedness						
Ancestry	Sample Size	(0,0.05]	(0.05,0.25]	(0.25,0.5]	(0.5,1]	> 0.5
African	4006	7,953,656	67,530	817	12	-
European	1753	1,525,070	9,818	657	82	1
Japanese	1770	1,557,485	7,711	335	13	21
Latino	1708	1,443,409	13,863	440	37	29
Common Relatedness						
Ancestry	Sample Size	(0,0.05]	(0.05,0.25]	(0.25,0.5]	(0.5,1]	> 0.5
African	4006	7,695,080	326,930	5	-	-
European	1753	1,454,989	80,625	14	-	-
Japanese	1770	1,357,348	208,210	7	-	-
Latino	1708	1,302,503	155,273	2	-	-
Array Relatedness						
Ancestry	Sample Size	(0,0.05]	(0.05,0.25]	(0.25,0.5]	(0.5,1]	> 0.5
African	3078	4,716,054	19,449	-	-	-
European	1627	1,322,697	53	1	-	-
Japanese	1674	1,400,103	196	2	-	-
Latino	1642	1,343,538	3,722	1	-	-

Supplementary Table 10. Tally of the off-diagonal entries in GRMs binned by absolute value of relatedness. We see slightly increased relatedness across rare variation when compared against common and arrayed SNPs. The proportion of pairs having relatedness at most 0.05 accounts for greater than 0.99 for rare SNPs, 0.96 for common SNPs, and greater than 0.99 for the genome-wide array SNPs. Rare and common GRMs were estimated from dosage data, while the array-based GRM was estimated using hard calls.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	530	58699	63972	0.00	0.17	0.20	0.15

Supplementary Table 11. Constrained AI-REML SNP-heritability estimates at various quality control criteria from genotype dosage calls for the Ugandan ancestry group. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. Since no Ugandan samples had array results, samples were not filtered on relatedness.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	530	58699	63972	0.01	0.16	0.19	0.15

Supplementary Table 12. Constrained AI-REML SNP-heritability estimates at various quality control criteria from genotype hard calls prior to imputation for the Ugandan ancestry group. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. Since no Ugandan samples had array results, samples were not filtered on relatedness.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	1753	33606	53164	0.00	0.06	0.27	0.06
W/o Diff Miss	0.5	1627	33028	51859	0.00	0.08	0.25	0.06
W/o Diff Miss	0.25	1626	33028	51859	0.00	0.09	0.25	0.06
W/o Diff Miss	0.1	1620	33028	51859	0.00	0.09	0.24	0.06
W/o Diff Miss	0.075	1615	33028	51859	0.00	0.09	0.25	0.06
W/o Diff Miss	0.05	1582	33028	51859	0.00	0.09	0.22	0.06
W/o Diff Miss	0.025	1132	33028	51859	0.00	0.14	0.21	0.07

Supplementary Table 13. Constrained AI-REML SNP-heritability estimates at various quality control criteria from genotype dosage calls for the European ancestry group. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. The large drop in sample size from initial set relatedness threshold is the result of only having n=1627 samples with array data. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 in the hard calls before imputation ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	1753	28950	57820	0.00	0.06	0.28	0.06
W/o Diff Miss	0.5	1627	28487	56409	0.00	0.08	0.26	0.06
W/o Diff Miss	0.25	1626	28487	56409	0.00	0.08	0.26	0.06
W/o Diff Miss	0.1	1620	28487	56409	0.00	0.08	0.26	0.06
W/o Diff Miss	0.075	1615	28487	56409	0.00	0.08	0.27	0.06
W/o Diff Miss	0.05	1582	28487	56409	0.00	0.09	0.24	0.06
W/o Diff Miss	0.025	1132	28487	56409	0.00	0.12	0.27	0.08

Supplementary Table 14. Constrained AI-REML SNP-heritability estimates at various quality control criteria from genotype hard calls prior to imputation for the European ancestry group. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. The large drop in sample size from initial set relatedness threshold is the result of only having n=1627 samples with array data. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	1770	29121	40742	0.05	0.07	0.13	0.04
W/o Diff Miss	0.5	1674	29029	40545	0.06	0.08	0.13	0.04
W/o Diff Miss	0.25	1662	29029	40545	0.06	0.08	0.13	0.04
W/o Diff Miss	0.1	1596	29029	40545	0.07	0.08	0.12	0.04
W/o Diff Miss	0.075	1575	29029	40545	0.06	0.08	0.11	0.04
W/o Diff Miss	0.05	1519	29029	40545	0.08	0.09	0.10	0.04
W/o Diff Miss	0.025	1357	29029	40545	0.09	0.10	0.11	0.05

