The American Journal of Human Genetics, Volume *110*

Supplemental information

Impact of cross-ancestry genetic architecture on GWASs in admixed populations

Rachel Mester, Kangcheng Hou, Yi Ding, Gillian Meeks, Kathryn S. Burch, Arjun Bhattacharya, Brenna M. Henn, and Bogdan Pasaniuc

Supplemental Material

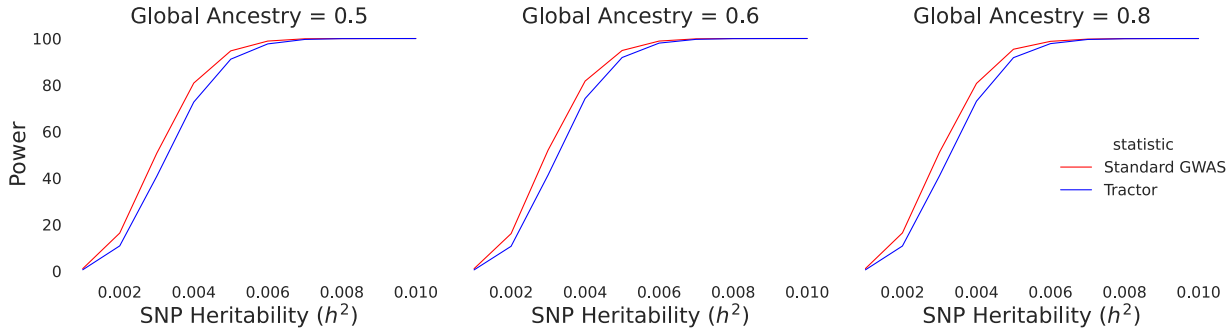

Figure S1: Global ancestry does not have a large impact on power compared to the choice of test statistic and SNP heritability. Power curves of Standard GWAS and Tractor as SNP heritability varies. In this case where neither frequency nor causal effect size vary by local ancestry, Standard GWAS has increased power over Tractor, especially at small levels of SNP heritability. Simulation results of 1,000 replicates with N = 10,000 individuals with causal allele frequency CAF_1 = CAF_2 = 0.5, and causal effect sizes $\beta_1 = \beta_2 = 1.0$. 95% confidence interval too narrow for display.

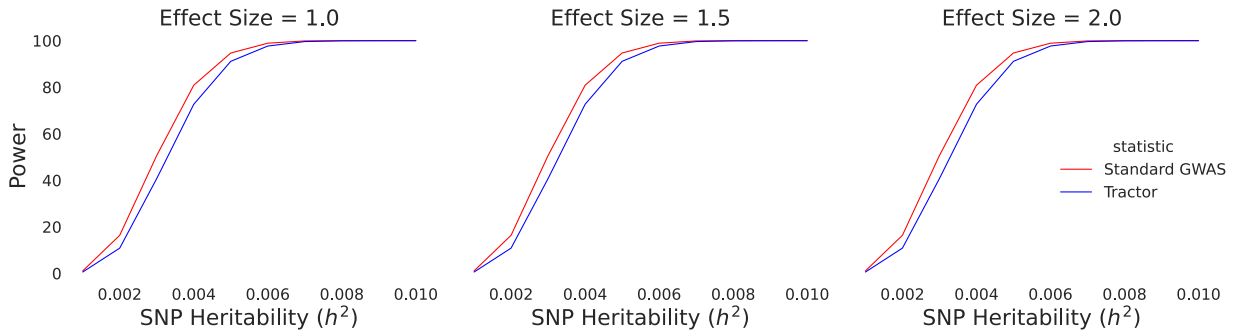

Figure S2: Effect size does not have a large impact on power compared to the choice of test statistic and SNP heritability. Power curves of Standard GWAS and Tractor as SNP heritability varies. In this case where neither frequency nor causal effect size vary by local ancestry, Standard GWAS has increased power over Tractor, especially at small levels of SNP heritability. Simulation results of 1,000 replicates with N = 10,000 individuals with causal allele frequency $CAF_1 = CAF_2 = 0.5$, global ancestry proportions at 50/50, and causal effect sizes $\beta_1 = \beta_2$. 95% confidence interval too narrow for display.

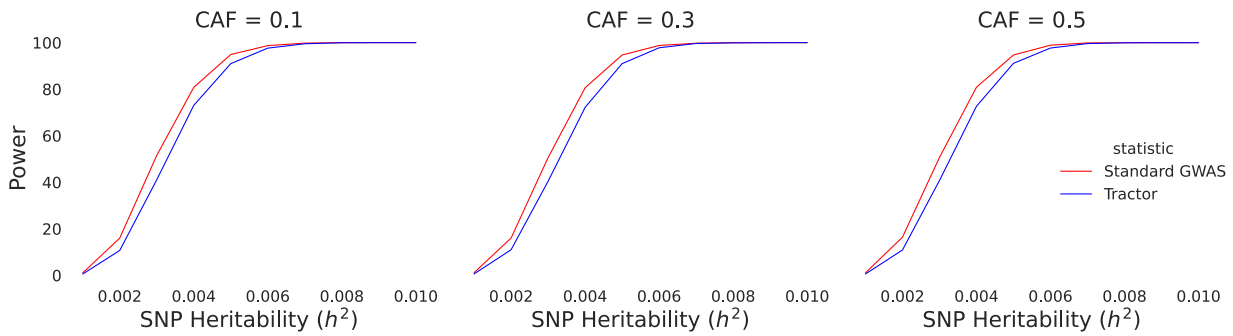

Figure S3: Causal allele frequency does not have a large impact on power compared to the choice of test statistic and SNP heritability. Power curves of Standard GWAS and Tractor as SNP heritability varies. In this case where neither frequency nor causal effect size vary by local ancestry, Standard GWAS has increased power over Tractor, especially at small levels of SNP heritability. Simulation results of 1,000 replicates with N = 10,000 individuals with causal allele frequency $CAF_1 = CAF_2$, global ancestry proportions at 50/50, SNP heritability h^2 = 0.005, and causal effect sizes $\beta_1 = \beta_2 = 1.0$. 95% confidence interval too narrow for display.

