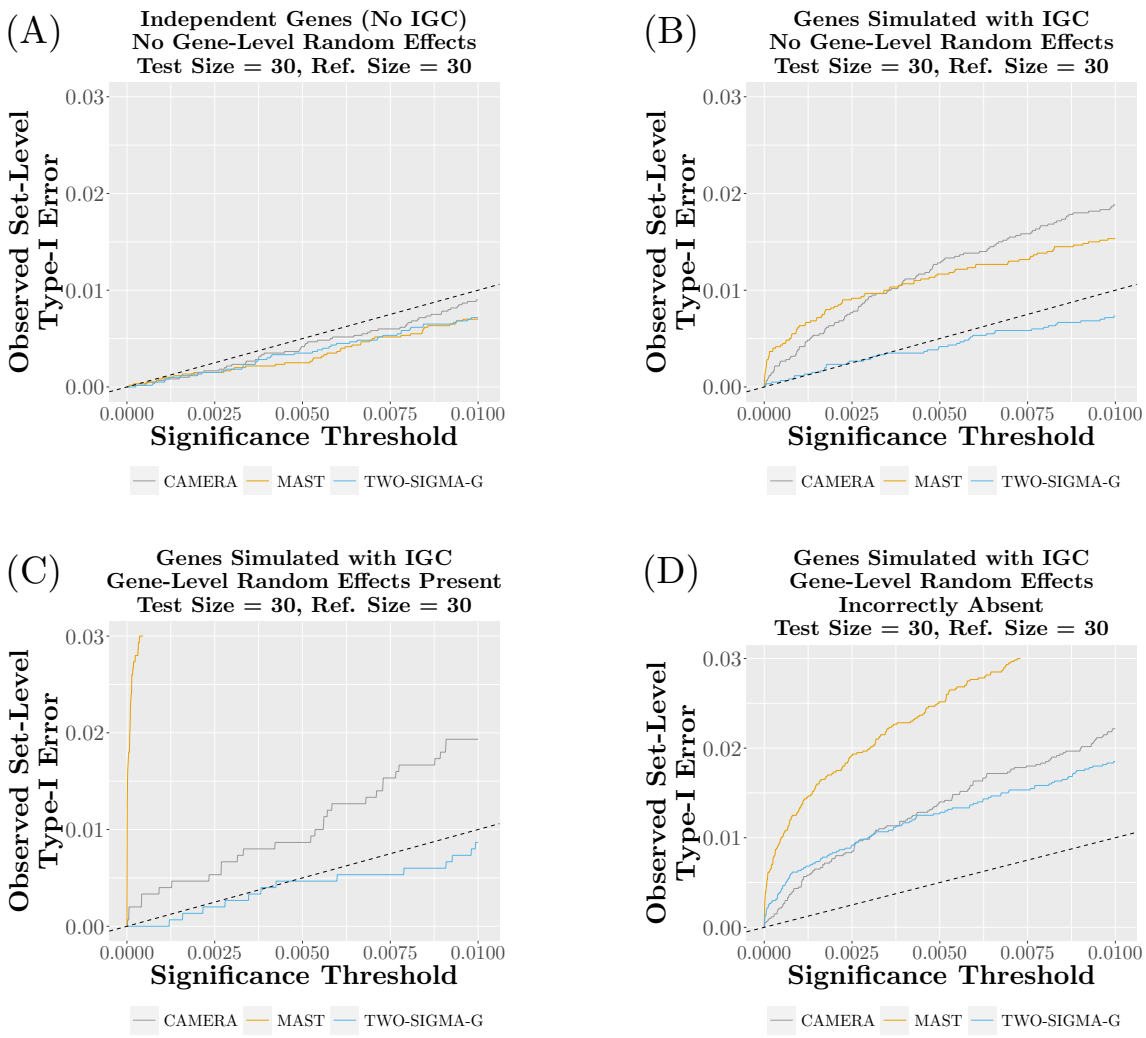
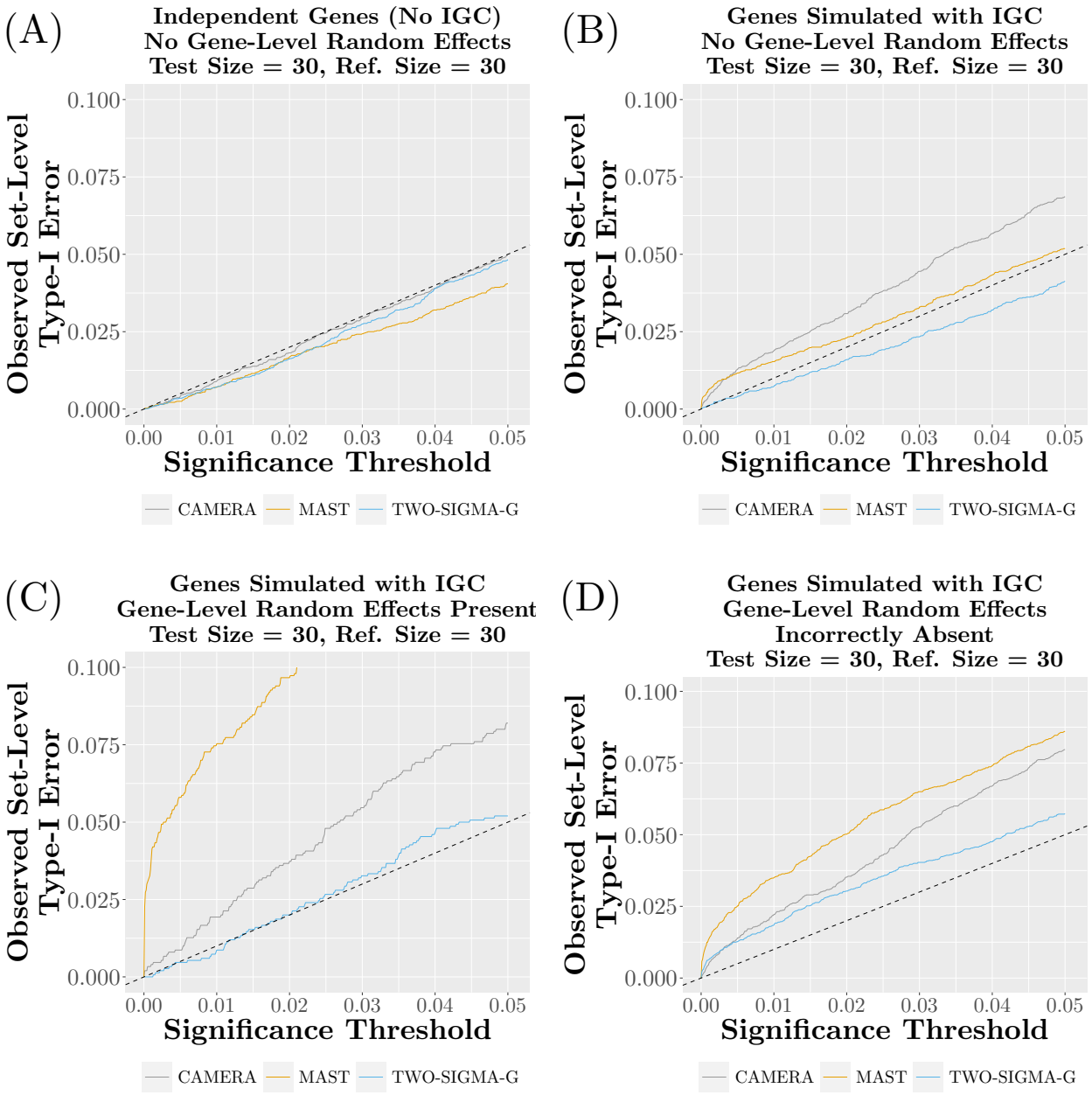
Supplement for TWO-SIGMA-G

Eric Van Buren, Ming Hu, Liang Chen, John Wrobel, Kirk Wilhelmsen, Lishan Su, Yun Li, Di Wu

