The American Journal of Human Genetics, Volume *104*

Supplemental Data

Genome-wide Significance Thresholds for Admixture Mapping Studies

Kelsey E. Grinde, Lisa A. Brown, Alexander P. Reiner, Timothy A. Thornton, and Sharon R. Browning

Supplemental Figures

Figure S1. Correlation of admixture mapping test statistics with binary versus quantitative traits. To confirm the validity of our theoretical work for binary traits, we simulated traits and local ancestry at pairs of loci for admixed individuals in a variety of populations. For each admixed individual $i = 1, \dots, n$, we first drew global ancestry proportions from a pre-specified distribution F representing different population structure scenarios: no structure ($\boldsymbol{\pi_i} = \boldsymbol{\pi}\ \forall i$), subpopulations ($\boldsymbol{\pi_i} = \{\boldsymbol{\pi_1}, \boldsymbol{\pi_2}, \dots, \boldsymbol{\pi_P}\}$ with probability $\{p_1, p_2, \dots, p_P\}$, where $\sum_{l=1}^{P} p_l = 1$), or Dirichlet global ancestry ($\boldsymbol{\pi_i} \sim_{iid} Dirichlet(\boldsymbol{\alpha})$). We considered various choices of number of individuals (n), number of ancestral populations (K), and hyperparameters for the distribution of global ancestry proportions F. For each individual's two haplotypes, we independently simulated crossover events between two loci separated by recombination fraction θ across g generations according to a Poisson process. We simulated ancestry at the first locus according to a Multinoulli (categorical) distribution with probabilities equal to the global ancestry vector $\boldsymbol{\pi_i}$. Using the simulated crossover history, we determined whether any recombination had occurred between the two loci; if so, we independently simulated ancestry at the second locus according to the same Multinoulli distribution; if not, we set ancestry at the second locus equal to ancestry at the first. We simulated binary traits for each individual according to the model $y_i \sim_{iid} Bernoulli(0.2)$, and quantitative traits according to $y_i \sim_{iid} N(0,1)$. Finally, we paired the individuals' haplotypes, recorded the local ancestry vectors for each individual, and calculated admixture mapping test statistics at each locus using the simulated traits. We repeated this process 10,000 times, calculated the correlation of admixture mapping test statistics at the two loci across simulation replicates, then compared the observed patterns of correlation to the expected correlation given by our theoretical results (Theorem 1). Panels (A) and (B) present the observed versus theoretical correlation of admixture mapping test statistics (testing the first ancestry component at two loci separated by recombination fraction θ) in an admixed population with $K = 3, n = 10{,}000, \boldsymbol{\pi_i} \sim_{iid} Dirichlet([1,1,1])$, and either binary (Panel A) or quantitative (Panel B) traits.