Supplementary Table 15. Constrained AI-REML SNP-heritability estimates at various quality control criteria from genotype dosage calls for the Japanese ancestry group. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. The large drop in sample size from initial set relatedness threshold is the result of only having n=1674 samples with array data. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 in the hard calls before imputation ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	1770	26413	43450	0.01	0.07	0.14	0.04
W/o Diff Miss	0.5	1674	26339	43233	0.01	0.08	0.15	0.04
W/o Diff Miss	0.25	1662	26339	43233	0.01	0.08	0.14	0.04
W/o Diff Miss	0.1	1596	26339	43233	0.03	0.08	0.13	0.04
W/o Diff Miss	0.075	1575	26339	43233	0.01	0.08	0.13	0.04
W/o Diff Miss	0.05	1519	26339	43233	0.03	0.08	0.12	0.04
W/o Diff Miss	0.025	1357	26339	43233	0.01	0.09	0.14	0.05

Supplementary Table 16. Constrained AI-REML SNP-heritability estimates at various quality control criteria from genotype hard calls prior to imputation for the Japanese ancestry group. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. The large drop in sample size from initial set relatedness threshold is the result of only having n=1674 samples with array data. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	1708	46373	45932	0.00	0.06	0.14	0.05
W/o Diff Miss	0.5	1642	46282	45762	0.00	0.07	0.14	0.05
W/o Diff Miss	0.25	1641	46282	45762	0.00	0.07	0.13	0.05
W/o Diff Miss	0.1	1616	46282	45762	0.00	0.07	0.13	0.05
W/o Diff Miss	0.075	1579	46282	45762	0.00	0.07	0.12	0.05
W/o Diff Miss	0.05	1505	46282	45762	0.00	0.07	0.15	0.05
W/o Diff Miss	0.025	1132	46282	45762	0.00	0.09	0.14	0.06

Supplementary Table 17. Constrained AI-REML SNP-heritability estimates at various quality control criteria from genotype dosage calls for the Latino ancestry group. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. The large drop in sample size from initial set relatedness threshold is the result of only having n=1642 samples with array data. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 in the hard calls before imputation ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	1708	43786	48520	0.00	0.06	0.15	0.05
W/o Diff Miss	0.5	1642	43682	48351	0.00	0.07	0.14	0.05
W/o Diff Miss	0.25	1641	43682	48351	0.00	0.07	0.14	0.05
W/o Diff Miss	0.1	1616	43682	48351	0.00	0.07	0.14	0.05
W/o Diff Miss	0.075	1579	43682	48351	0.00	0.07	0.12	0.05
W/o Diff Miss	0.05	1505	43682	48351	0.00	0.07	0.16	0.06
W/o Diff Miss	0.025	1132	43682	48351	0.00	0.10	0.13	0.07

Supplementary Table 18. Constrained AI-REML SNP-heritability estimates at various quality control criteria from genotype hard calls prior to imputation for the Latino ancestry group. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. The large drop in sample size from initial set relatedness threshold is the result of only having n=1642 samples with array data. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 in the hard calls before imputation ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Coverage	# Rare SNPs	# Common SNPs	$h^2_{g,rare}$ (SE)		$h^2_{g,rare}$ per SNP	$h^2_{g,common}$ (SE)		$h^2_{g,common}$ per SNP
≥ 0	58699	63972	0.12	(0.05)	2.13E-06	0.17	(0.03)	2.71E-06
≥ 2	58427	60734	0.12	(0.05)	2.13E-06	0.18	(0.03)	2.94E-06
≥ 5	49189	48068	0.11	(0.04)	2.27E-06	0.17	(0.03)	3.50E-06
≥ 7	41031	39806	0.08	(0.04)	1.85E-06	0.17	(0.03)	4.16E-06
≥ 10	29817	28776	0.07	(0.04)	2.39E-06	0.14	(0.03)	4.99E-06

