Supplementary Figures

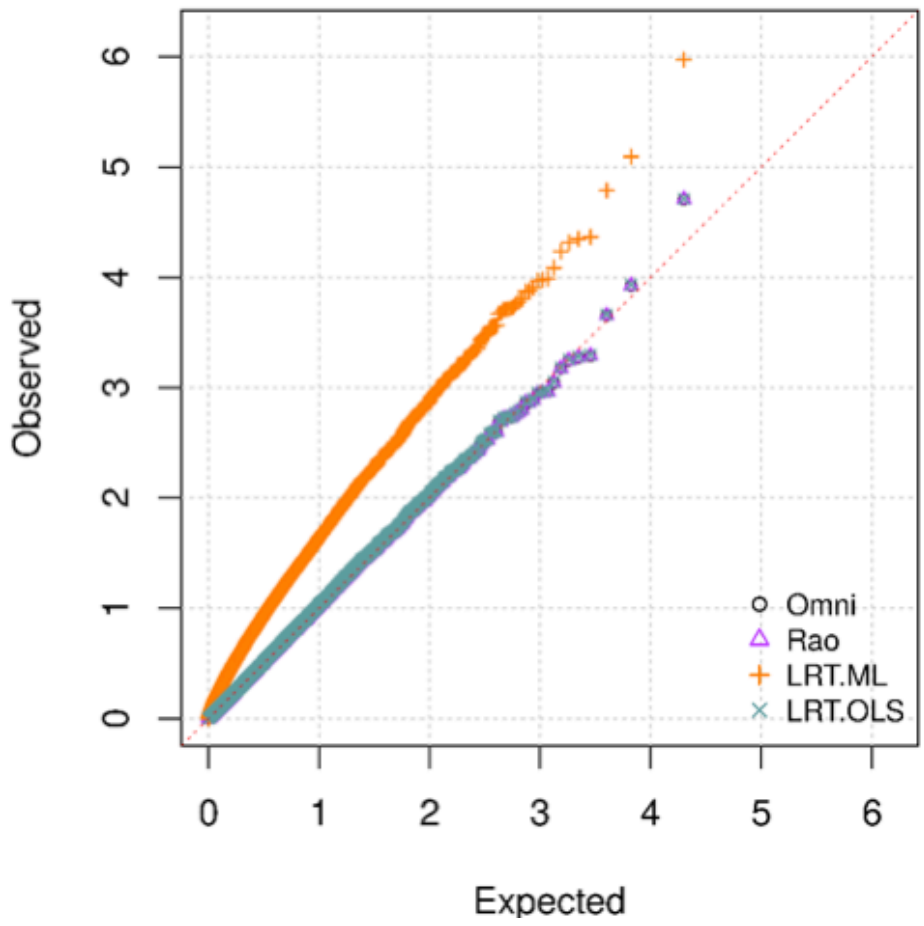


Figure S1. Q-Q plot of the p-values for the nullity of the interaction effects in a linear regression framework. We conducted a series of 10,000 simulations to compare the performances of the different tests classically performed in linear regression to assess the significance of effect sizes. For each simulation, we simulated 100 independent genotypes for 20,000 individuals, 10 independent exposures and a continuous phenotype independent from both the genotypes and the exposures (as well as their interactions). We considered the linear regression model including all the genotypes, all the exposures and the 10×100=1000 genotype-environment interaction terms. Finally, we computed the p-values obtained from testing the simultaneous nullity of all the interaction effects using the Omnibus test (Wald test statistic, black; the Rao's Score test, purple; the Likelihood Ratio Test with Maximum Likelihood estimator of the residual variance (LRT.ML), orange; and LRT with the Ordinary Least Squares estimator of the residual variance (LRT.OLS), blue).

