**A Bayesian framework for generalized linear mixed modeling identifies new candidate loci for late-onset Alzheimer’s disease**

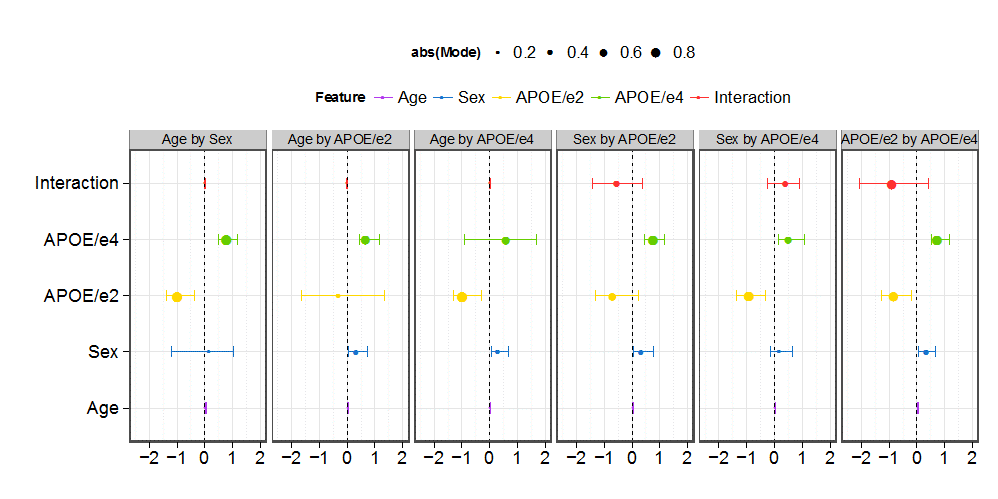
Xulong Wang, et al.

**SUPPLEMENTAL TABLE CAPTIONS**

**Supplemental Table 1 (SupplementalTable1.xls, 33 kb).** Ensembl variant annotations for the top 55 associated variants. Variants commonly mapped to multiple annotations.