Supplementary Table 19. Constrained AI-REML estimates of SNP-heritability for the African ancestry group controlled for coverage. Any SNP that had a sample mean coverage less than the threshold was removed from the GRM estimation. GRMs were estimated using the dosage data.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	# Array	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE	$h^2_{g,array}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	3078	58699	63972	251919	0.13	0.06	0.17	0.04	0.02	0.15
W/o Diff Miss	0.5	3073	54995	58366	251919	0.12	0.06	0.16	0.04	0.02	0.15
W/o Diff Miss	0.25	3066	54995	58366	251919	0.11	0.06	0.16	0.04	0.02	0.16
W/o Diff Miss	0.1	2982	54995	58366	251919	0.09	0.06	0.15	0.04	0.00	0.16
W/o Diff Miss	0.075	2909	54995	58366	251919	0.09	0.06	0.15	0.04	0.00	0.17
W/o Diff Miss	0.05	2762	54995	58366	251919	0.12	0.06	0.16	0.04	0.00	0.18
W/o Diff Miss	0.025	2073	54995	58366	251919	0.19	0.08	0.21	0.05	0.30	0.24

Supplementary Table 20. SNP-heritability estimated from multiple random effects model fitting rare, common and array genetic effects using constrained AI-REML for African Americans with rare, common and array components computed from imputed dosages. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 in the hard calls before imputation ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	# Array	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE	$h^2_{g,array}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	3078	58699	63972	251919	0.12	0.06	0.19	0.04	0.01	0.16
W/o Diff Miss	0.5	3073	54995	58366	251919	0.12	0.06	0.19	0.04	0.00	0.16
W/o Diff Miss	0.25	3066	54995	58366	251919	0.12	0.06	0.19	0.04	0.00	0.16
W/o Diff Miss	0.1	2982	54995	58366	251919	0.08	0.06	0.18	0.04	0.00	0.17
W/o Diff Miss	0.075	2909	54995	58366	251919	0.07	0.06	0.17	0.04	0.00	0.18
W/o Diff Miss	0.05	2762	54995	58366	251919	0.11	0.06	0.19	0.04	0.00	0.18
W/o Diff Miss	0.025	2073	54995	58366	251919	0.18	0.09	0.25	0.06	0.30	0.25

Supplementary Table 21. SNP-heritability estimated from multiple random effects model fitting rare, common and array genetic effects using constrained AI-REML results for African Americans with rare, common and array components computed from hard calls prior to imputation. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	# Array	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE	$h^2_{g,array}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	1627	33606	53164	182983	0.00	0.09	0.25	0.06	0.18	0.16
W/o Diff Miss	0.5	1627	33028	51859	182983	0.00	0.09	0.25	0.06	0.18	0.16
W/o Diff Miss	0.25	1626	33028	51859	182983	0.00	0.09	0.25	0.06	0.17	0.16
W/o Diff Miss	0.1	1620	33028	51859	182983	0.00	0.09	0.25	0.06	0.22	0.17
W/o Diff Miss	0.075	1615	33028	51859	182983	0.00	0.09	0.26	0.06	0.23	0.17
W/o Diff Miss	0.05	1582	33028	51859	182983	0.00	0.09	0.23	0.06	0.20	0.17
W/o Diff Miss	0.025	1132	33028	51859	182983	0.00	0.14	0.25	0.08	0.16	0.23

Supplementary Table 22. SNP-heritability estimated from multiple random effects model fitting rare, common and array genetic effects using constrained AI-REML for Europeans with rare, common and array components computed from imputed dosages. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 in the hard calls before imputation ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	# Array	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE	$h^2_{g,array}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	1627	28950	57820	182983	0.00	0.08	0.28	0.06	0.18	0.17
W/o Diff Miss	0.5	1627	28487	56409	182983	0.00	0.08	0.28	0.06	0.18	0.17
W/o Diff Miss	0.25	1626	28487	56409	182983	0.00	0.08	0.28	0.06	0.17	0.17
W/o Diff Miss	0.1	1620	28487	56409	182983	0.00	0.08	0.28	0.06	0.22	0.17
W/o Diff Miss	0.075	1615	28487	56409	182983	0.00	0.08	0.29	0.06	0.22	0.17
W/o Diff Miss	0.05	1582	28487	56409	182983	0.00	0.09	0.26	0.06	0.19	0.17
W/o Diff Miss	0.025	1132	28487	56409	182983	0.00	0.12	0.30	0.09	0.18	0.24