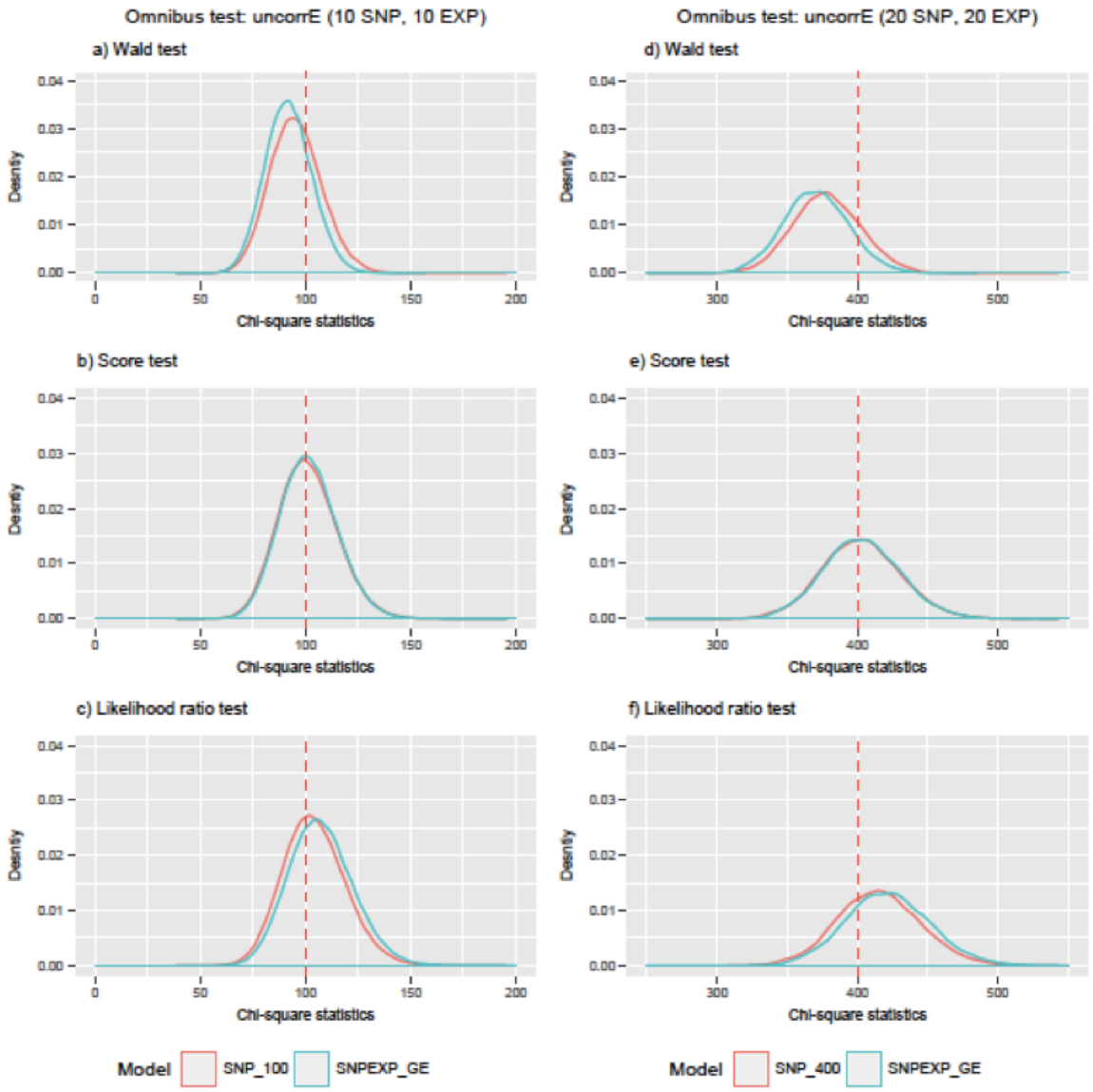


Figure S2. Test statistics of joint analysis for multiple parameters in logistic models. We simulated data with 10,000 replications holding the same number of EPV (EPV=5) to compare test statistics of two models: (1) models with SNP only and (2) models with G-E interactions. We compared three test statistics (*Wald*, *Score*, and *LRT*) varying the number of df (i.e. 100 and 400). We derived the empirical distribution of the test statistics under the null (i.e. no genetic effect or no interaction effect). We used different numbers of sample size and parameter; panels (A), (B), and (C) had 120 SNPs (red) or 10 SNPs and 10 exposures (blue) to test 100 parameters jointly (n = 2,000, 30% prevalence); panels (D), (E) and (F) had 440 SNPs or 20 SNPs and 20 exposures to test 400 parameters jointly (n = 7,000, 30% prevalence).