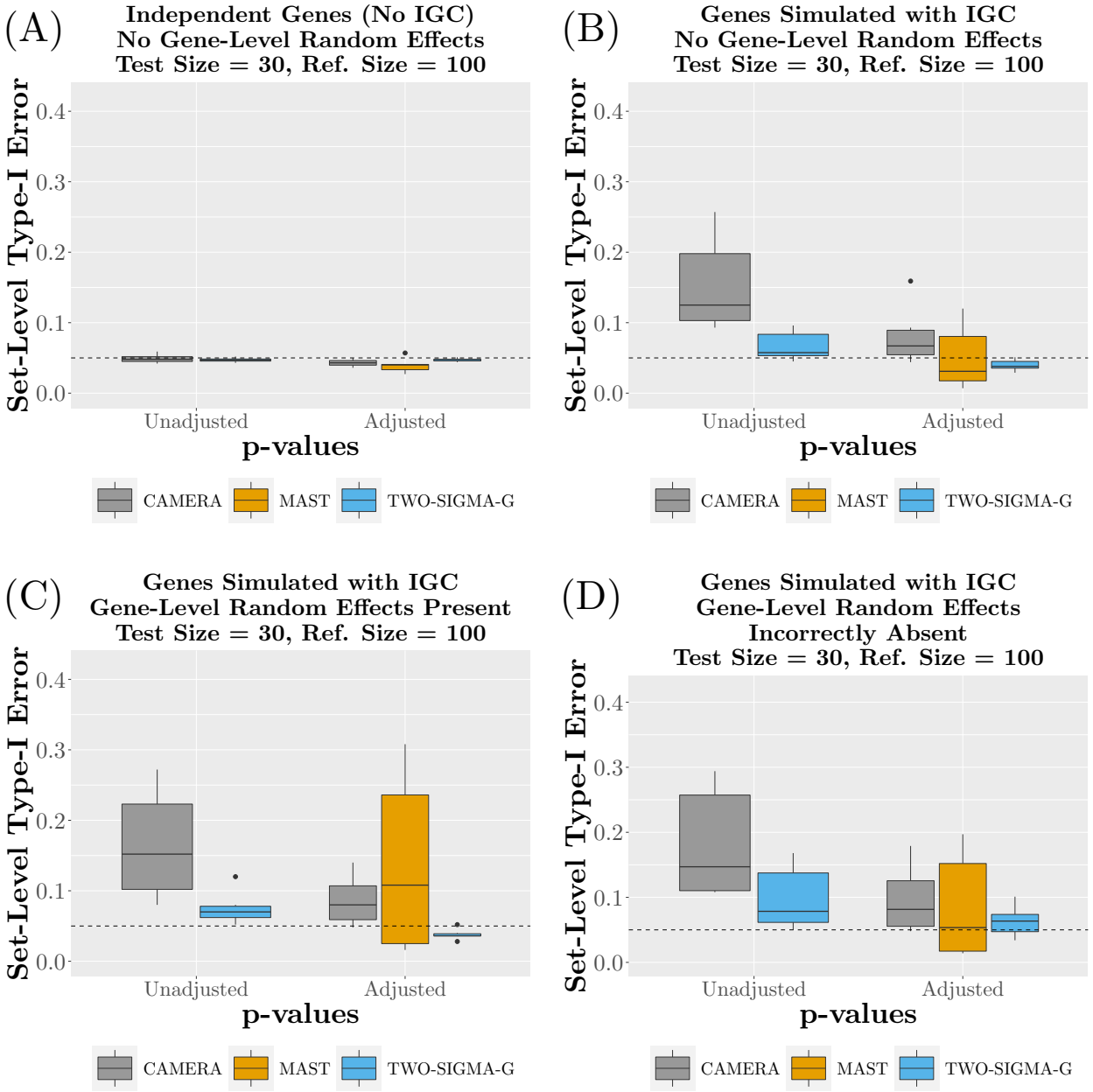
S1 Additional Type-I Error and Power Results



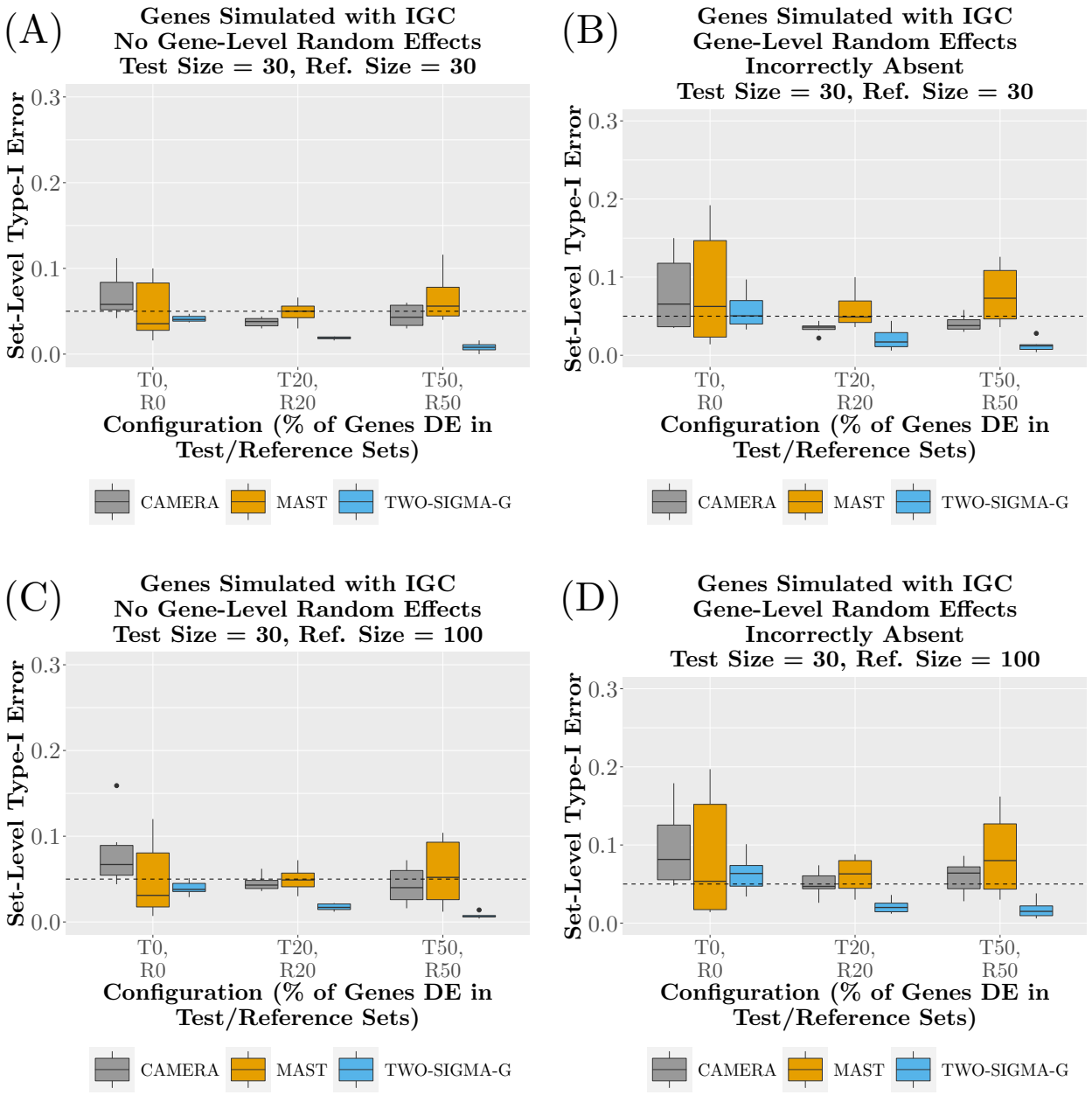
Supplementary Figure S1: **Type-I error performance of CAMERA, MAST, and TWO-SIGMA-G as significance threshold varies from 0 to .01**. Each panel varies the existence of IGC between genes in the test set and the presence of gene-level random effect terms in gene-level model (CAMERA never includes gene-level random effect terms). Each plot combines six different settings, and 10 replicates per setting, which vary both the magnitude of the average inter-gene correlation (where applicable) in the test set and the nature of the correlation structure via the introduction of other individual-level covariates. See the Methods section of the main text for more details regarding the simulation procedure.



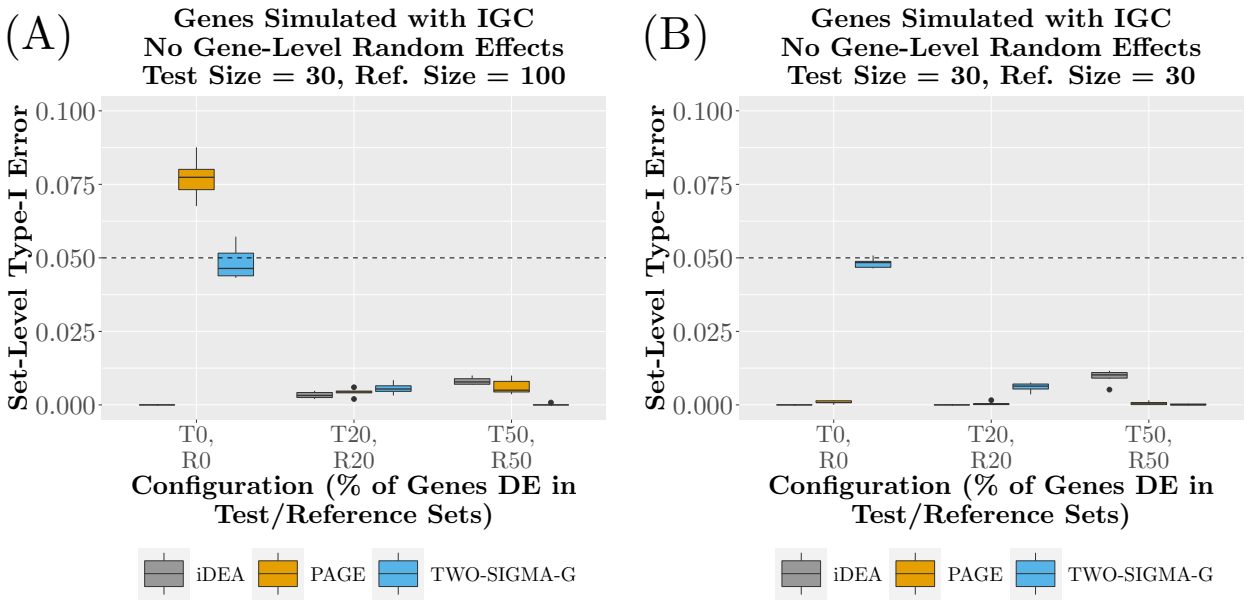
Supplementary Figure S2: **Type-I error performance of CAMERA, MAST, and TWO-SIGMA-G as significance threshold varies from 0 to .05**. Each panel varies the existence of IGC between genes in the test set and the presence of gene-level random effect terms in gene-level model (CAMERA never includes gene-level random effect terms). Each plot combines six different settings, and 10 replicates per setting, which vary both the magnitude of the average inter-gene correlation (where applicable) in the test set and the nature of the correlation structure via the introduction of other individual-level covariates. See the Methods section of the main text for more details regarding the simulation procedure.