Supplementary Table 23. SNP-heritability estimated from multiple random effects model fitting rare, common and array genetic effects using constrained AI-REML for Europeans with rare, common and array components computed from hard calls prior to imputation. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	# Array	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE	$h^2_{g,array}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	1674	29121	40742	96711	0.06	0.08	0.13	0.04	0.02	0.14
W/o Diff Miss	0.5	1674	29029	40545	96711	0.06	0.08	0.13	0.04	0.02	0.14
W/o Diff Miss	0.25	1662	29029	40545	96711	0.06	0.08	0.13	0.04	0.00	0.13
W/o Diff Miss	0.1	1596	29029	40545	96711	0.07	0.08	0.12	0.04	0.07	0.17
W/o Diff Miss	0.075	1575	29029	40545	96711	0.06	0.08	0.11	0.04	0.05	0.16
W/o Diff Miss	0.05	1519	29029	40545	96711	0.08	0.09	0.10	0.04	0.07	0.17
W/o Diff Miss	0.025	1357	29029	40545	96711	0.09	0.10	0.11	0.05	0.04	0.17

Supplementary Table 24. SNP-heritability estimated from multiple random effects model fitting rare, common and array genetic effects using constrained AI-REML for Japanese with rare, common and array components computed from imputed dosages. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 in the hard calls before imputation ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	# Array	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE	$h^2_{g,array}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	1674	26413	43450	96711	0.01	0.08	0.15	0.04	0.03	0.14
W/o Diff Miss	0.5	1674	26339	43233	96711	0.01	0.08	0.15	0.04	0.03	0.14
W/o Diff Miss	0.25	1662	26339	43233	96711	0.01	0.08	0.14	0.04	0.01	0.13
W/o Diff Miss	0.1	1596	26339	43233	96711	0.03	0.08	0.14	0.04	0.07	0.17
W/o Diff Miss	0.075	1575	26339	43233	96711	0.01	0.08	0.13	0.04	0.05	0.16
W/o Diff Miss	0.05	1519	26339	43233	96711	0.03	0.08	0.12	0.04	0.07	0.18
W/o Diff Miss	0.025	1357	26339	43233	96711	0.01	0.09	0.14	0.05	0.04	0.17

Supplementary Table 25. SNP-heritability estimated from multiple random effects model fitting rare, common and array genetic effects using constrained AI-REML for Japanese with rare, common and array components computed from hard calls prior to imputation. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	# Array	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE	$h^2_{g,array}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	1642	46373	45932	109118	0.00	0.07	0.09	0.04	0.00	0.19
W/o Diff Miss	0.5	1642	46282	45762	109118	0.00	0.07	0.09	0.04	0.00	0.19
W/o Diff Miss	0.25	1641	46282	45762	109118	0.00	0.07	0.08	0.04	0.00	0.19
W/o Diff Miss	0.1	1616	46282	45762	109118	0.00	0.06	0.01	0.03	0.00	0.20
W/o Diff Miss	0.075	1579	46282	45762	109118			No Convergence			
W/o Diff Miss	0.05	1505	46282	45762	109118			No Convergence			
W/o Diff Miss	0.025	1132	46282	45762	109118	0.00	0.09	0.14	0.06	0.12	0.29

Supplementary Table 26. SNP-heritability estimated from multiple random effects model fitting rare, common and array genetic effects using constrained AI-REML for Latino with rare, common and array components computed from imputed dosages. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 in the hard calls before imputation ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds. Optimization failed for several values by oscillating endlessly between the same several values.