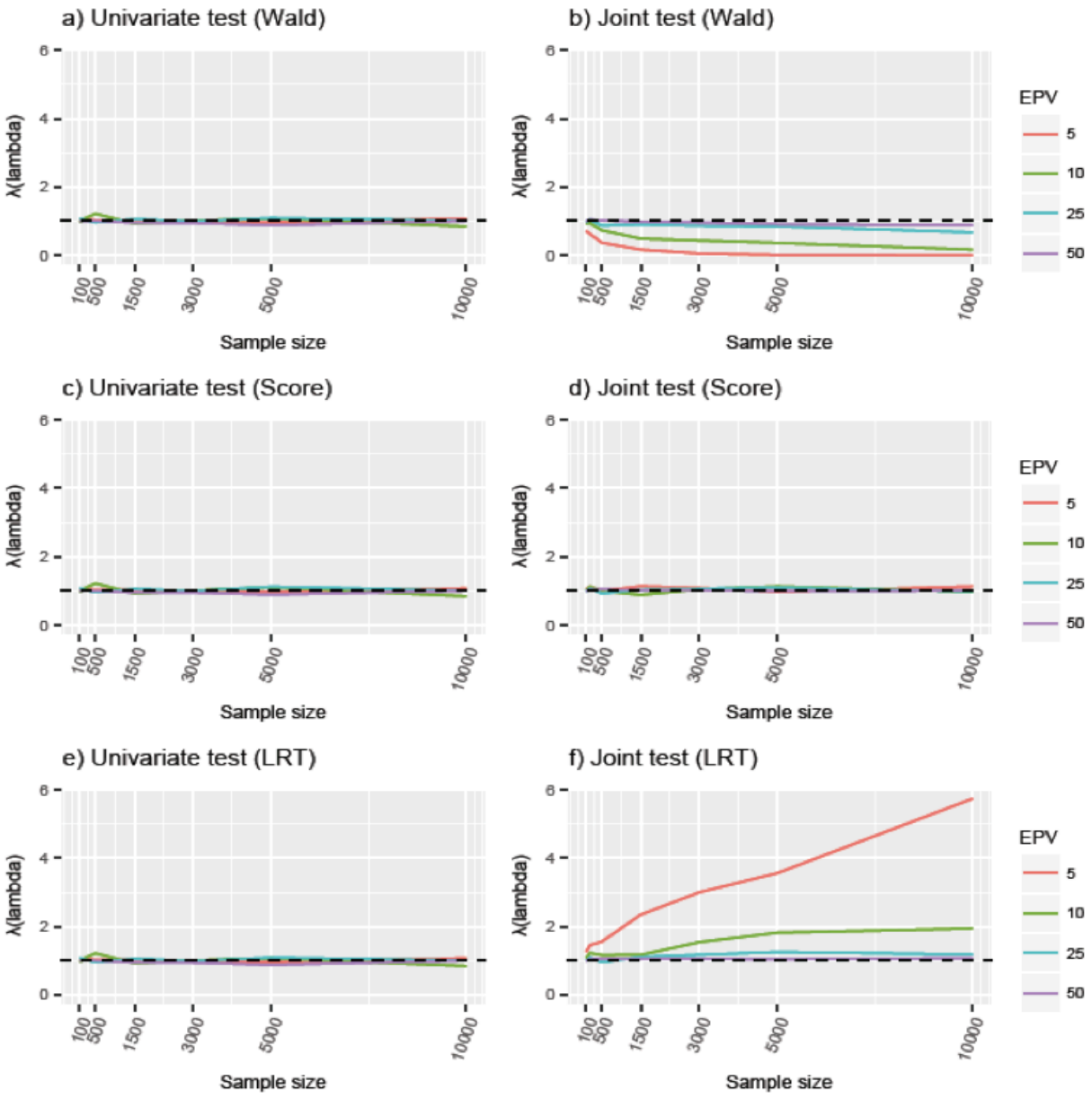


Figure S3. Robustness comparison between univariate and omnibus tests by different number of event per variable. We generated simulations of common SNPs with 1,000 replicates varying the number of event per variable (EPV) (range = 5 - 50) in logistic regression. For simplicity, we set prevalence of 0.5 for all tests. We performed univariate tests (A, C, and E) and joint tests (i.e. omnibus) (B, D, and F) with three test statistics (*Wald*, *Score*, and *LRT*) under the null and compared genomic inflation factor (λ) values. For univariate tests, we selected the p-value of a random SNP based on chi-square distribution with df=1. Y-axis was the estimated lambda value and x-axis was the sample size (n = 100-10,000). Each color line represents the number of EPV.