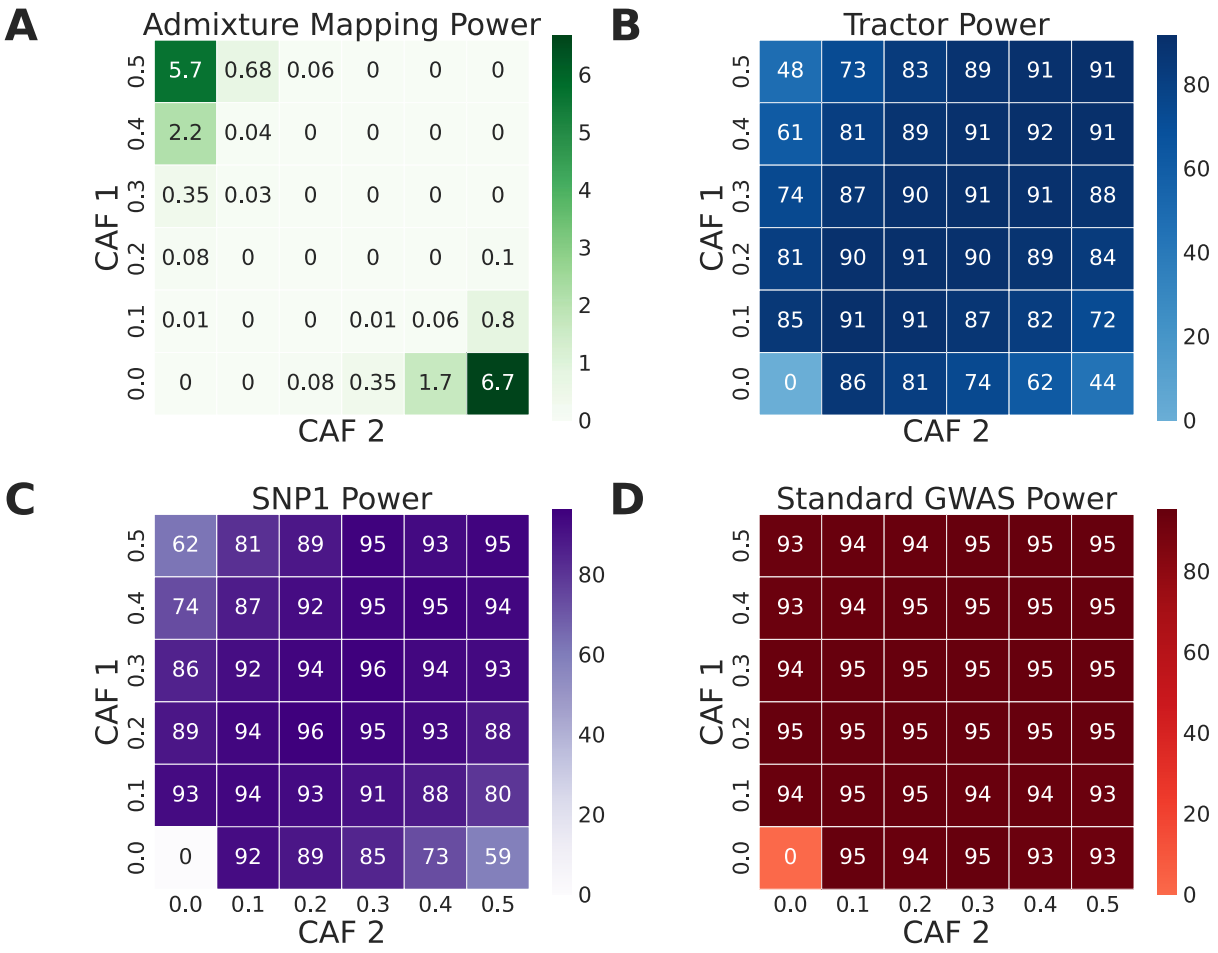

Figure S4: Association statistic power at differing levels of causal allele frequency difference. (a) Admixture mapping has maximum power when causal allele frequency difference by local ancestry is increased. **(b)** Tractor has drastically decreased power when causal allele frequency difference by local ancestry is increased. In this case where causal effect size does not vary by local ancestry, the decrease in Tractor power at high levels of minor allele frequency difference by local ancestry is driven by the increase in power for admixture mapping, which serves as the null hypothesis against which Tractor tests SNP-level effects. **(c)** SNP1 has higher power than Tractor generally but also suffers from drastically decreased power when causal allele frequency difference by local ancestry is increased, likely due to its identical null hypothesis. **(d)** Standard GWAS has slightly decreased power when causal allele frequency difference by local

ancestry is increased. Standard GWAS does not suffer from using ADM as its null hypothesis as Tractor does, but the decrease in power is likely due to increased correlation between global and local ancestry at high levels of allele frequency difference. All panels are simulation results of 1,000 replicates with N = 10,000 individuals with global ancestry proportions at 50/50, SNP heritability h^2 = 0.005, and causal effect sizes $\beta_1 = \beta_2 = 1.0$.