Variant Filtering	Relatedness Threshold	# Samples	# Rare	# Common	# Array	$h^2_{g,rare}$	SE	$h^2_{g,common}$	SE	$h^2_{g,array}$	SE
Initial Set $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001	-	1642	43786	48520	109118	0.00	0.07	0.10	0.05	0.00	0.20
W/o Diff Miss	0.5	1642	43682	48351	109118	0.00	0.07	0.10	0.05	0.00	0.20
W/o Diff Miss	0.25	1641	43682	48351	109118	0.00	0.07	0.09	0.04	0.00	0.20
W/o Diff Miss	0.1	1616	43682	48351	109118	0.00	0.07	0.03	0.04	0.00	0.20
W/o Diff Miss	0.075	1579	43682	48351	109118			No Convergence			
W/o Diff Miss	0.05	1505	43682	48351	109118			No Convergence			
W/o Diff Miss	0.025	1132	43682	48351	109118			No Convergence			

Supplementary Table 27. SNP-heritability estimated from multiple random effects model fitting rare, common and array genetic effects using constrained AI-REML for Latino with rare, common and array components computed from hard calls prior to imputation. The Initial Set refers to the set of variants that passed $\hat{r}^2 \geq 0.6$ and MAF ≥ 0.001. We proceeded to remove the set of variants that were found to have missing-ness correlated with disease status at nominal p-value of 0.01 ("W/o Diff Miss"). Estimates were recomputed using stricter relatedness thresholds. Optimization failed for several values by oscillating endlessly between the same several values.

Ancestry	Sample Size	h_g^2 Index SNPs		$h^2_{g,rare}$ (0.1% ≤ MAF < 1%)		$h^2_{g,common}$ (MAF ≥ 1%)	
African	4,006	0.06	(0.01)	0.12	(0.05)	0.17	(0.03)
Non-Ugandan	3,476	0.07	(0.01)	0.13	(0.05)	0.18	(0.03)
Ugandan	530	0.08	(0.04)	0.00	(0.17)	0.20	(0.15)
African (No 8q24)	4,006	0.04	(0.01)	0.12	(0.04)	0.09	(0.03)
Non-Ugandan	3,476	0.04	(0.01)	0.13	(0.05)	0.09	(0.03)
Ugandan	530	0.03	(0.04)	0.01	(0.16)	0.11	(0.14)
African (non-standardized genotypes in GRM)	4,006	-	-	0.10	(0.03)	0.11	(0.02)

Supplementary Table 28. Constrained AI-REML estimates of h_g^2 (standard errors) using sequencing data when adjusting for possible population stratification from Ugandan ancestry. The analyses were repeated with the 8q24 region removed. Lastly, to account for possible bias in differing underlying effect distributions, REML analysis was performed on non-standardized genotypes for GRM computation.

Ancestry	Sample Size	$h^2_{g,rare}$ (0.1% ≤ MAF < 1%)		$h^2_{g,low-frequency}$ (1% ≤ MAF < 5%)		$h^2_{g,common}$ (MAF ≥ 5%)	
African	4006	0.12	(0.05)	0.10	(0.03)	0.08	(0.02)
European	1753	0.00	(0.06)	0.10	(0.06)	0.14	(0.04)
Japanese	1770	0.06	(0.08)	0.00	(0.04)	0.10	(0.03)
Latino	1708	0.00	(0.06)	0.04	(0.04)	0.08	(0.03)

Supplementary Table 29. Constrained AI-REML estimates of h^2_g (standard errors) with three variance components (corresponding to rare sequenced SNPs, uncommon sequenced SNPs, and common sequenced SNPs).

Ancestry	Sample Size	Jack-knife $h^2_{g,rare}$ (SE)		Jack-knife $h^2_{g,common}$ (SE)	
African	4006	0.13	(0.06)	0.17	(0.04)
European	1753	-0.07	(0.05)	0.31	(0.06)
Japanese	1770	0.05	(0.11)	0.13	(0.05)
Latino	1708	-0.04	(0.07)	0.17	(0.06)

Supplementary Table 30. Leave-one-out jack-knife estimates for rare and common SNP-heritability. Estimates were obtained by dropping each sample once and re-estimating SNP heritability using standard AI-REML. The reported values are the mean SNP-heritability across estimates and jack-knife adjusted standard error given by $SE\left(h_g^2\right) = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}\left(h_{g_i}^2 - h_{g_{jack}}^2\right)^2}$.