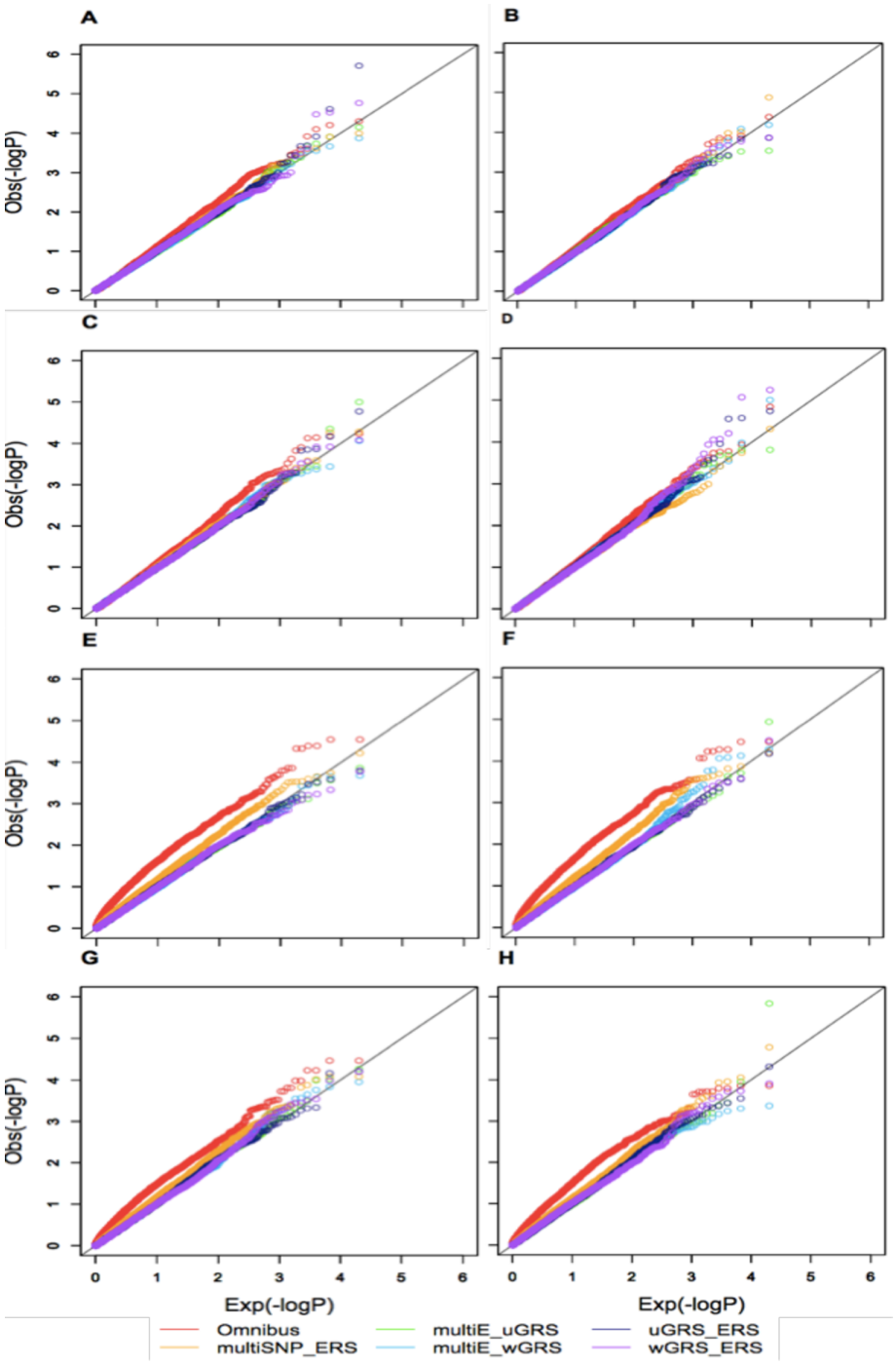