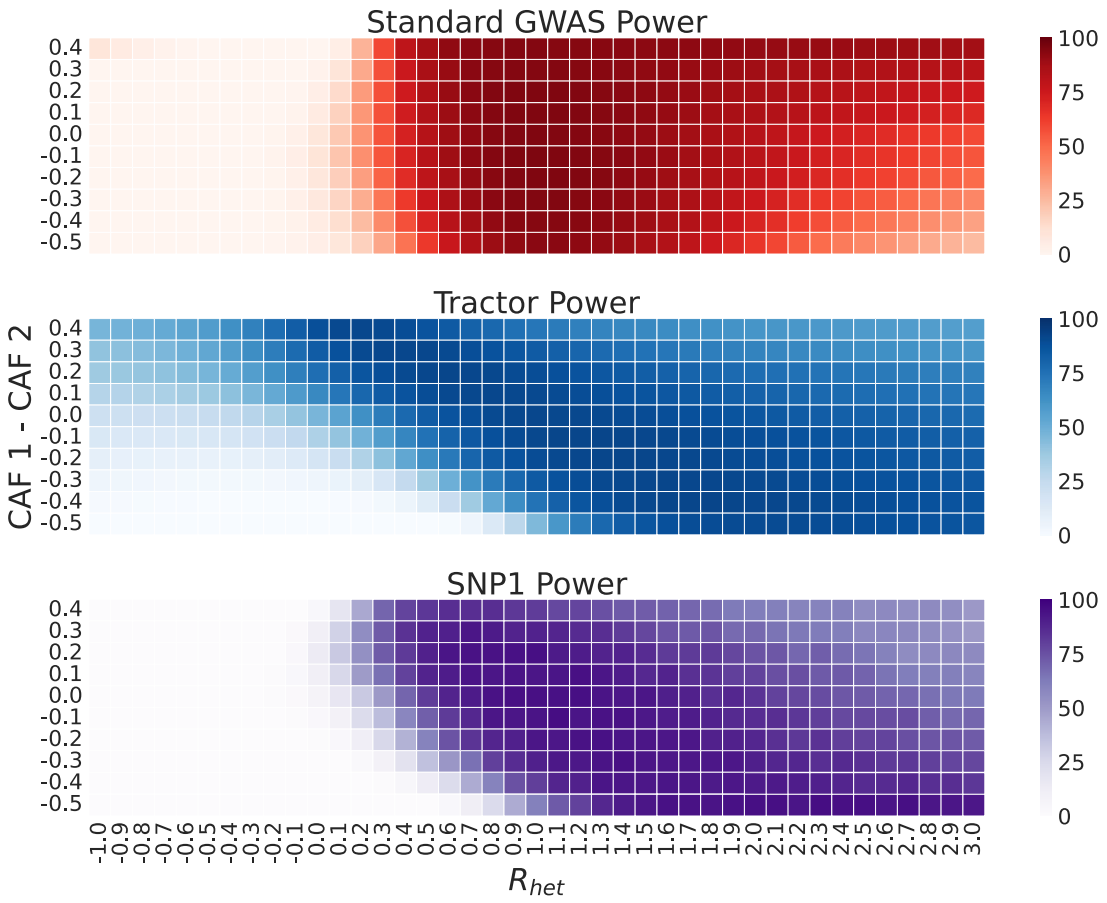

Figure S5: Impact of HetLanc and CAF difference on power of Standard GWAS, Tractor, and SNP1 individually. As HetLanc increases, Standard GWAS power decreases, especially when causal effects are in opposite directions. CAF difference impacts Tractor and SNP1 more drastically than Standard GWAS. Simulation results of 1,000 replicates with N = 10,000 individuals with minor allele frequency CAF_1 = 0.5, global ancestry proportions at 50/50, heritability h^2 = 0.005, and causal effect size β_2 = 1.0.

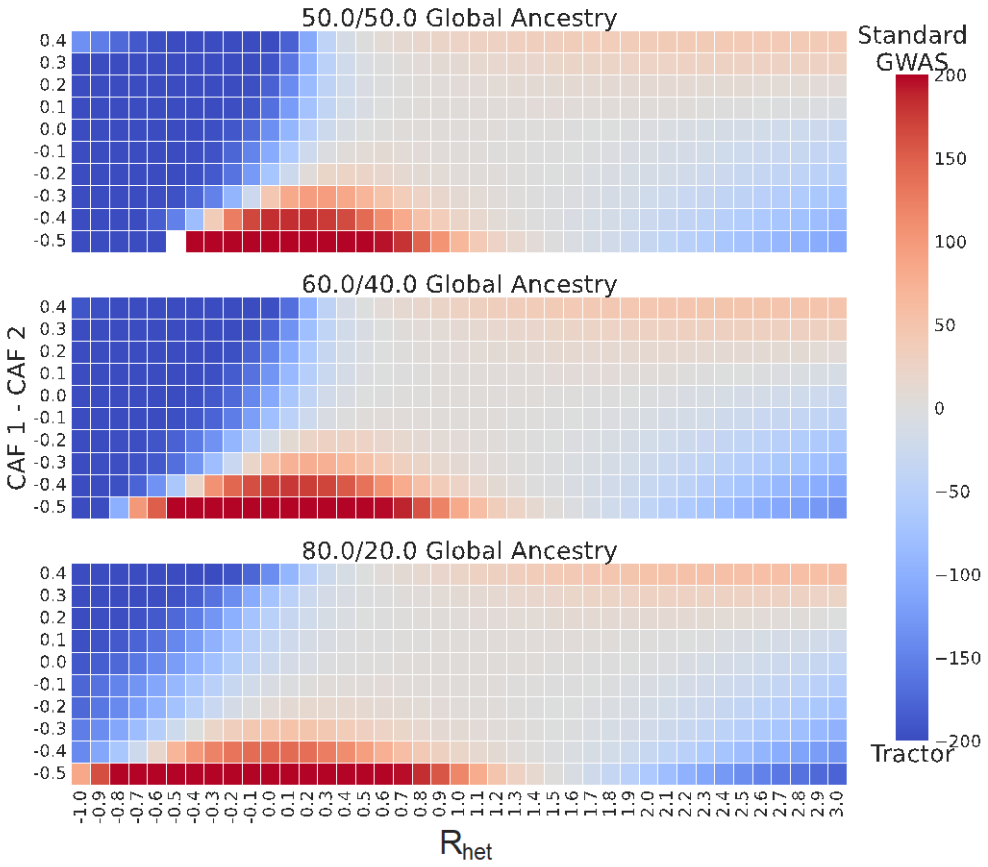

Figure S6: Impact of HetLanc and CAF difference on percent difference in power depends on global ancestry ratios. Heatmap of percent difference in power for Standard GWAS vs Tractor. Red indicates where $Power_{Standard\ GWAS} > Power_{Tractor}$**.** As global ancestry ratios become further from 50%, the range of HetLanc and CAF difference in which Standard GWAS has more power than Tractor increases. Simulation results of 1,000 replicates with N = 10,000 individuals with minor allele frequency CAF_1 **=** 0.5, heritability h^2 = 0.005, and causal effect size $\beta_2 = 1.0$.