Ancestry	Sample Size	$h^2_{g,rare}$ (0.1% ≤ MAF < 1%)		$h^2_{g,common}$ (MAF ≥ 1%)	
Non-Ugandan African	1,753	0.13	(0.06)	0.15	(0.04)

Supplementary Table 31. Mean unconstrained REML estimates of h^2_g (standard errors) for down-sampled sequencing data on the African-ancestry group. The estimates shown are the mean values of estimates across 100 subsamples. Each instance contained 1,753 genotype/phenotype pairs from the 3,476 non-Ugandan Africans.

Ancestry	Sample Size	$h^2_{g,rare}$ (0.1% ≤ MAF < 1%)		$h^2_{g,common}$ (MAF ≥ 1%)	
African	4006	0.12	(0.05)	0.17	(0.03)
European	1753	-0.07	(0.04)	0.31	(0.06)
Japanese	1770	0.05	(0.08)	0.14	(0.04)
Latino	1708	-0.04	(0.05)	0.15	(0.04)

Supplementary Table 32. Standard AI-REML estimates of h^2_g (standard errors) using dosages from sequencing data.

Ancestry	Sample Size	$h^2_{g,rare}$ (0.1% ≤ MAF < 1%)		$h^2_{g,common}$ (MAF ≥ 1%)	
African	4006	0.12	(0.05)	0.17	(0.03)
European	1753	0.00	(0.06)	0.26	(0.05)
Japanese	1770	0.03	(0.07)	0.13	(0.04)
Latino	1708	0.00	(0.06)	0.14	(0.04)

Supplementary Table 33. Constrained AI-REML estimates of h^2_g (standard errors) using best-guess calls from sequencing data.

Ancestry	Sample Size	$h^2_{g,rare}$ (0.1% ≤ MAF < 1%)		$h^2_{g,common}$ (MAF ≥ 1%)	
African	4006	0.11	(0.05)	0.16	(0.03)
European	1753	-0.07	(0.04)	0.30	(0.06)
Japanese	1770	0.03	(0.07)	0.13	(0.04)
Latino	1708	-0.04	(0.05)	0.17	(0.05)

Supplementary Table 34. Standard AI-REML estimates of h^2_g (standard errors) using best-guess genotypes from sequencing data.

Ancestry	Sample Size	$h^2_{g,array}$		Joint Rare + Common				Joint Rare + Common + Array					
				$h^2_{g,rare}$ (0.1% ≤ MAF < 1%)		$h^2_{g,common}$ (MAF ≥ 1%)		$h^2_{g,rare}$ (0.1% ≤ MAF < 1%)		$h^2_{g,common}$ (MAF ≥ 1%)		$h^2_{g,array}$	
African	3078	0.11	(0.16)	0.13	(0.06)	0.17	(0.04)	0.13	(0.06)	0.17	(0.04)	0.02	(0.15)
European	1627	0.27	(0.17)	0.00	(0.08)	0.25	(0.06)	0.00	(0.09)	0.26	(0.06)	0.19	(0.16)
Japanese	1674	0.01	(0.14)	0.06	(0.08)	0.13	(0.04)	0.06	(0.08)	0.13	(0.04)	0.02	(0.14)
Latino	1642	0.00	(0.19)	0.00	(0.07)	0.14	(0.05)	0.00	(0.07)	0.09	(0.04)	0.00	(0.19)

Supplementary Table 35. Constrained AI-REML estimates of h^2_g (standard errors) with three variance components (corresponding to rare sequenced SNPs, common sequenced SNPs, and genotyped array SNPs). Sample sizes differ from earlier estimates, as not all individuals were genotyped utilizing arrays in addition to sequencing. Baseline estimation was performed with rare and common variance components limited to the individuals with array-based SNPs.

Effective Population Size	Sample Size	Mean τ	95% Confidence Interval	
5000	4006	0.49	0.18	0.77
7500	4006	0.46	0.21	0.78
10000	4006	0.48	0.17	0.79
15000	4006	0.49	0.18	0.77

Supplementary Table 36. Sensitivity analysis for effective population size in estimating τ for the African ancestry group under our simulation-based pipeline.