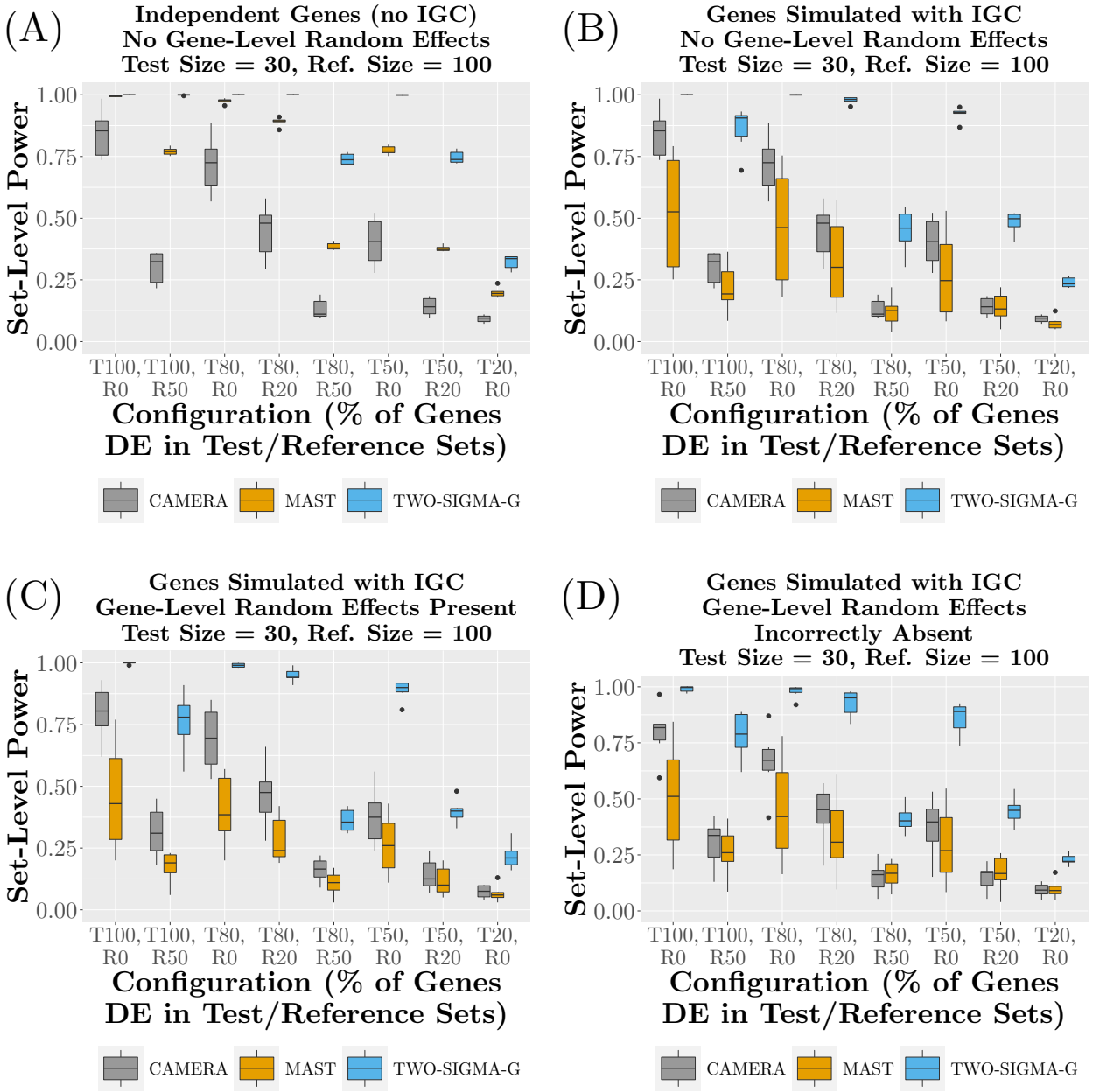
Supplementary Figure S3: **Type-I error performance of CAMERA, MAST, and TWO-SIGMA-G using a reference set size of 100 genes**. Each panel varies the existence of IGC between genes in the test set and the presence of gene-level random effect terms in gene-level model (CAMERA never includes gene-level random effect terms). Within each panel, both unadjusted and adjusted set-level p-values are plotted (unadjusted p-values are unavailable for MAST). Each boxplot aggregates six different settings which vary both the magnitude of the average inter-gene correlation (where applicable) in the test set and the nature of the correlation structure via the introduction of other individual-level covariates. Such settings are intended to represent the diversity seen in real data sets to paint an accurate picture of testing properties over a wide range of gene sets. Each of the six settings is further composed of 10 replicates which vary only random seed to mimic the impact of a different starting pool of cells from which genes were simulated. See the Methods section of the main text for more details regarding the simulation procedure.



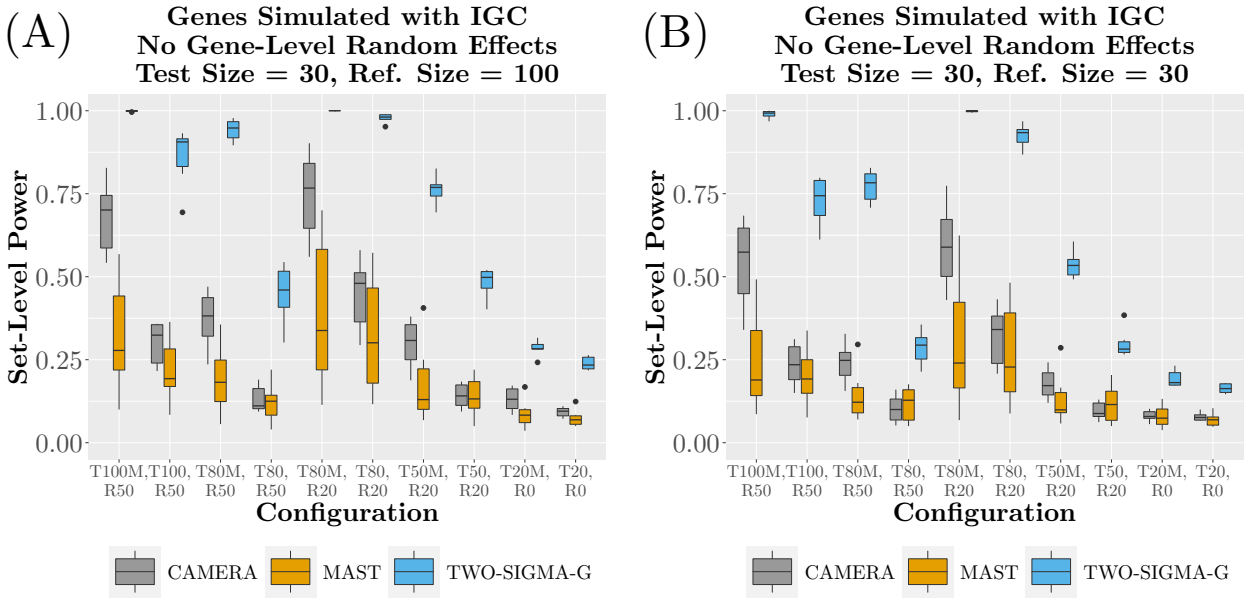
Supplementary Figure S4: **Type-I error performance of TWO-SIGMA-G, CAMERA, and MAST for various set-level null hypotheses**. Each panel varies the reference set size and the presence of gene-level random effect terms in gene-level model. Scenarios along the x-axis of each panel vary the percentage of genes that are differentially expressed (with the same effect size) in the test and reference sets. Each boxplot aggregates six different settings which vary both the magnitude of the average inter-gene correlation in the test set and the nature of the correlation structure via the introduction of other individual-level covariates. Such settings are intended to represent the diversity seen in real data sets to paint an accurate picture of testing properties over a wide range of gene sets. Each of the six settings is further composed of 10 replicates which vary only random seed to mimic the impact of a different starting pool of cells from which genes were simulated. See the Methods section of the main text for more details regarding the simulation procedure.