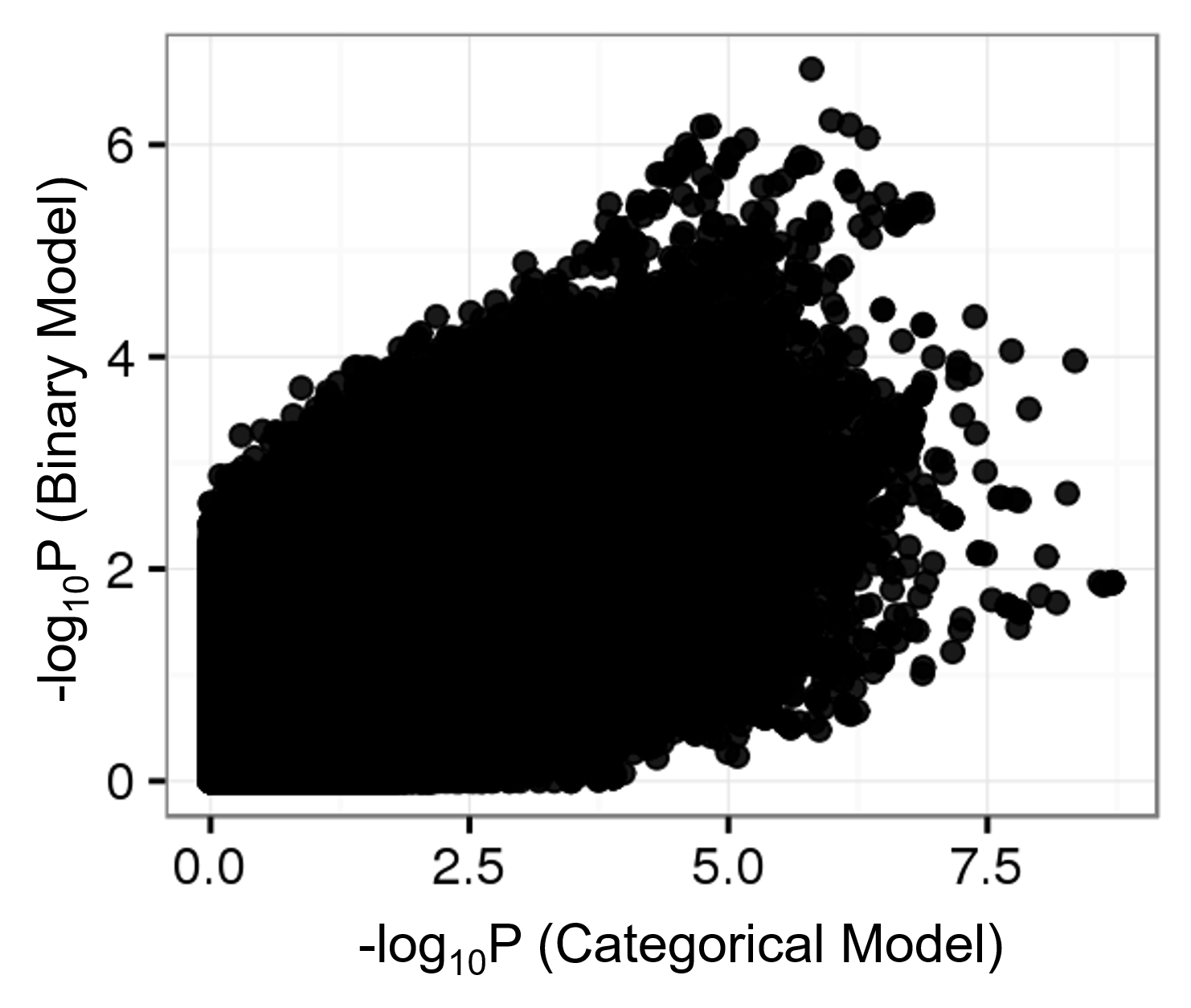
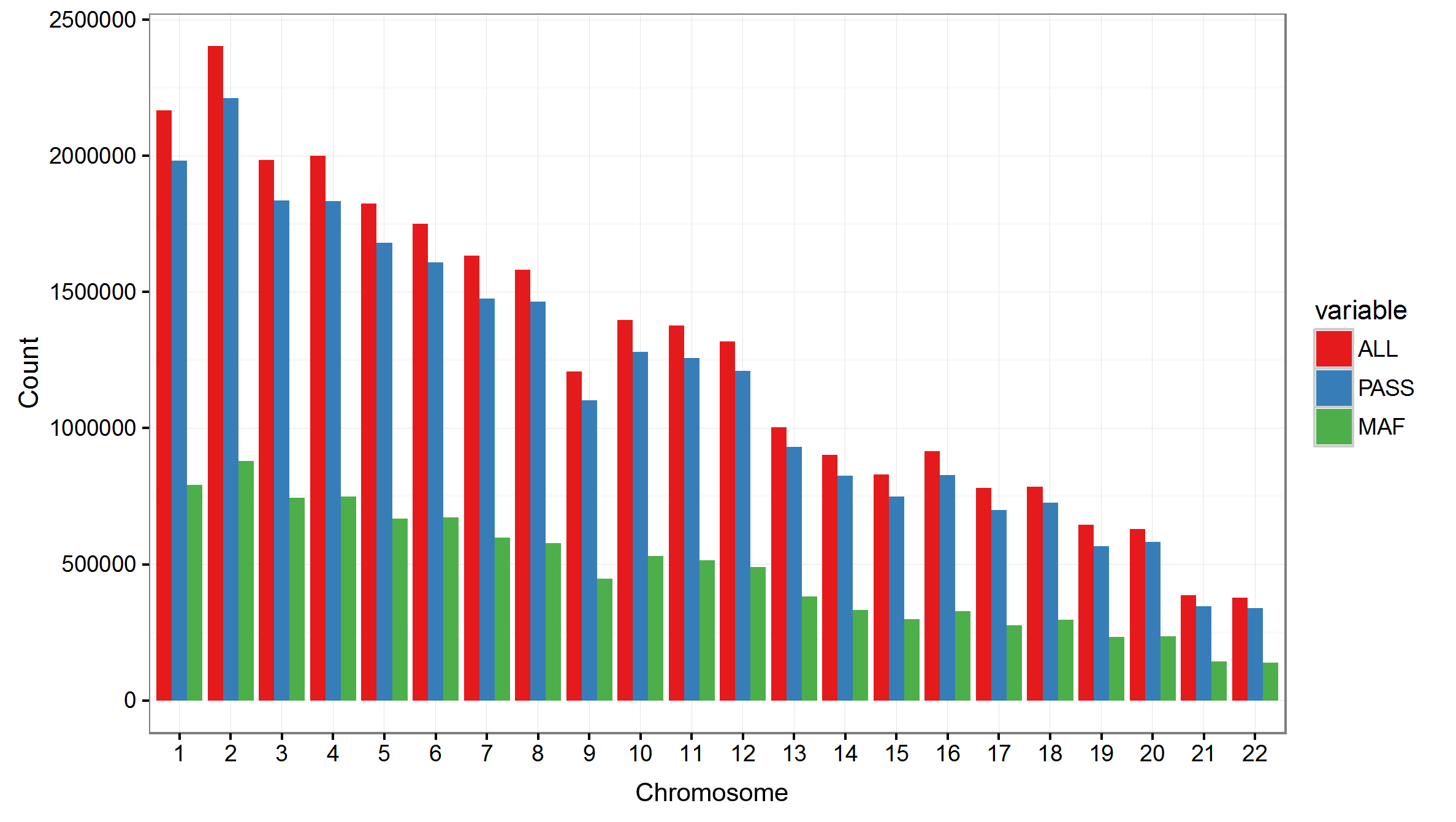
**Supplemental Table 2 (SupplementalTable2.xls, 15 kb).** NHGRI GWAS annotations for the top 55 associated variants.