Ancestry	Mean number of SNPs per locus		Mean -$\log_{10}$ p-value Index variant	Mean -$\log_{10}$ p-value at top variant	
	0.1% ≤ MAF < 1%	MAF ≥ 1%		No Conditioning	After Conditioning on Index SNPs
African	1,239	1,353	1.2	3.7	3.4
European	701	1,123	1.1	3.2	2.9
Japanese	606	861	1.2	2.9	2.5
Latino	965	971	0.9	2.9	2.7
Meta-analysis	2,257	1,898	2.5	4.4	3.4

Supplementary Table 38. Average association signal at the fine-mapped loci across the 4 ethnic groups using two thresholds on MAF.

RS ID	Chromosome & Position	Posterior Probability	P Value	Annotations
rs78416326	3:170074517	>0.99	2.51E-05	LNCaP.DNaseI.ENCODE.plusminusAndrogen, LNCaP.H3K4me3.DHT, LNCaP.FOXA1.plusminus.DHT
rs7808935	7:27977363	>0.99	4.72E-06	LNCaP.FOXA1.plusminus.DHT
NA	14:53309727	>0.99	6.51E-05	
rs7642887	3:87172632	0.99	3.41E-05	
rs2005983	22:43432319	0.99	6.08E-03	LNCaP.DNaseI.ENCODE.plusminusAndrogen, LNCaP.FOXA1.plusminus.DHT
NA	1:204462686	0.93	4.35E-03	LNCaP.DNaseI.ENCODE.plusminusAndrogen, LNCaP.H3K4me3.DHT
rs10866527	5:1891800	0.91	2.11E-05	LNCaP.DNaseI.ENCODE.plusminusAndrogen
rs115020225	6:32371731	0.91	8.27E-03	
rs6470499	8:128123394	0.91	1.15E-17	LNCaP.FOXA1.plusminus.DHT

Supplementary Table 43. Sequenced variants with posterior probabilities for causality > 0.90 under a joint model consisting of the top 4 functional categories (those with p-value < 0.05). Each row corresponds to a variant described by its rs ID (if available) in addition to chromosome and position, its posterior probability to be causal, and association-statistic p-value, along with the associated functional annotations in the joint PAINTOR model.

	African		Japanese		Latino		European	
Disease Stage	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**
Localized Disease	1427	0.66	819	0.86	730	0.82	780	0.85
Regional Disease	236	0.11	83	0.09	90	0.10	80	0.09
Metastatic Disease	60	0.03	23	0.02	24	0.03	26	0.03
Unspecified	93	0.04	26	0.03	38	0.04	25	0.03
Missing	336	0.16	3	<0.01	8	0.01	7	0.01
Gleason Grade								
<5	73	0.03	45	0.05	50	0.06	71	0.08
5-7	1081	0.50	442	0.46	560	0.63	497	0.54
≥8	749	0.35	450	0.47	244	0.27	324	0.35
Missing	249	0.12	17	0.02	36	0.04	26	0.03

Supplementary Table 44. Stage and grade of all prostate cancer cases brought forward for QC and analysis split by ancestry.

Ancestry	r Rare	(SE)	p-value	r Common	(SE)	p-value
African	0.07	0.02	2.11E-06	0.15	0.02	2.20E-16
European	-0.02	0.02	4.83E-01	0.18	0.02	1.08E-13
Japanese	-0.02	0.02	5.11E-01	0.11	0.02	5.54E-06
Latino	-0.01	0.02	8.17E-01	0.10	0.02	2.72E-05

Supplementary Table 45. Pearson correlation (r) between total risk score as computed from BLUP estimates with true dichotomous case/control status. Allelic effects were estimated from a single GRM over all SNPs and partitioned when computing risk scores using 10-fold cross-validation.

References

1. Gilmour, A. R., et al. (1995). "Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models." Biometrics 51(4): 1440-1450.

2. Dempster, A. P., et al. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm." Journal of the Royal Statistical Society. Series B (Methodological) 39(1): 1-38.

3. Searle, S. R., et al. (2008). Variance Components, John Wiley & Sons, Inc.

4. Speed, D., et al. "Improved Heritability Estimation from Genome-wide SNPs." The American Journal of Human Genetics 91(6): 1011-1021.

5. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience, 4:7.

6. Dempster E, Lerner I. 1950. Heritability of threshold characters. Genetics 35: 212–236.