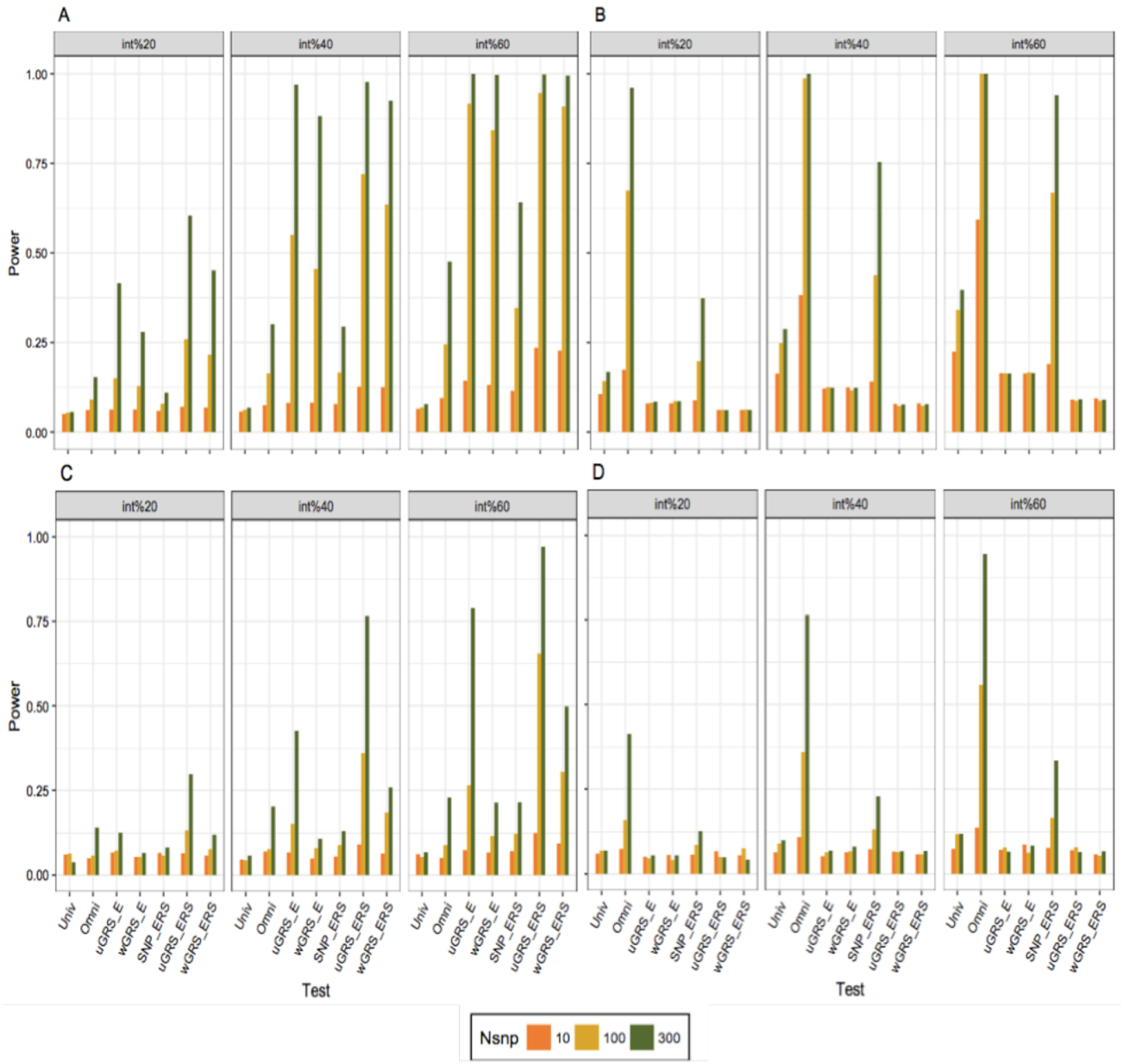