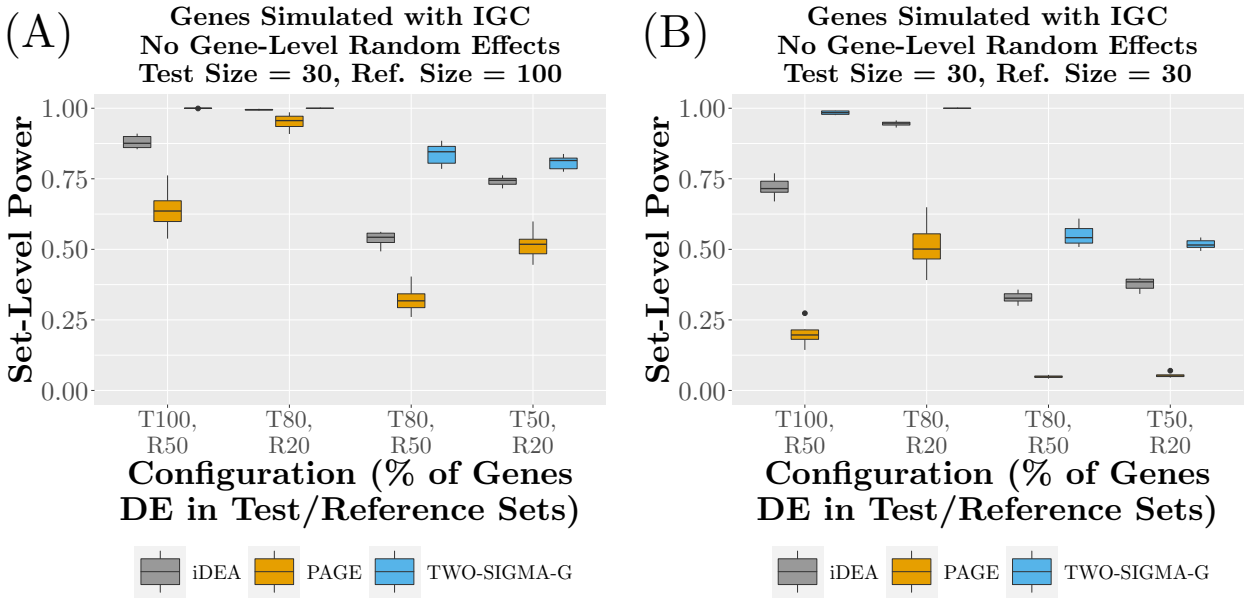
Supplementary Figure S5: **Type-I error performance of iDEA, PAGE, and TWO-SIGMA for various set-level null hypotheses for genes simulated with IGC**. Reference set sizes of 100 and 30 are shown. Each boxplot aggregates six different settings which vary both the magnitude of the average inter-gene correlation in the test set and the nature of the correlation structure via the introduction of other individual-level covariates. Such settings are intended to represent the diversity seen in real data sets to paint an accurate picture of testing properties over a wide range of gene sets. Each of the six settings is further composed of 10 replicates which vary only random seed to mimic the impact of a different starting pool of cells from which genes were simulated. See the Methods section of the main text for more details regarding the simulation procedure.



Supplementary Figure S6: **Set-level power of CAMERA, MAST, and TWO-SIGMA-G using a reference set size of 100 genes (corresponds to Figure 3 in main text)**. Each panel varies the existence of IGC between genes in the test set and the presence of gene-level random effect terms in gene-level model (CAMERA never includes gene-level random effect terms). Scenarios along the x-axis of each panel vary the percentage of genes that are differentially expressed (with the same effect size) in the test and reference sets. For example, "T80,R50" corresponds to the configuration under the alternative hypothesis in which 80% of test set genes are DE and 50% of reference set genes are DE. Each boxplot aggregates six different settings which vary both the magnitude of the average inter-gene correlation in the test set and the nature of the correlation structure via the introduction of other individual-level covariates. Such settings are intended to represent the diversity seen in real data sets to paint an accurate picture of testing properties over a wide range of gene sets. Each of the six settings is further composed of 10 replicates which vary only random seed to mimic the impact of a different starting pool of cells from which genes were simulated. See the Methods section of the main text for more details regarding the simulation procedure.

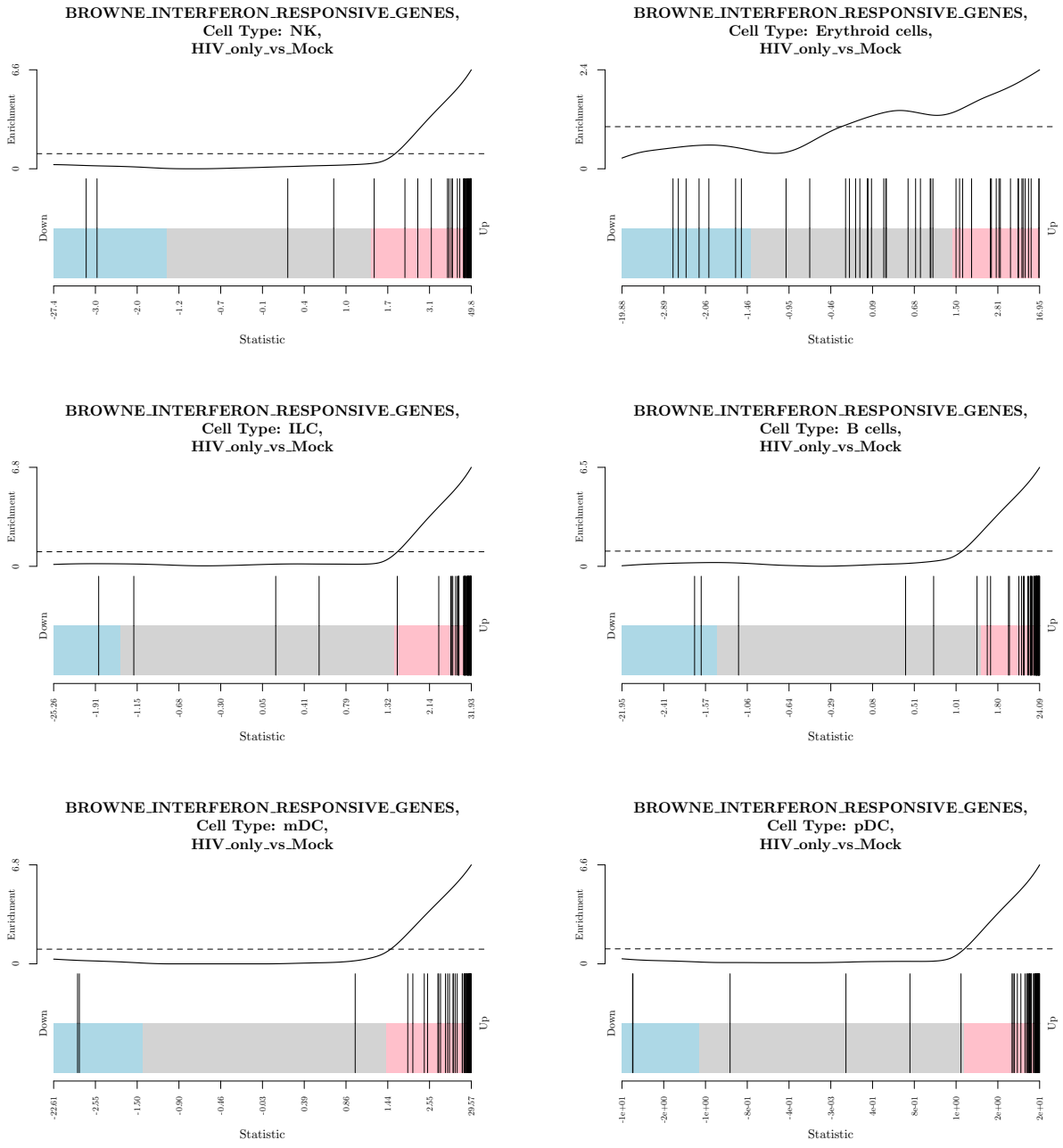


Supplementary Figure S7: **Set-level power of CAMERA, MAST, and TWO-SIGMA-G using differing DE magnitudes at the gene-level**. Genes are simulated with IGC, and reference set sizes of 100 and 30 are used for gene set testing. Scenarios along the x-axis of each panel vary the percentage of genes that are differentially expressed (with the same effect size) in the test and reference sets. For example, "T80,R50" corresponds to the configuration under the alternative hypothesis in which 80% of test set genes are DE and 50% of reference set genes are DE. Within some test sets, the amount of DE is mixed: with 50% of genes having twice as large of an effect size as the other half (see, e.g., "T100M, R50"). Each boxplot aggregates six different settings which vary both the magnitude of the average inter-gene correlation in the test set and the nature of the correlation structure via the introduction of other individual-level covariates. Such settings are intended to represent the diversity seen in real data sets to paint an accurate picture of testing properties over a wide range of gene sets. Each of the six settings is further composed of 10 replicates which vary only random seed to mimic the impact of a different starting pool of cells from which genes were simulated. See the Methods section of the main text for more details regarding the simulation procedure.



