nature portfolio

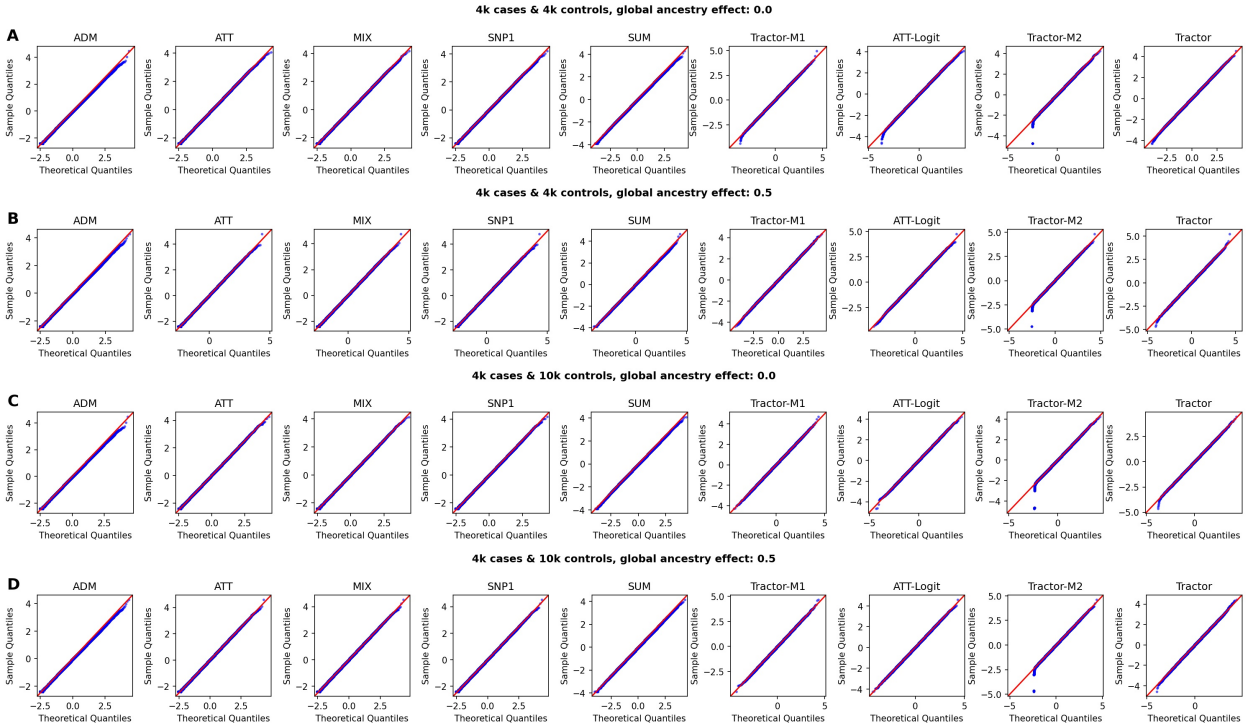
https://doi.org/10.1038/s41588-021-00953-5