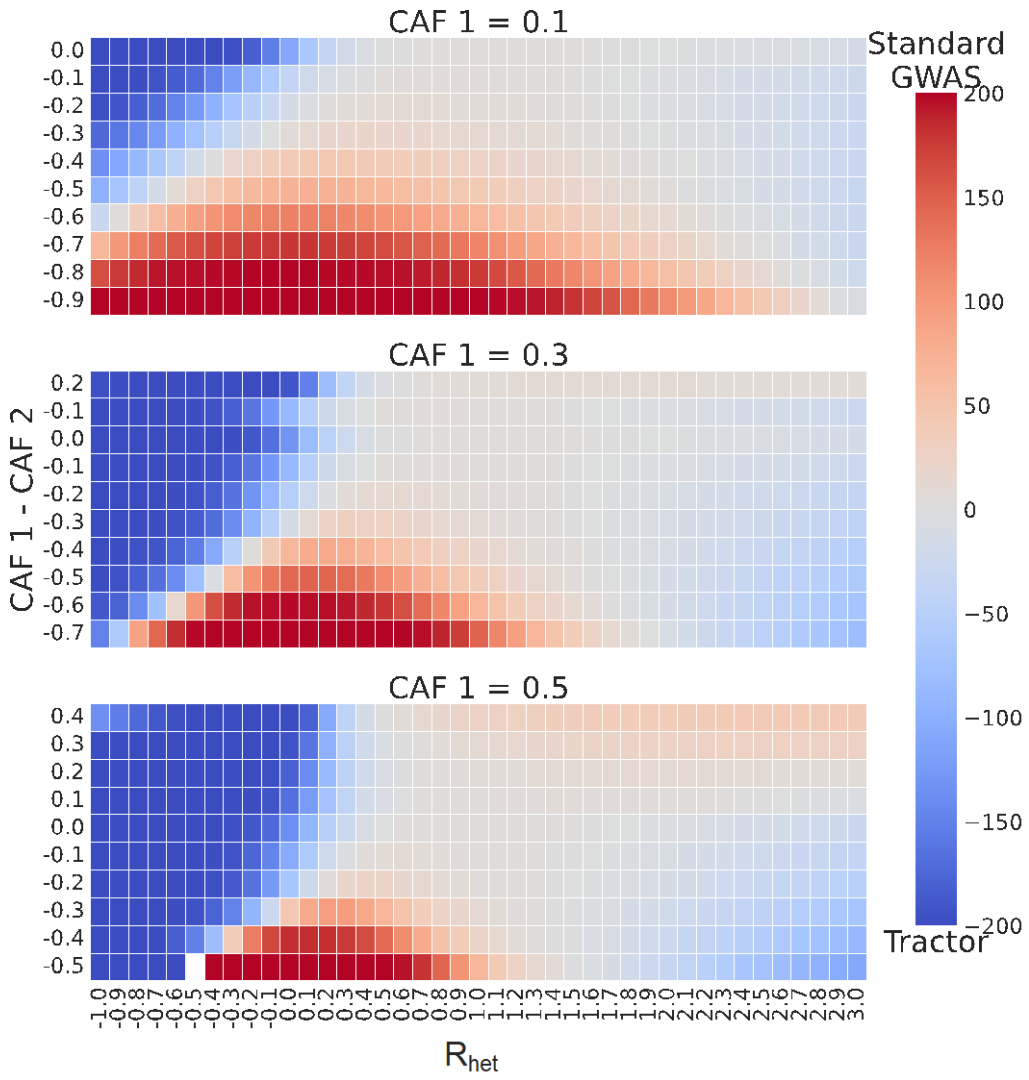

Figure S7: Impact of HetLanc and CAF difference on percent difference in power depends on CAF. Heatmap of percent difference in power for Standard GWAS vs Tractor. Red indicates where $\text{Power}_{\text{Standard GWAS}} > \text{Power}_{\text{Tractor}}$. As CAF becomes further from 0.5, the range of HetLanc and CAF difference in which Standard GWAS has more power than Tractor increases. Simulation results of 1,000 replicates with N = 10,000 individuals with global ancestry proportions at 50/50, SNP heritability h^2 = 0.005, and causal effect size β_2 = 1.0.