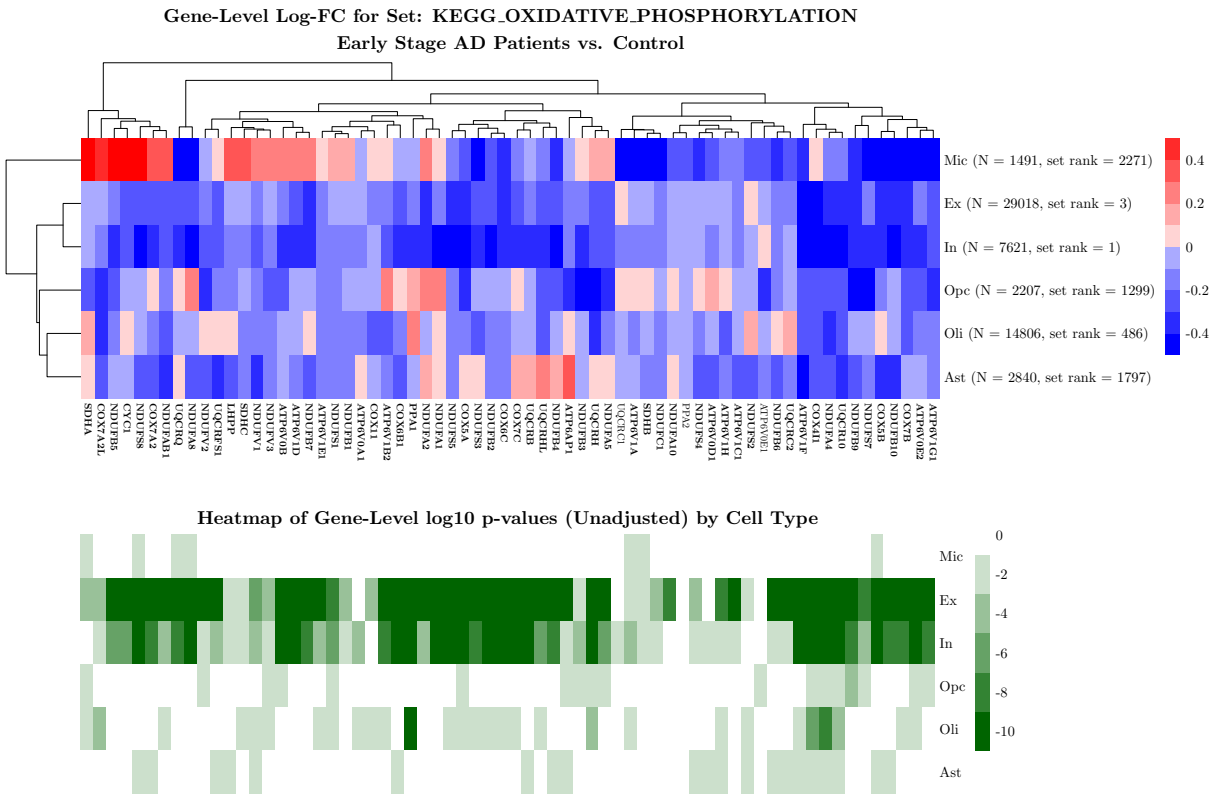
Supplementary Figure S8: **Set-level power of iDEA, PAGE, and TWO-SIGMA-G using differing DE magnitudes at the gene-level**. Genes are simulated with IGC using reference set sizes of 100 and 30. Scenarios along the x-axis of each panel vary the percentage of genes that are differentially expressed (with the same effect size) in the test and reference sets. For example, "T80,R50" corresponds to the configuration under the alternative hypothesis in which 80% of test set genes are DE and 50% of reference set genes are DE. Because iDEA performed poorly in scenarios involving "R0", they were excluded. Each boxplot aggregates six different settings which vary both the magnitude of the average inter-gene correlation in the test set and the nature of the correlation structure via the introduction of other individual-level covariates. Such settings are intended to represent the diversity seen in real data sets to paint an accurate picture of testing properties over a wide range of gene sets. Each of the six settings is further composed of 10 replicates which vary only random seed to mimic the impact of a different starting pool of cells from which genes were simulated. See the Methods section of the main text for more details regarding the simulation procedure.

S2 Additional HIV Data Results

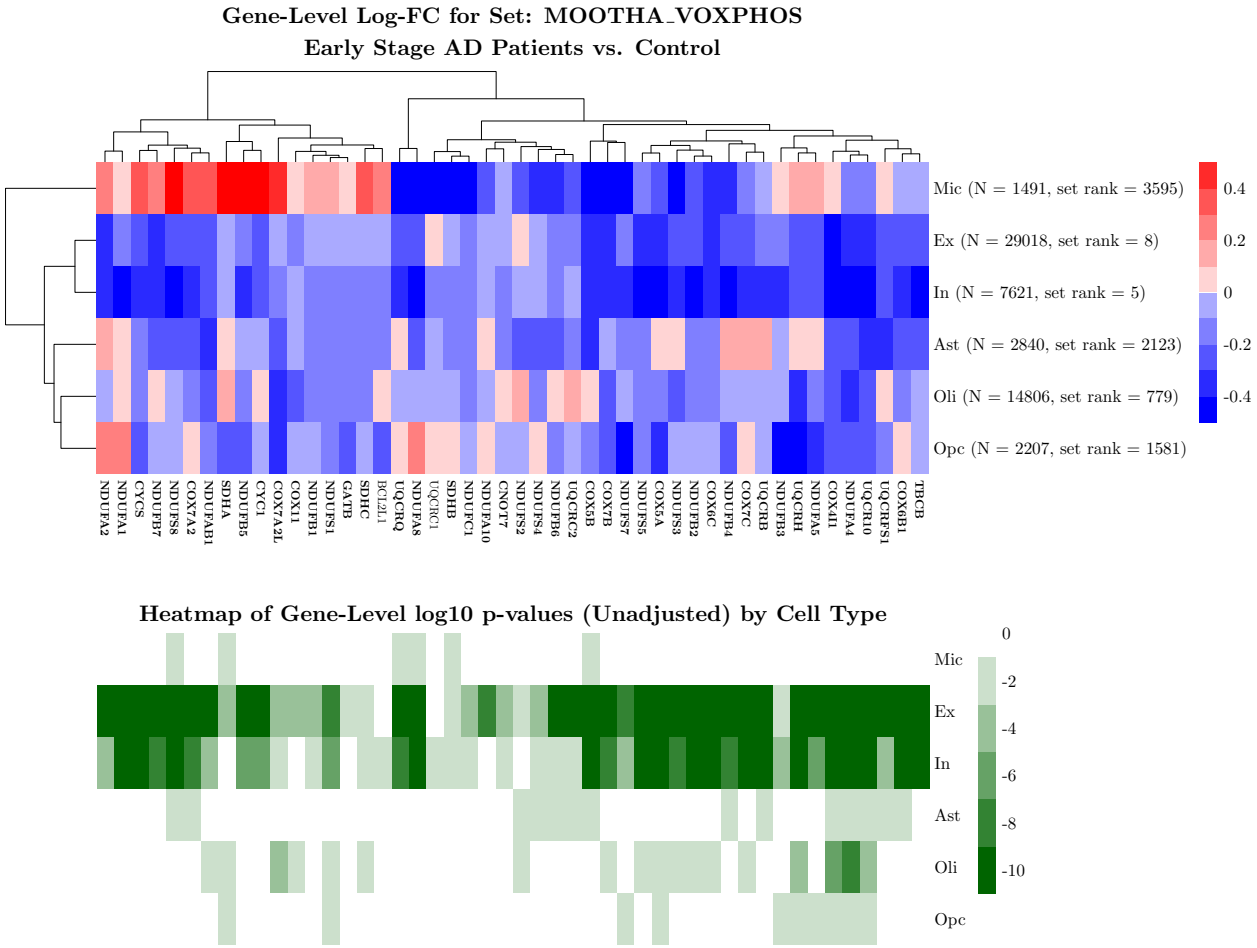


Supplementary Figure S9: Barcode plots showing gene-level statistics for the six most common cell types in the HIV dataset for the gene set "BROWNE_INTERFERON_RESPONSIVE_GENES." The x-axis shows the concentration of gene-level statistics (each bar represents one gene-level statistic), and the enrichment shown in the y-axis demonstrates that gene-level statistics tend to be overwhelmingly positive and large in magnitude for all cell types except Erythroid cells.

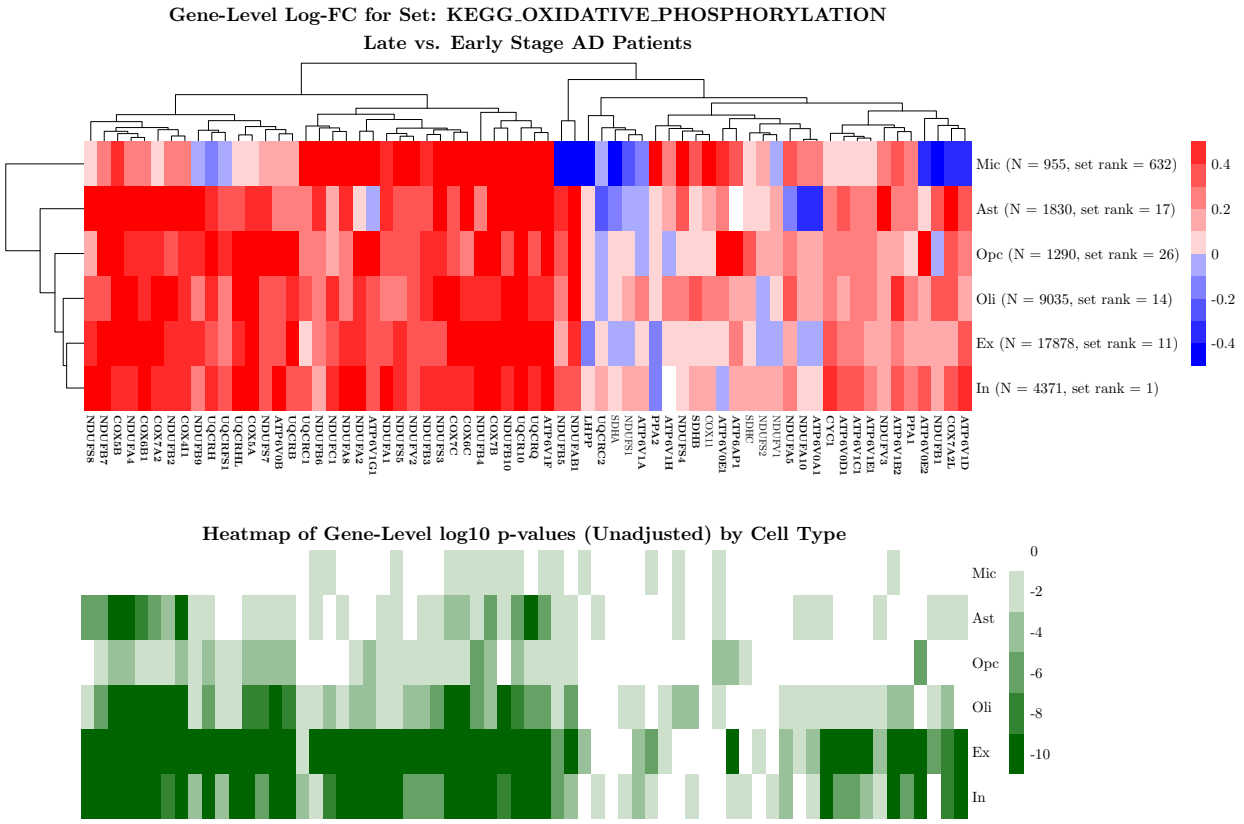
S3 Comparing Early Stage AD Patients to Control