Supplementary information

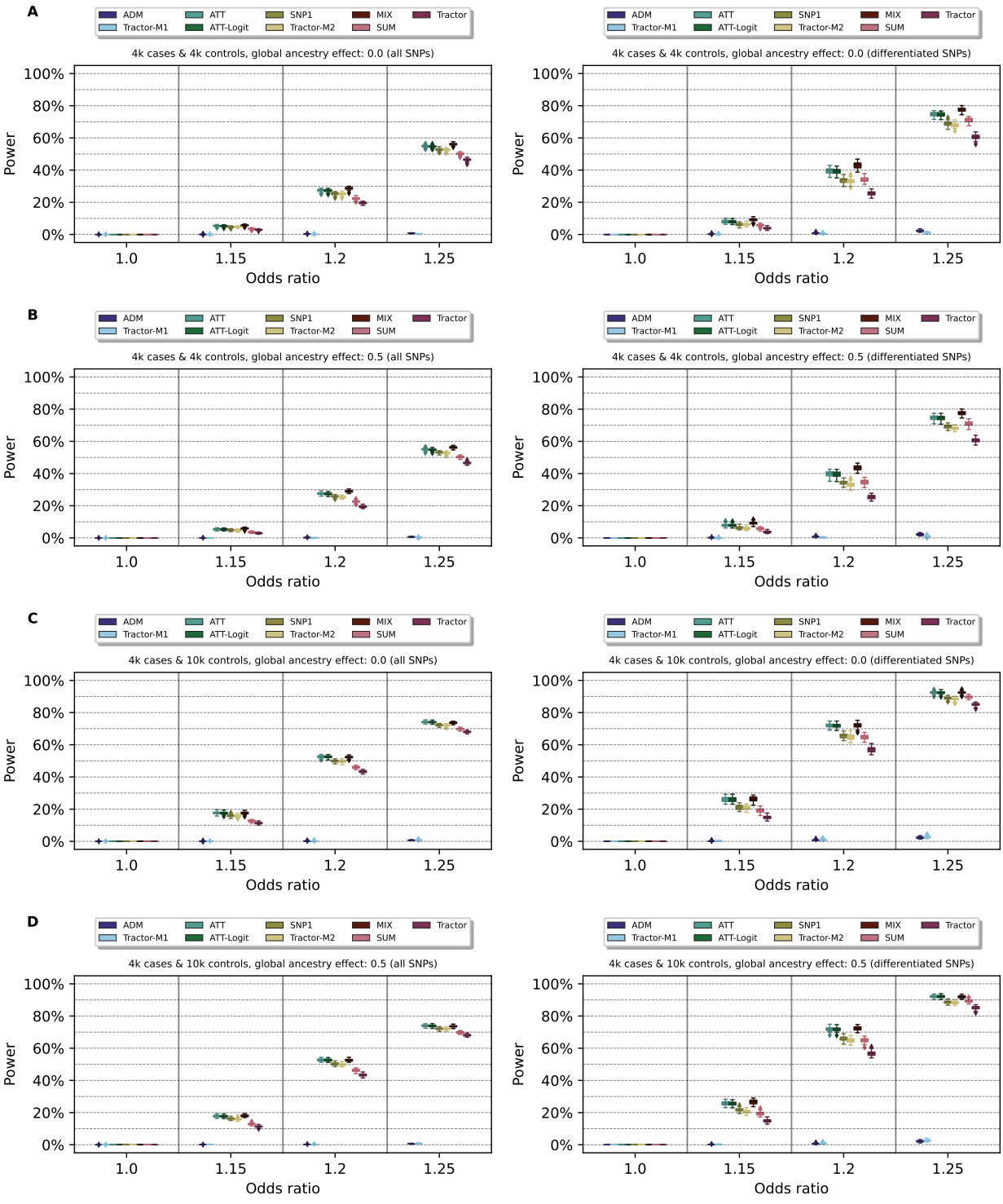
On powerful GWAS in admixed populations

In the format provided by the authors and unedited

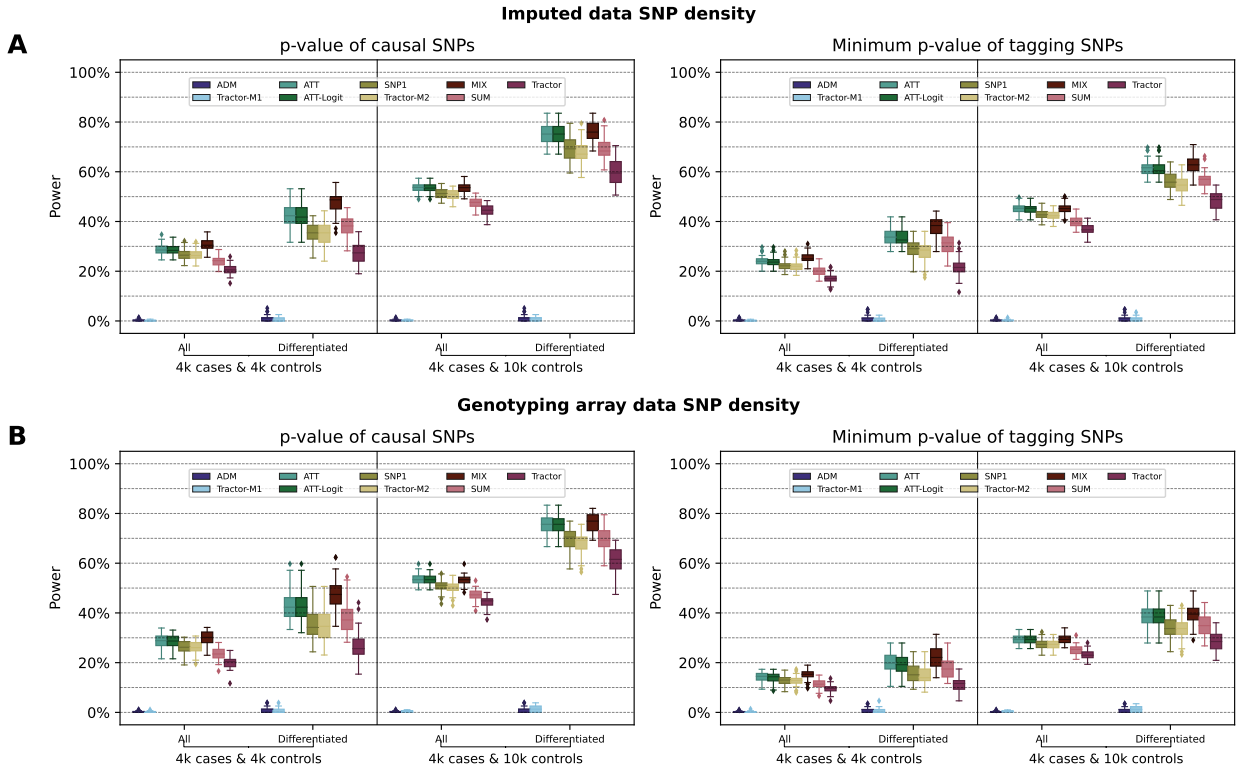
Supplementary Information for "On powerful GWAS in admixed populations"



Supplementary Figure 1: **All tests appropriately control false positive rates under the null simulations.** Quantile-quantile plots of p-values under null simulations. We simulate under four scenarios. **(A)** 4,000 cases and 4,000 controls, no global ancestry effect is simulated. **(B)** 4,000 cases and 4,000 controls, global ancestry effect is set to 0.5. **(C)** 4,000 cases and 10,000 controls, no global ancestry effect is simulated. **(D)** 4,000 cases and 10,000 controls, global ancestry effect is set to 0.5. Global ancestry effects are simulated by adding "[global ancestry effect] $\times$ AFR admixture proportion" to the logit model, similar to Atkinson et al. 2021.



Supplementary Figure 2: **Power simulation with varying odds ratio and global ancestry effects.** We simulate under four scenarios (one scenario per row). For each row, the left sub-figure represents distributions of power estimates from 50 simulation replicates and 3,000 causal SNPs uniformly drawn from the set of all SNPs (50 points per boxplot), while the right sub-figure restricts to the subset of SNPs (904 out of 3,000) with absolute allele frequency difference > 0.2 between Europeans and Africans (50 points per boxplot). See more descriptions in captions of Figure 1A and Supplementary Figure 1.



Supplementary Figure 3: **Power simulation when the causal variant is untyped.** We start with 546k SNPs on chromosome 2 with MAF $> 1\%$ in both the European- and African-ancestries individuals in 1000 Genomes, and uniformly sample 300 SNPs as causal SNPs. For each causal SNP, we select 40 nearby SNPs (as well the causal SNP) to form a region of 41 SNPs. We consider two options to select these nearby SNPs: **(A)** nearest 40 SNPs of each selected causal SNP, which corresponds to a SNP density with 546k SNPs on chromosome 2, roughly matching the imputed data SNP density. **(B)** among 400 nearest SNPs of each selected causal SNP, uniformly sample one SNP every 10 SNPs, which corresponds to a SNP density with 55k SNPs on chromosome 2, roughly matching the genotyping array data SNP density. For each formed region, we simulate phenotypes with odds ratio 1.2 only on causal SNP (the same as Figure 1A) (50 simulation replicates) and perform association testing on the causal SNP as well as the 40 tagging SNPs. We show the proportion of SNPs exceeding genome-wide significance (power) with **(Left)** p-value of causal SNPs **(Right)** minimum p-value of the 40 tagging SNPs (masking the true causal SNP). See caption of Figure 1A, Methods for more description of simulation setup.