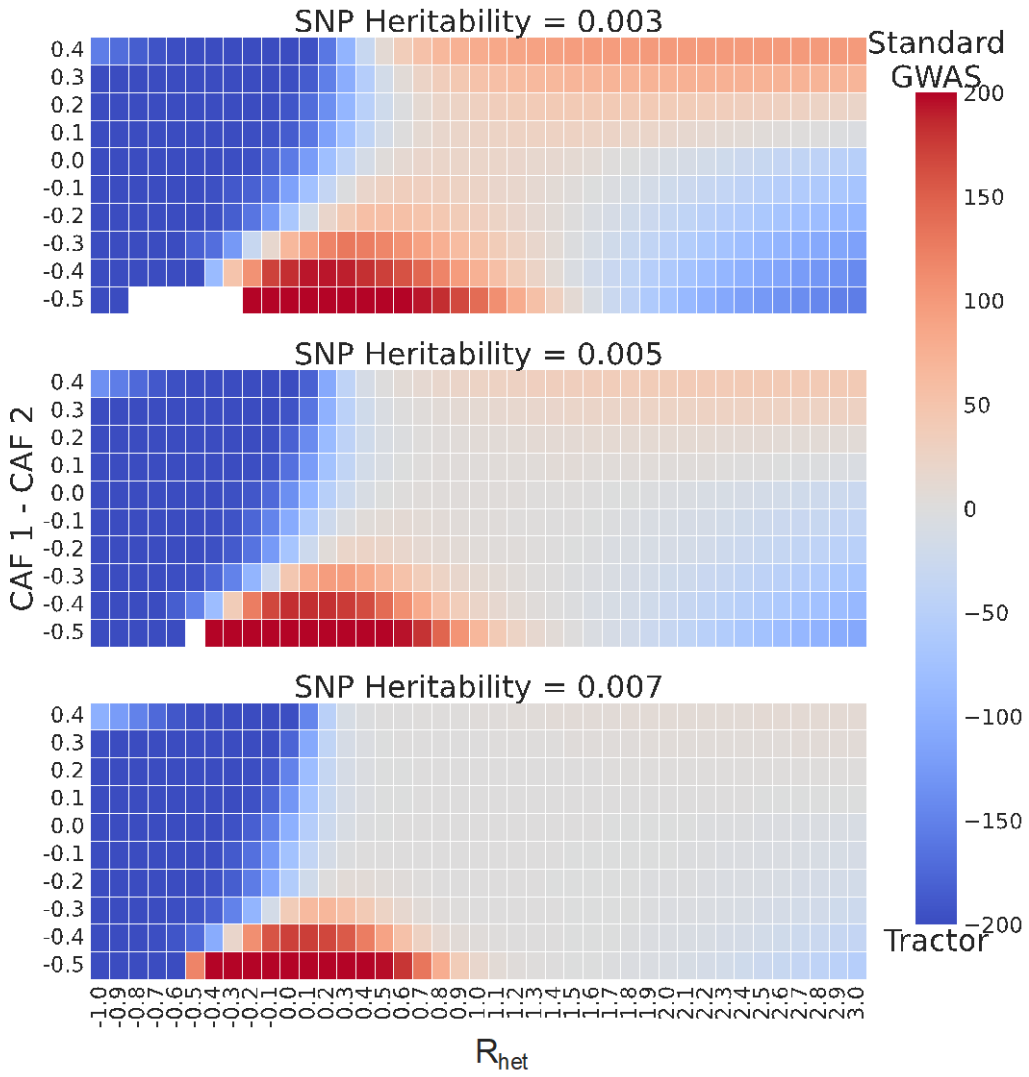

Figure S8: Impact of HetLanc and CAF difference on percent difference in power depends on heritability. Heatmap of percent difference in power for Standard GWAS vs Tractor. Red indicates where $\text{Power}_{\text{Standard GWAS}} > \text{Power}_{\text{Tractor}}$. As heritability decreases, the percent difference in power between Standard GWAS and Tractor increases. Simulation results of 1,000 replicates with N = 10,000 individuals with causal allele frequency $\text{CAF}_1 = 0.5$, global ancestry proportions at 50/50, and causal effect size $\beta_2 = 1.0$.