**SUPPLEMENTAL FIGURES**

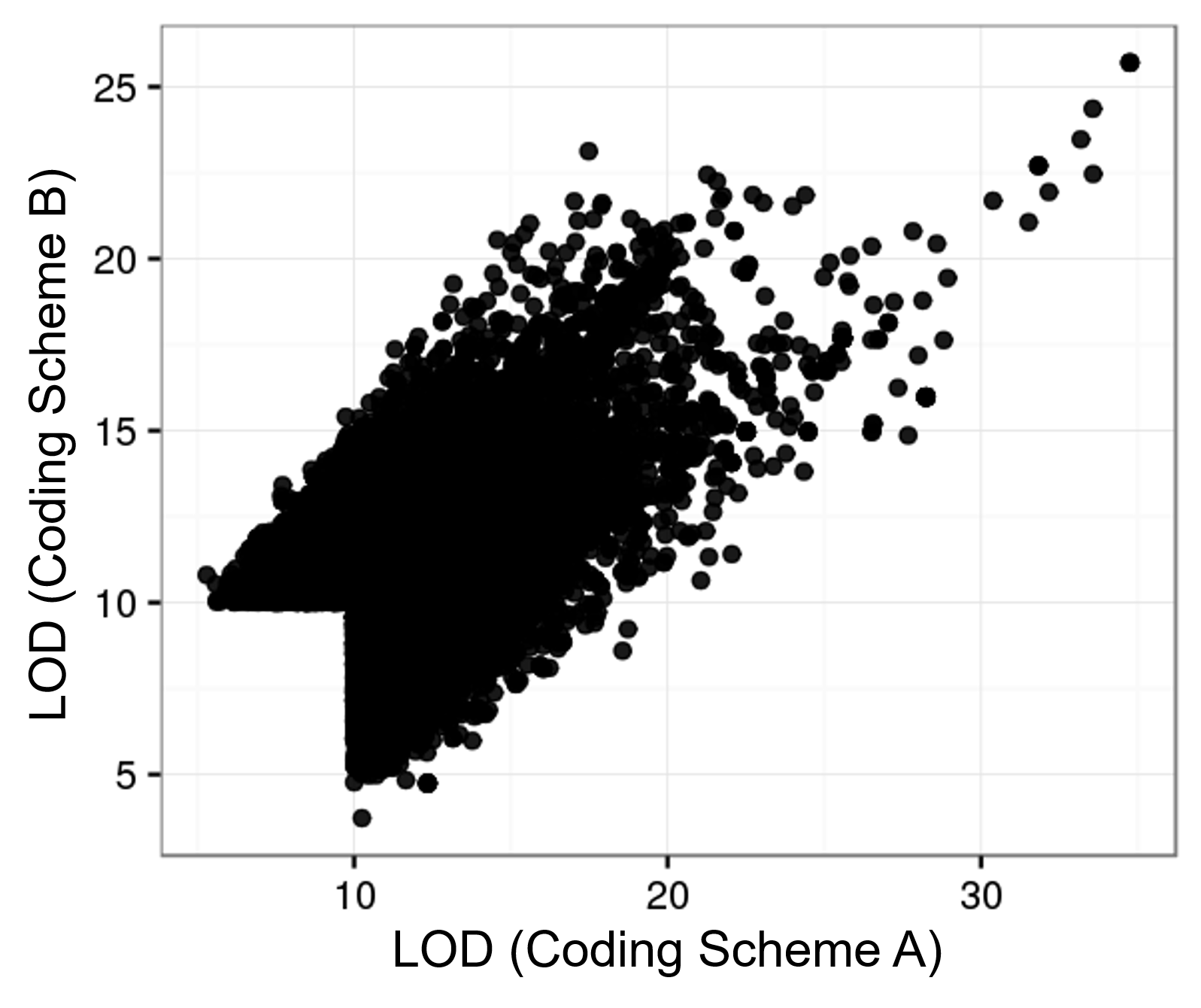


**Supplemental Figure 1.** Interaction tests for covariates estimated by MCMC sampling in *Bayes-GLMM*. No interactions were detected between any pair of covariates (age, sex, APOE/e2, APOE/e4).

**Supplemental Figure 2.** Genome-wide variants in the ADSP WGS cohort stratified by chromosomes. Red bars represent the number of total variants, blue bars represent the number of variants that passed the quality check (Methods), and green bars represent variants that pass the quality check and occur at MAF greater than 0.01.



**Supplemental Figure 3.** P-values of the GWAS were distorted by collapsing the ADSP four-level categorical diagnostic variables as binary. Bayes-GLMM was used to test 10.3 million genomic variants with MAF greater than 0.01 generalized linear model, using either the four-level categorical variable or a collapsed binary variable to mimic case-control status.



**Supplemental Figure 4.** Results of GWAS using numerical variables shows inconsistency due to numeric coding scheme. The four categorical levels of AD diagnosis in ADSP (no, possible, probable, definite) were transformed to numerical by two rules: Coding Scheme A with 0, 0.25, 0.5, 1; and Coding Scheme B with 0, 0.33, 0.66, 1. Variants with LOD results greater than 10 in either of the two schemes were plotted (23,388 variants). Analysis was performed using a linear mixed model in *R/QTLRel*.