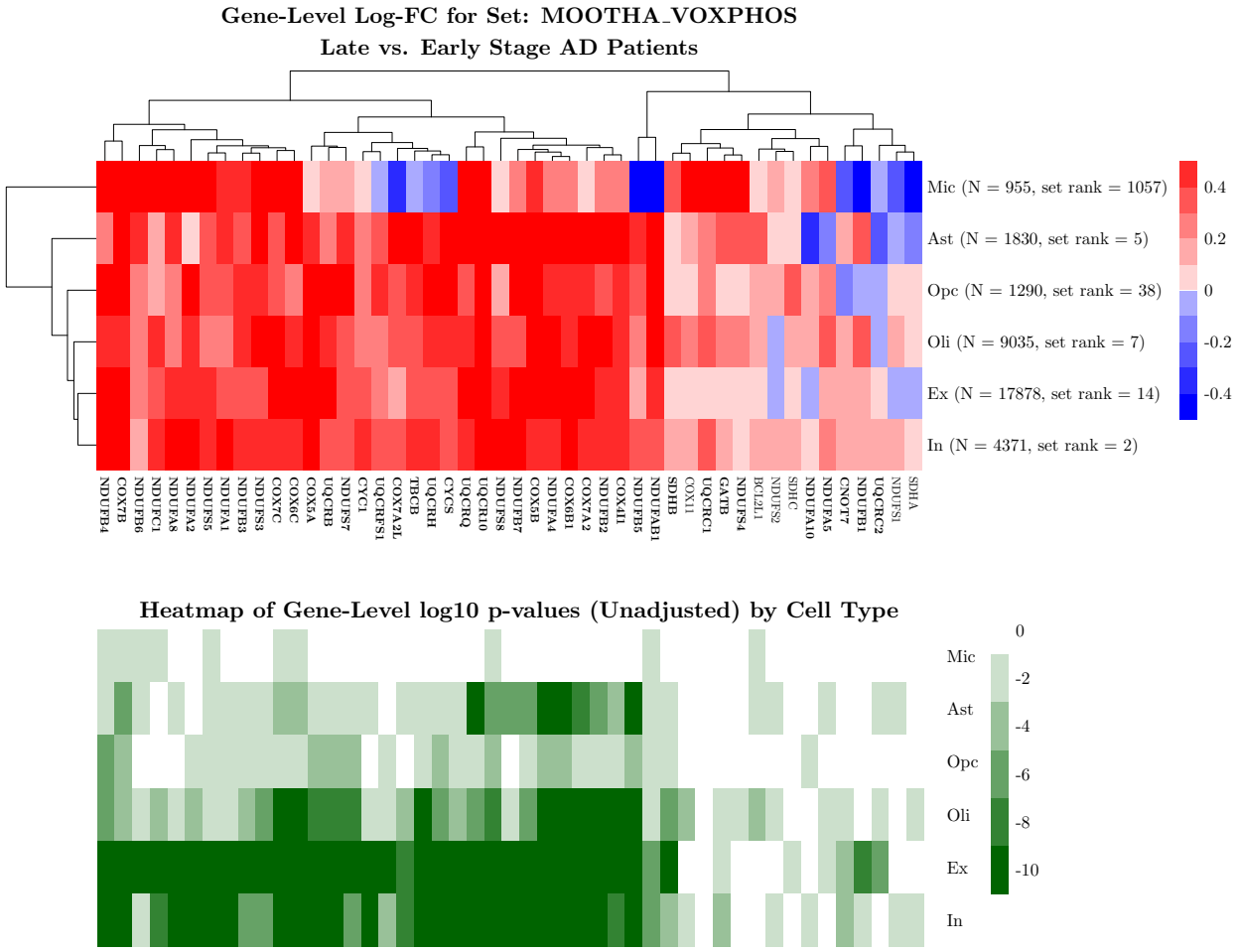
Supplementary Figure S10: **Cell-type specific variation in gene-level significance for genes in the KEGG_OXIDATIVE_PHOSPHORYLATION pathway comparing early stage AD patients to controls**. Gene names that are bolded are signifcant over all cell types after FDR-adjustment of the Fisher's method p-value.



Supplementary Figure S11: **Cell-type specific variation in gene-level significance for genes in the MOOTHA_VOXPHOS pathway comparing early stage AD patients to controls**. Gene names that are bolded are signifcant over all cell types after FDR-adjustment of the Fisher's method p-value.

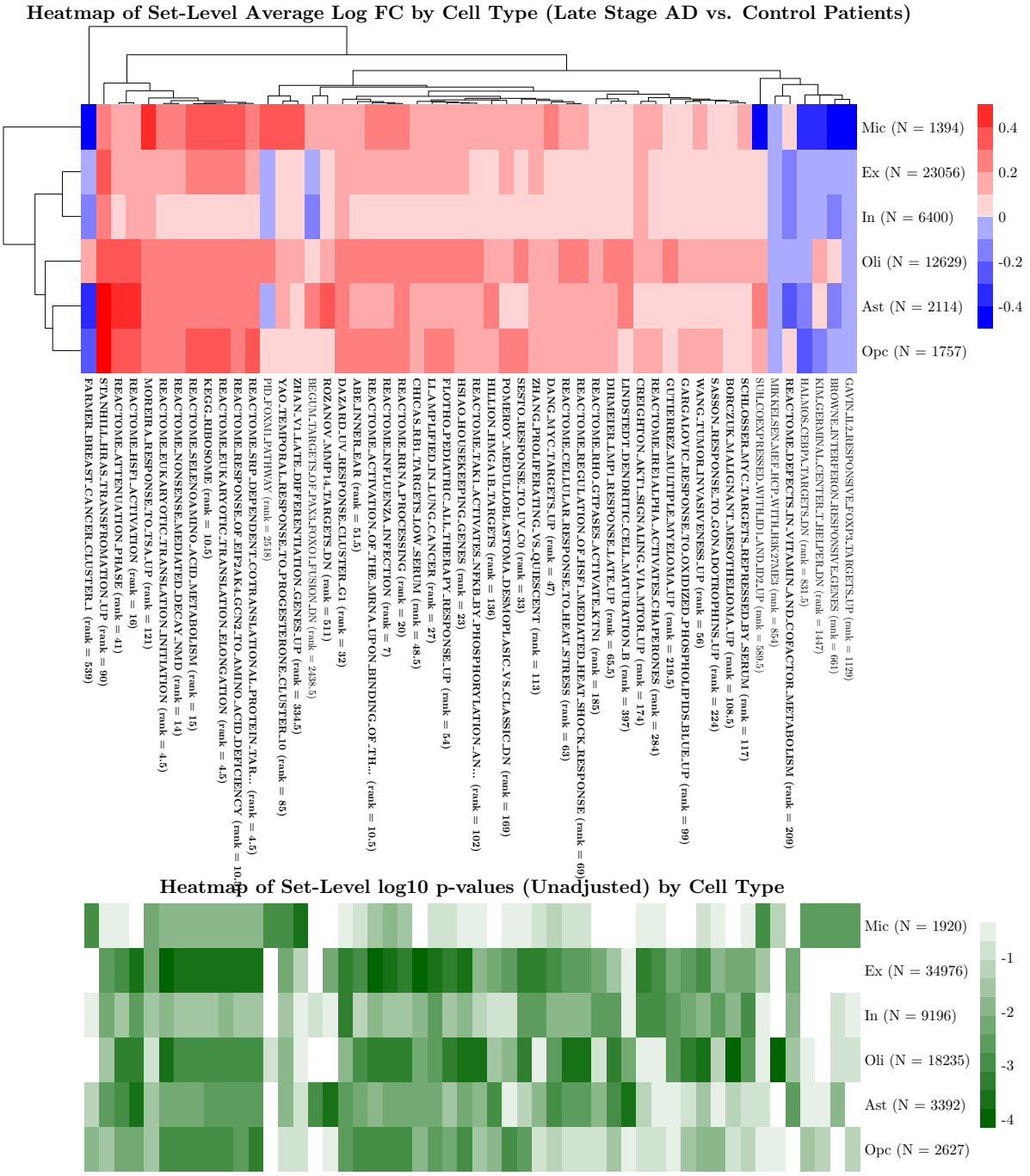


Supplementary Figure S12: **Cell-type specific variation in gene-level significance for genes in the KEGG_OXIDATIVE_PHOSPHORYLATION pathway comparing late stage AD patients to early stage AD patients**. Gene names that are bolded are signifcant over all cell types after FDR-adjustment of the Fisher's method p-value.