Figure S4. Q-Q plots of null models assuming no interaction effect between 300 SNPs and 10 exposures (Table 1). (A) Linear models of rare SNPs with independence between G and E. (B) Linear models of rare SNPs with no independence between G and E. (C) Linear models of common SNPs with independence between G and E. (D) Linear models of common SNPs with no independence between G and E. (E) Logistic models of rare SNPs with independence between G and E. (F) Logistic models of rare SNPs with no independence between G and E. (G) Logistic models of common SNPs with independence between G and E. (H) Logistic models of common SNPs with no independence between G and E.



Figiure S5. Power comparison of G-E interaction approaches with normal distributed and moderate correlated exposures. This figure represents power plots of simulation data (20,000 samples, 10 exposures, varying number of SNP) with 10,000 replications (expect for univariate and omnibus tests in logistic models using 1,000 replicates). (A) Linear models assuming all G-E interactions effects correlated with marginal effects. (B) Linear models assuming no correlations between G-E interaction effects and marginal effects, (C) Logistic models assuming correlations between G-E interaction effects and marginal effects. (D) Logistic models assuming no correlations between G-E interaction effects and marginal effects.

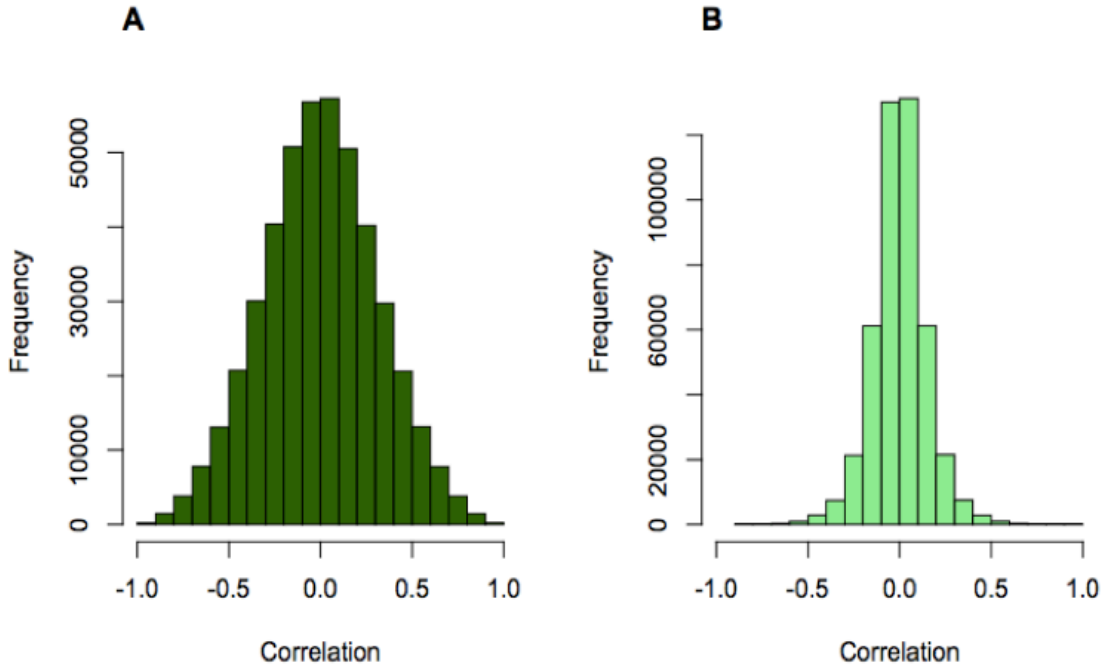


Figure S6. Distribution of correlations across exposures. (A) Relatively strong correlations across exposures (mean of pairwise correlation = 0.10). (B) Moderate correlations across exposures (mean of pairwise correlation = 0.02).