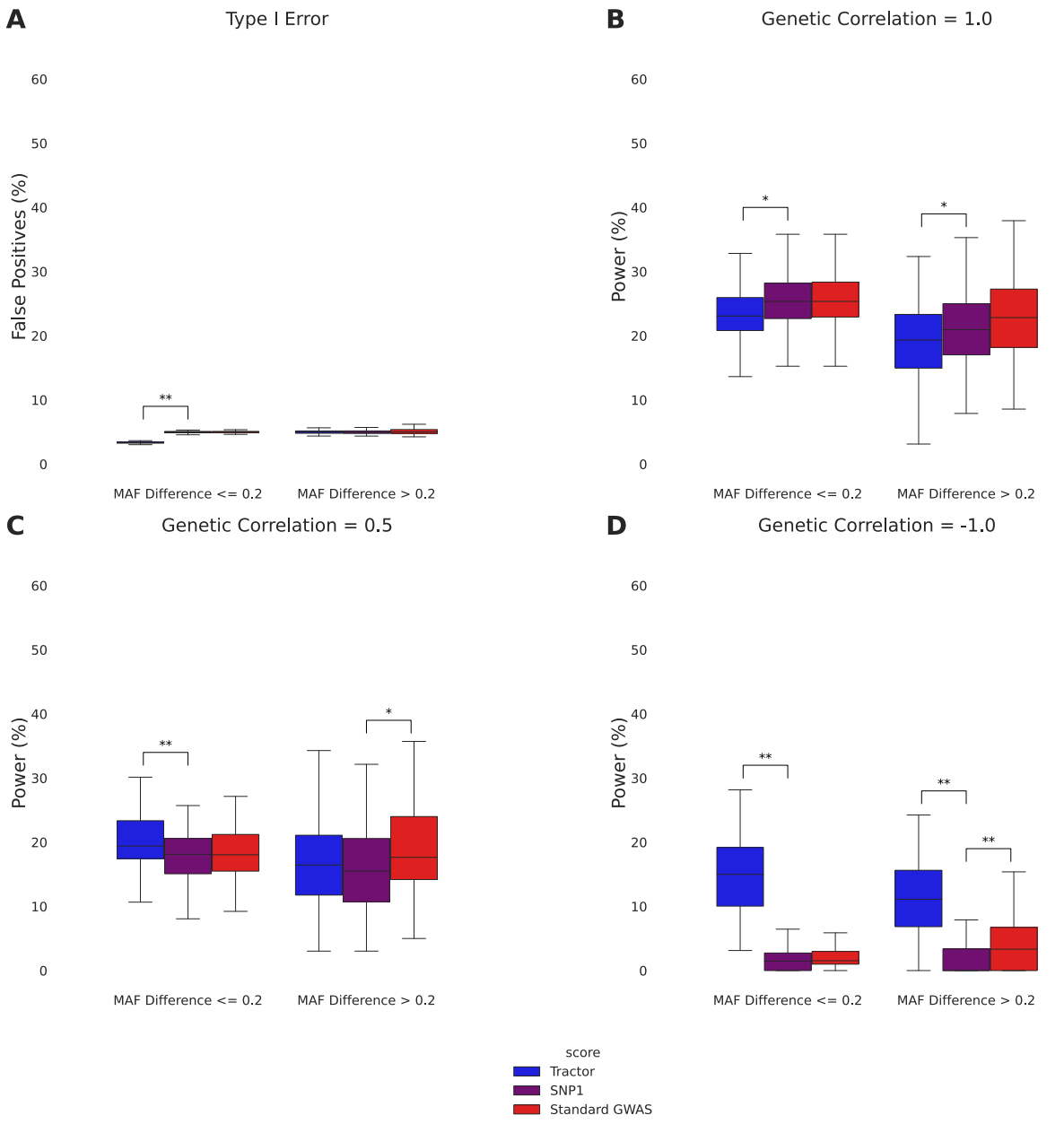

Figure S9: Effect Size Heterogeneity of Tractor, SNP1, and Standard GWAS in the Context of Polygenicity (a) Box plot of Type I error for Tractor, SNP1, and Standard GWAS split by non-differentiated (MAF difference ≤ 0.2) and differentiated (MAF difference > 0.2) SNPs. **(b)** Box plot of power for Tractor, SNP1, and Standard GWAS in the case of no effect size heterogeneity split by non-differentiated and differentiated SNPs. **(c)** Box plot of power for Tractor, SNP1, and Standard GWAS in the case of effect size heterogeneity split by non-differentiated and

differentiated SNPs. **(d)** Box plot of power for Tractor, SNP1 and Standard GWAS in the case of opposite effect sizes split by non-differentiated and differentiated SNPs. **(a-d)** All simulations used real UKBB admixed genotypes and simulated phenotypes with 100 causal SNPs and a total additive genetic heritability of h^2 = 0.5 (see methods). "*" indicates a nominally significant p-value (<0.05). "**" indicates a Bonferroni-corrected significant p-value (<1.28 x 10^{-3}). The boxes show the inter-quartile range while the whiskers show the rest of the distribution (not including outliers).

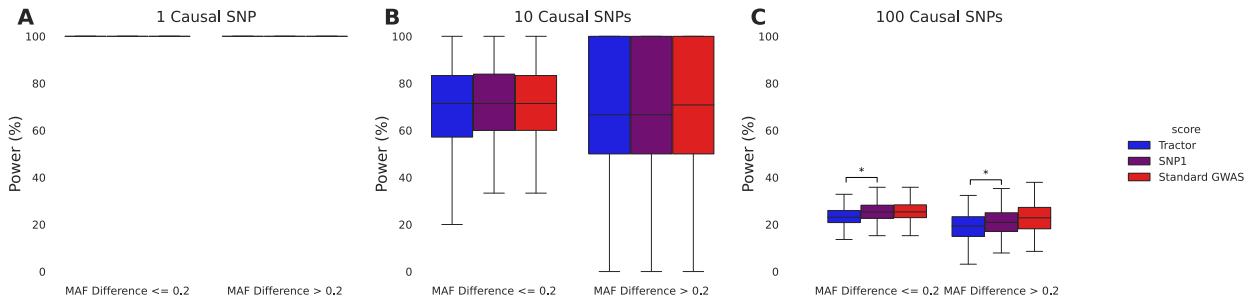

Figure S10: Effect Size Heterogeneity in the Context of Varying Levels of Polygenicity (a) Box plot of power for Tractor, SNP1, and Standard GWAS in the case of one causal SNP split by non-differentiated and differentiated SNPs. All methods had 100% power in this case due to a high SNP heritability of 50%. **(b)** Box plot of power for Tractor, SNP1, and Standard GWAS in the case of 10 causal SNPs split by non-differentiated and differentiated SNPs. **(c)** Box plot of power for Tractor, SNP1, and Standard GWAS in the case of 100 causal SNPs split by non-differentiated and differentiated SNPs. **(a-d)** All simulations used real UKBB admixed genotypes and simulated phenotypes with genetic correlation = 1.0 and a total additive genetic heritability of h^2 = 0.5 (see methods). "*" indicates a nominally significant p-value (<0.05). "**" indicates a Bonferroni-corrected significant p-value (<1.28 x 10^{-3}). The boxes show the interquartile range while the whiskers show the rest of the distribution (not including outliers).