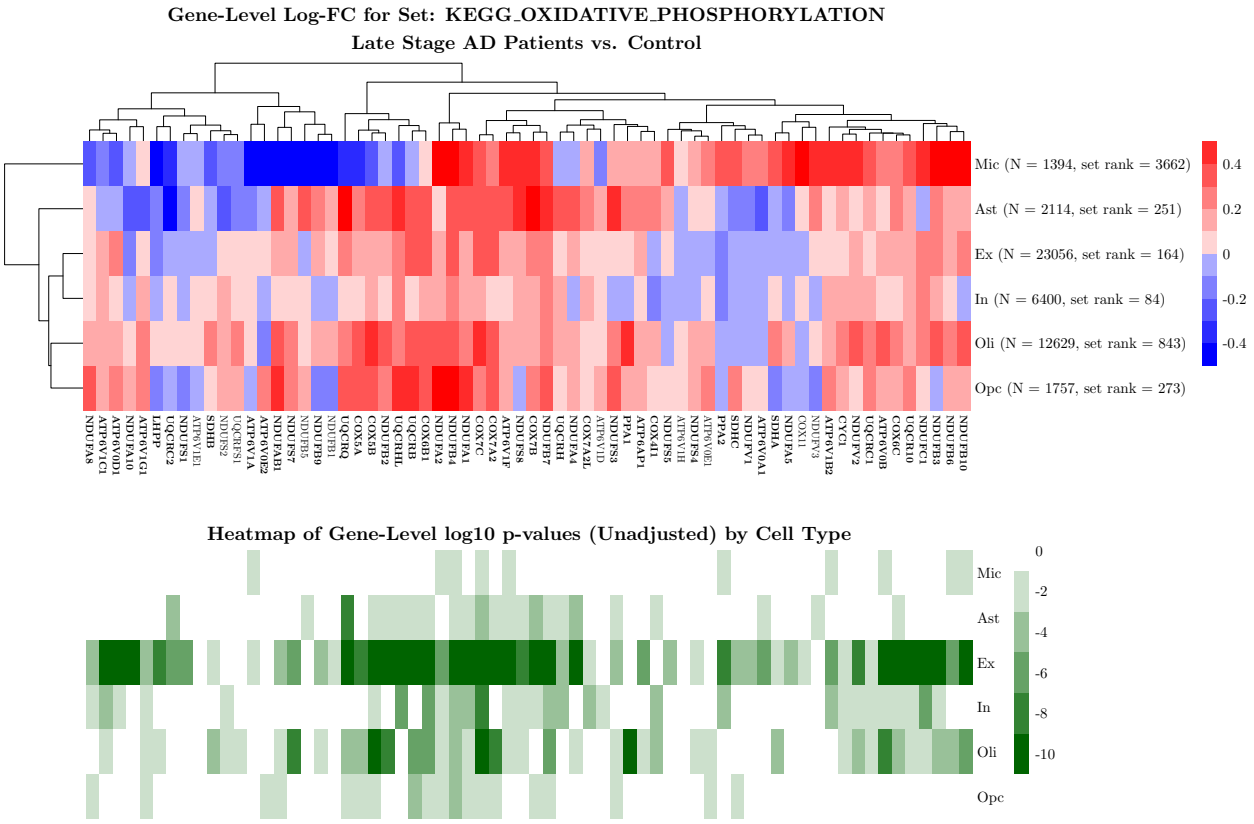


Supplementary Figure S13: **Cell-type specific variation in gene-level significance for genes in the MOOTHA_VOXPHOS pathway comparing late stage AD patients to early stage AD patients**. Gene names that are bolded are signifcant over all cell types after FDR-adjustment of the Fisher's method p-value.

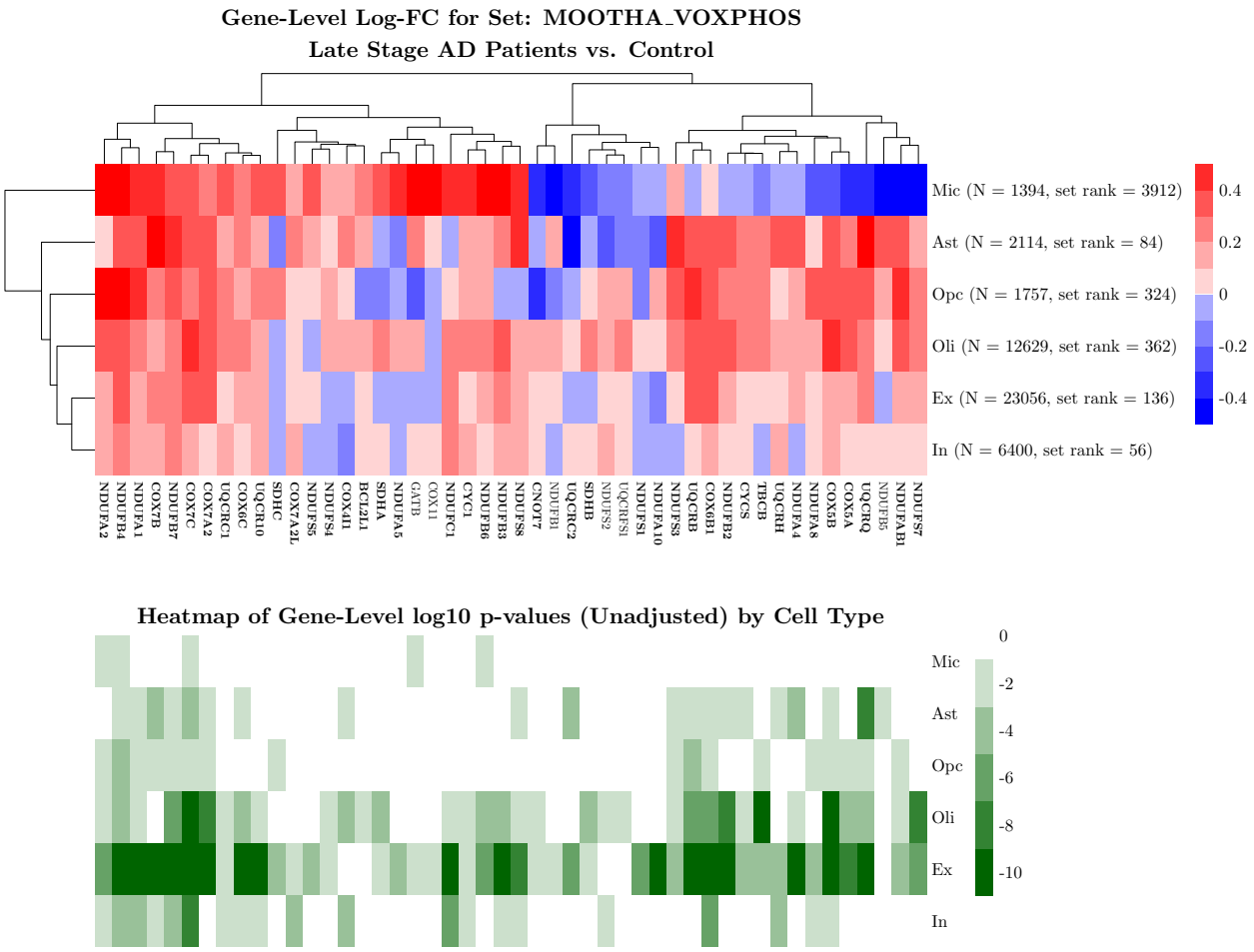
S4 Comparing Late Stage AD Patients to Control



Supplementary Figure S14: **Heatmap of the most significant gene sets (and their corresponding p-values) comparing late state AD patients to controls by cell type**. Sets plotted are among the top 10 in significance for at least once cell type. Sets in bold are significant over all cell types after FDR-adjustment of the Fisher's method p-value.