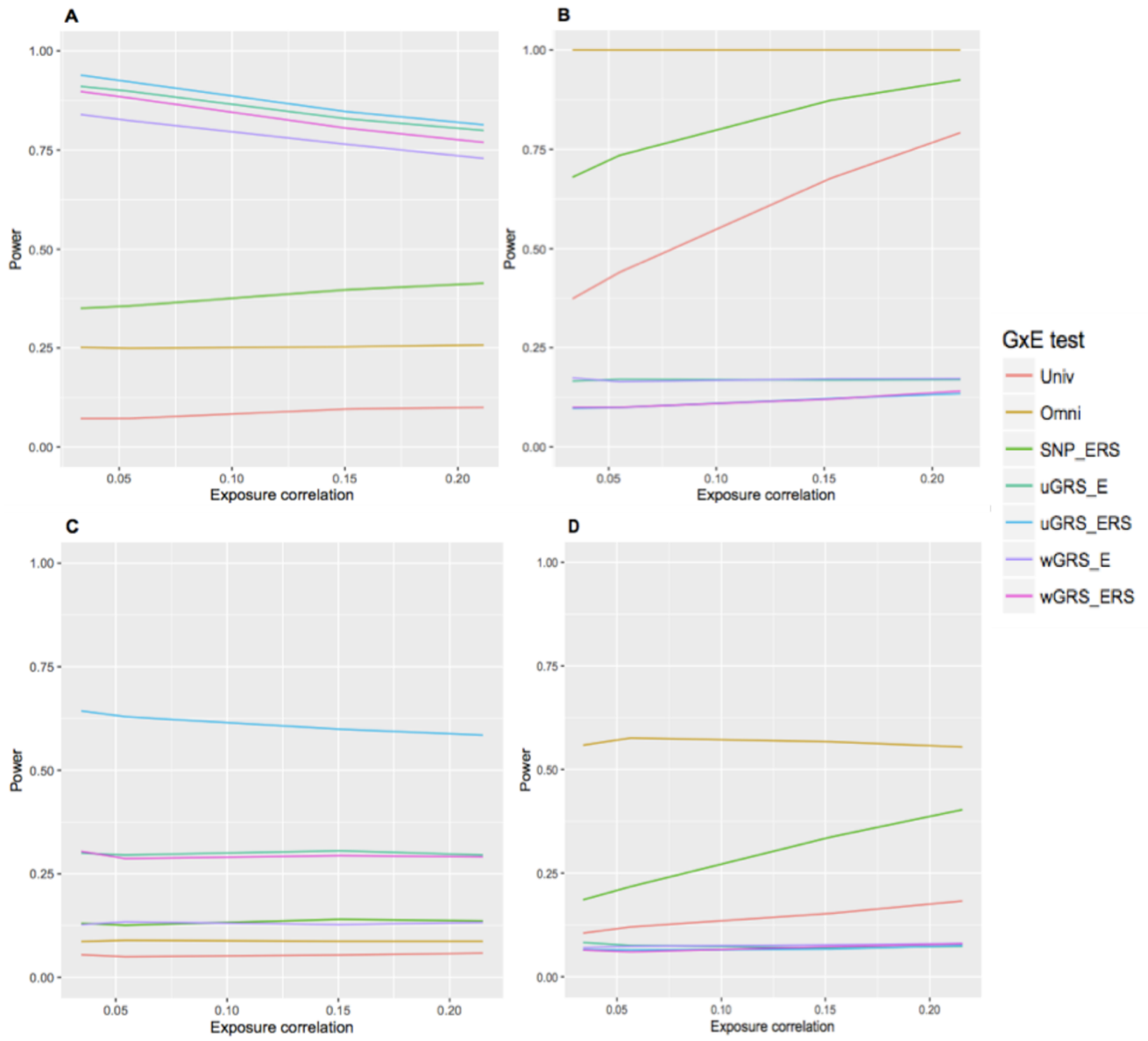


Figure S7. Change in power by size of exposure correlation. To illustrate the change in power by the amount of correlations between exposures, we simulated data of 20,000 samples, 10 exposures, and 100 SNPs (assuming 60% true G-E interactions and independence between SNPs and exposures) with 1,000 replications. (A) Linear models assuming all G-E interactions effects correlated with marginal effects. (B) Linear models assuming no correlations between G-E interaction effects and marginal effects, (C) Logistic models assuming correlations between G-E interaction effects and marginal effects. (D) Logistic models assuming no correlations between G-E interaction effects and marginal effects.