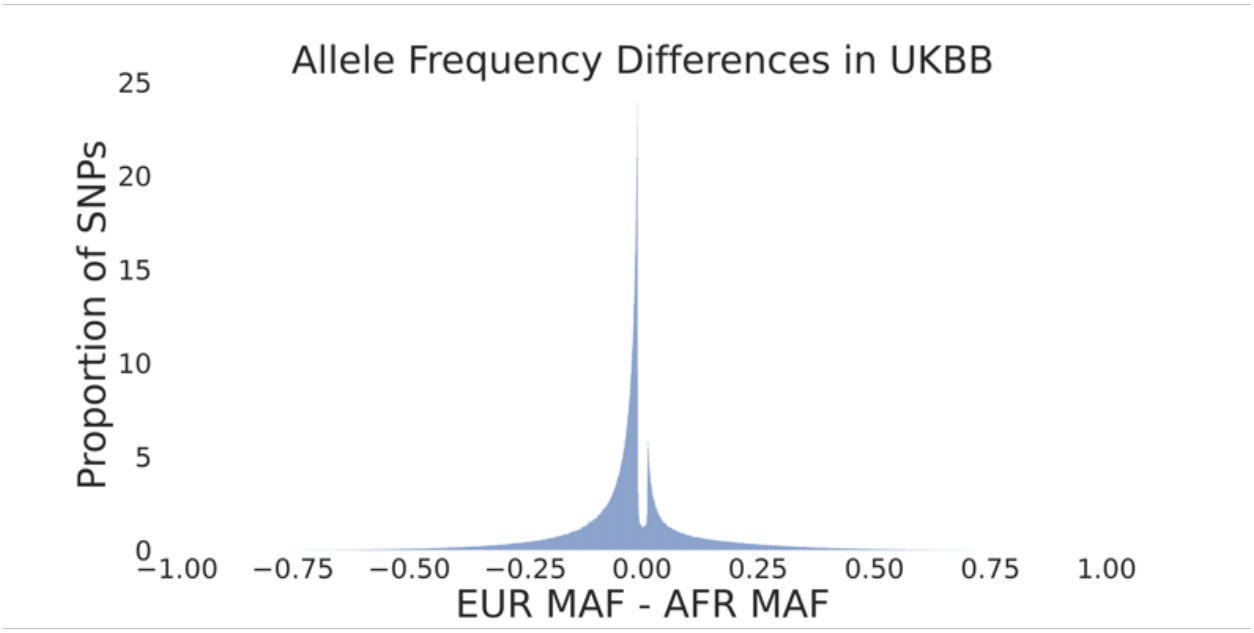

Figure S11: Minor allele frequency differences between European and African local ancestries in the African-European admixed population in the UKBB. Minor allele frequency differences center near zero, at -2.39 x 10^{-2}, indicating only a small systematic bias towards larger minor allele frequencies in the African local ancestry segments. Mean absolute value of minor allele frequency differences is 9.59 x 10^{-2}, indicating a small average allele frequency difference, with a standard deviation of 1.15 x 10^{-1}. Study population is 4,327 individuals from the UK Biobank with on average 58.9% African and 41.1% European admixed ancestry.

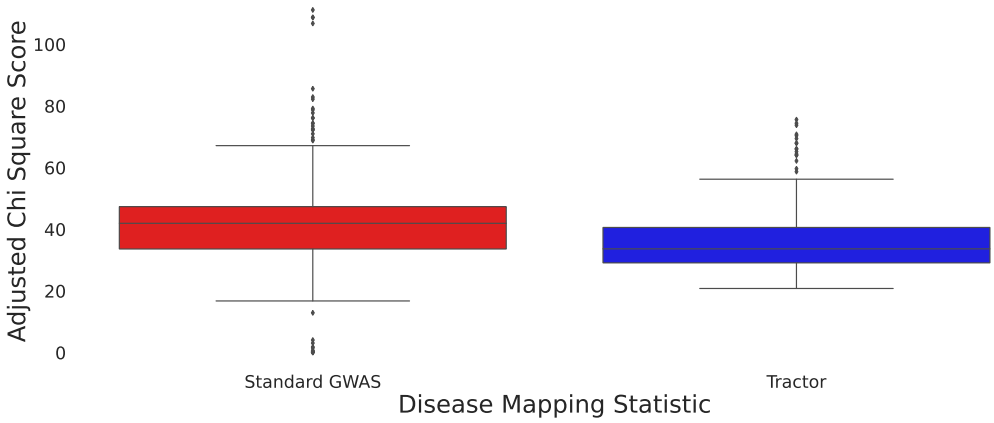

Figure S12: Adjusted Chi Square Statistics for significant SNPs for 12 traits in the UKBB. Standard GWAS χ_1^2 is significantly larger than the Tractor statistic (adjusted from χ_2^2 to χ_1^2). Mean Standard GWAS χ_1^2 for significant SNPs is 42.9, mean Tractor χ_2^2 for significant SNPs is 37.5, p-value 2.11 x 10^{-4}. Study population is 4,327 individuals from the UK Biobank with on average 58.9% African and 41.1% European admixed ancestry. Tractor and Standard GWAS statistics computed over 16,584,433 SNPs and 12 traits including AST, BMI, cholesterol, erythrocyte count, HDL, height, LDL, leukocyte count, lymphocyte count, monocyte count, platelet count, and triglycerides. See methods for chi-square adjustment. The boxes show the inter-quartile range while the whiskers show the rest of the distribution (not including outliers).

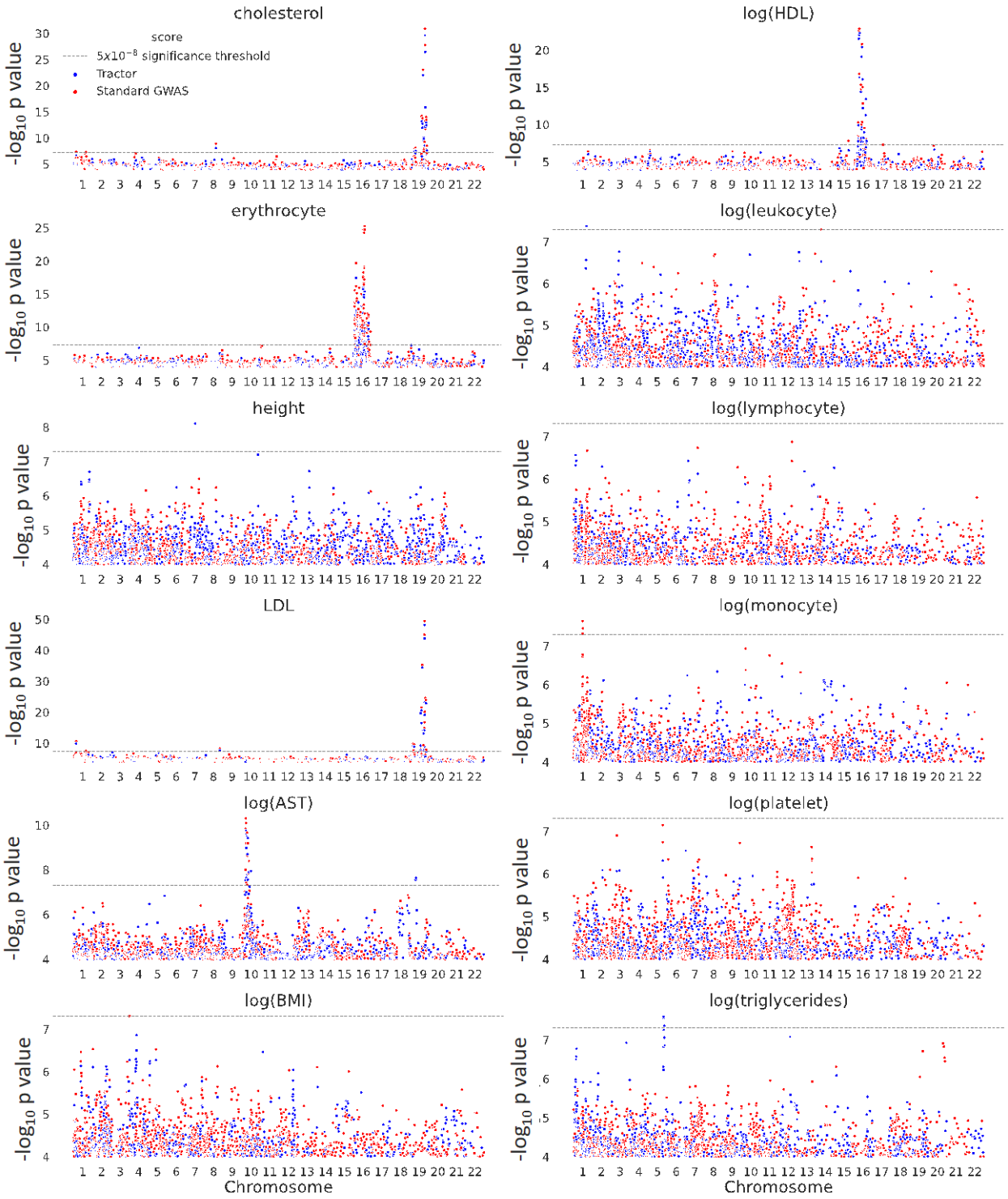

Figure S13: Manhattan plots for 12 quantitative traits in the UKBB African-European

admixed population. Study population is 4,327 individuals from the UK Biobank with on average 58.9% African and 41.1% European admixed ancestry. Manhattan plot SNPs shown filtered for p-value < 10^{-4} and SNPs are plotted based on post-filter indices.

Table S1: Number of Independent Significant Loci by Phenotype

Phenotype	# Loci Standard GWAS	# Loci Tractor	# Loci Shared
cholesterol	3	2	2
erythrocyte	3	3	2
Height	0	1	0
LDL	4	3	3
log(AST)	1	1	0
log(BMI)	1	0	0
log(HDL)	5	0	0
log(leukocyte)	1	0	0
log(lymphocyte)	0	0	0
log(monocyte)	1	0	0
log(platelets)	0	0	0
log(triglycerides)	0	0	0

Table S2: Independent Significant SNPs in UKBB Admixed Population

Phenotype	SNP (Reference Allele / Alternate Allele)	Standard GWAS p value	Tractor p value
cholesterol	chr1:55054772 (A / G)	3.72×10^{-8}	not significant
cholesterol	chr8:118543713 (A / T)	1.19×10^{-9}	8.31×10^{-9}
cholesterol	chr19:44908822 (C / T)	1.22×10^{-31}	2.31×10^{-30}
erythrocyte	chr16:261108 (G / A)	5.44×10^{-26}	not significant
erythrocyte	chr16:360054 (A / G)	9.15×10^{-13}	not significant
erythrocyte	chr16:50884914 (A / T)	4.92×10^{-10}	not significant
erythrocyte	chr16:117409 (C / T)	not significant	3.47×10^{-18}
erythrocyte	chr16:260355 (C / T)	not significant	6.34×10^{-18}
erythrocyte	chr16:384271 (G / A)	not significant	2.33×10^{-11}
Height	chr7:78824856 (G / A)	not significant	7.79×10^{-9}
LDL	chr1:55063542 (C / A)	2.47×10^{-11}	1.14×10^{-10}
LDL	chr1:88869866 (G / A)	3.01×10^{-8}	not significant
LDL	chr8:118543713 (A / T)	5.74×10^{-9}	2.45×10^{-8}
LDL	chr19:44908822 (C / T)	3.58×10^{-50}	6.24×10^{-49}
log(AST)	chr10:17819068 (G / A)	5.03×10^{-11}	not significant
log(AST)	chr19:17024164 (C / T)	not significant	2.30×10^{-8}
log(BMI)	chr3:196672134 (G / A)	4.83×10^{-8}	not significant
log(HDL)	chr15:76063105 (G / A)	1.54×10^{-8}	not significant
log(HDL)	chr16:56957451 (C / T)	1.34×10^{-8}	not significant
log(HDL)	chr17:58519260 (G / A)	4.30×10^{-8}	not significant
log(HDL)	chr17:58607316 (C / G)	4.94×10^{-8}	not significant
log(HDL)	chr17:58744530 (C / T)	4.94×10^{-8}	not significant
log(leukocyte)	chr14:30683993 (A / G)	4.96×10^{-8}	not significant
log(monocyte)	chr1:159092646 (G / A)	2.21×10^{-8}	not significant