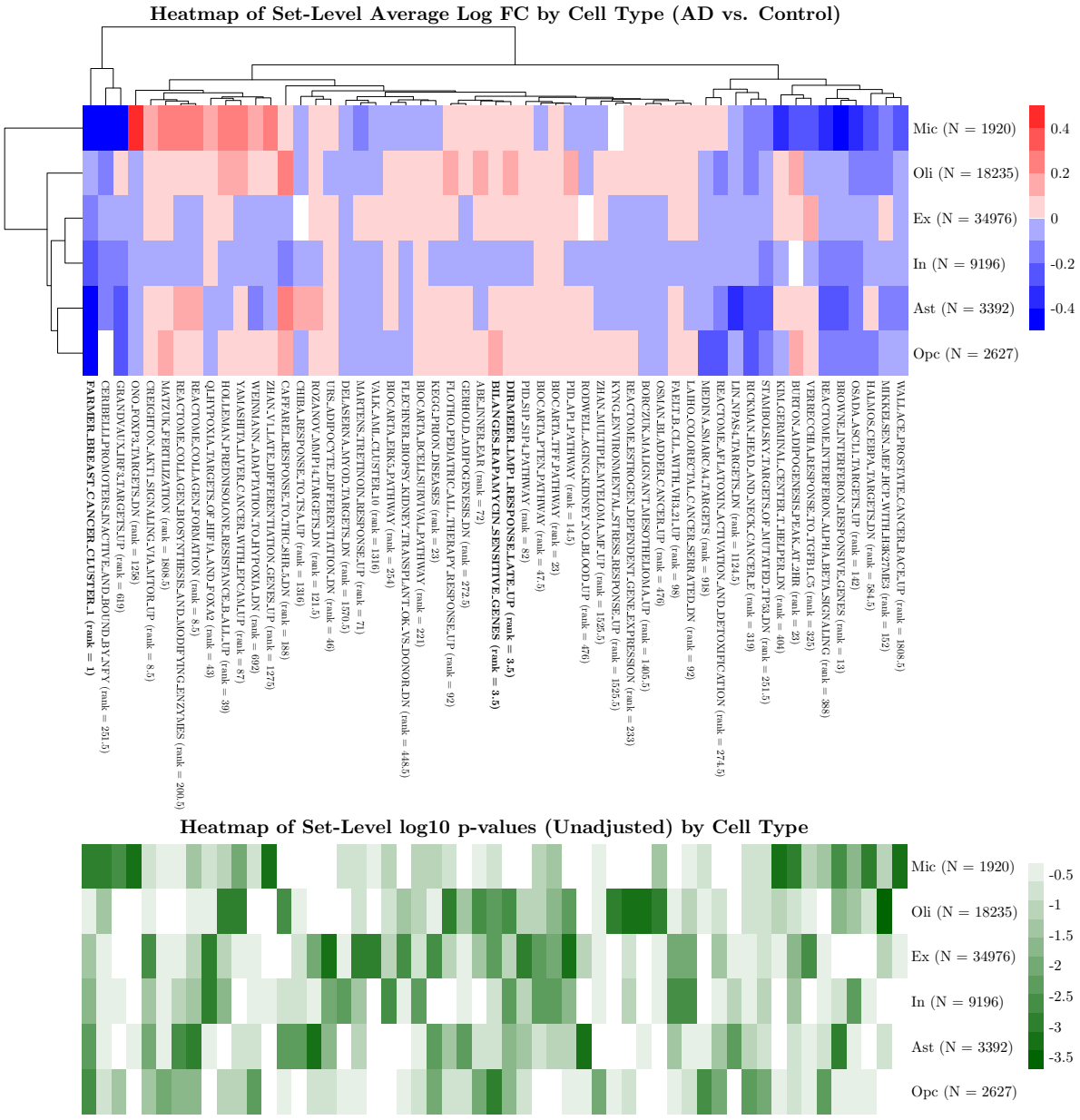


Supplementary Figure S15: **Cell-type specific variation in gene-level significance for genes in the KEGG_OXIDATIVE_PHOSPHORYLATION pathway comparing late stage AD patients to controls**. Gene names that are bolded are signifcant over all cell types after FDR-adjustment of the Fisher's method p-value.

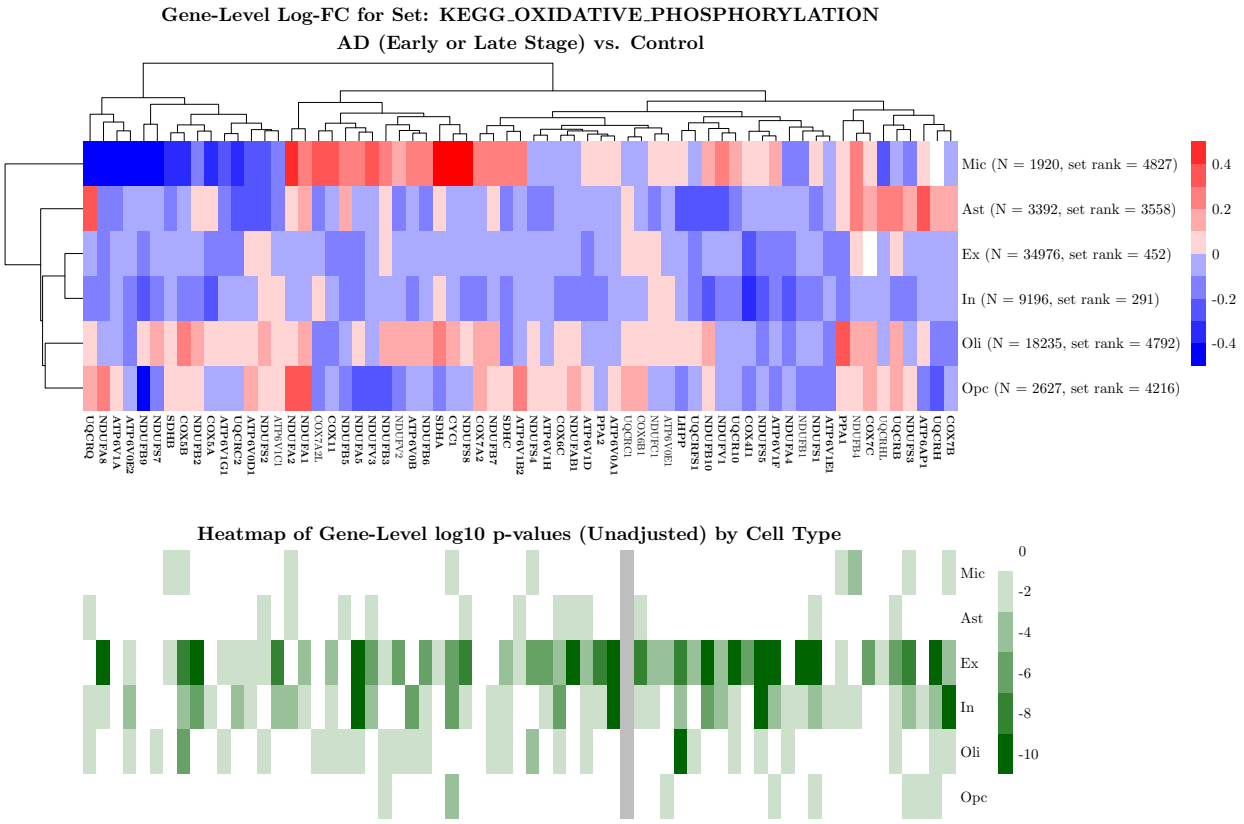


Supplementary Figure S16: **Cell-type specific variation in gene-level significance for genes in the MOOTHA_VOXPHOS pathway comparing late stage AD patients to controls**. Gene names that are bolded are signifcant over all cell types after FDR-adjustment of the Fisher's method p-value.

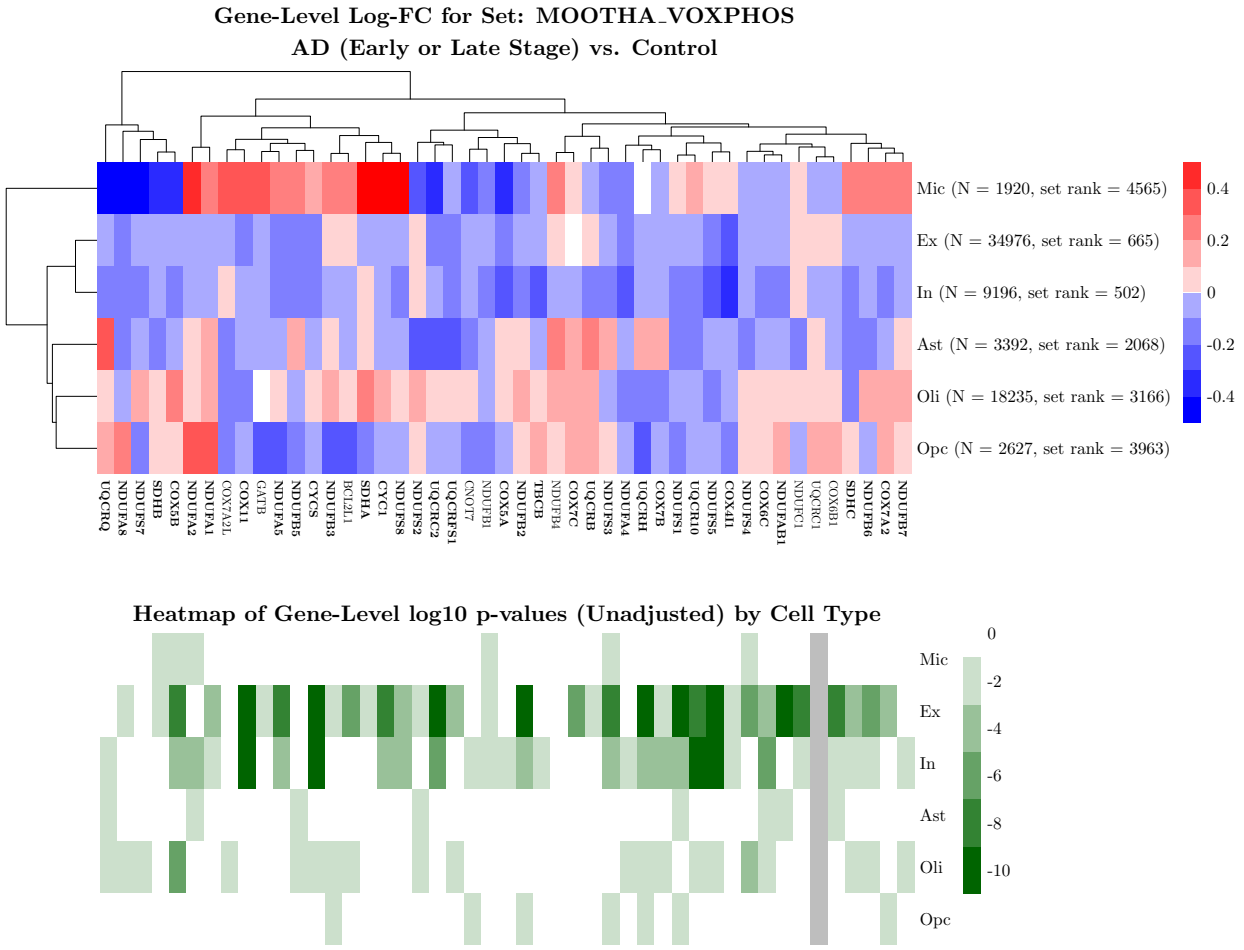
S5 Comparing AD Patients (Early and Late Stage) to Control



Supplementary Figure S17: **Heatmap of the most significant gene sets (and their corresponding p-values) comparing AD patients (early and late stage) to controls by cell type**. Sets plotted are among the top 10 in significance for at least once cell type. Sets in bold are significant over all cell types after FDR-adjustment of the Fisher's method p-value.

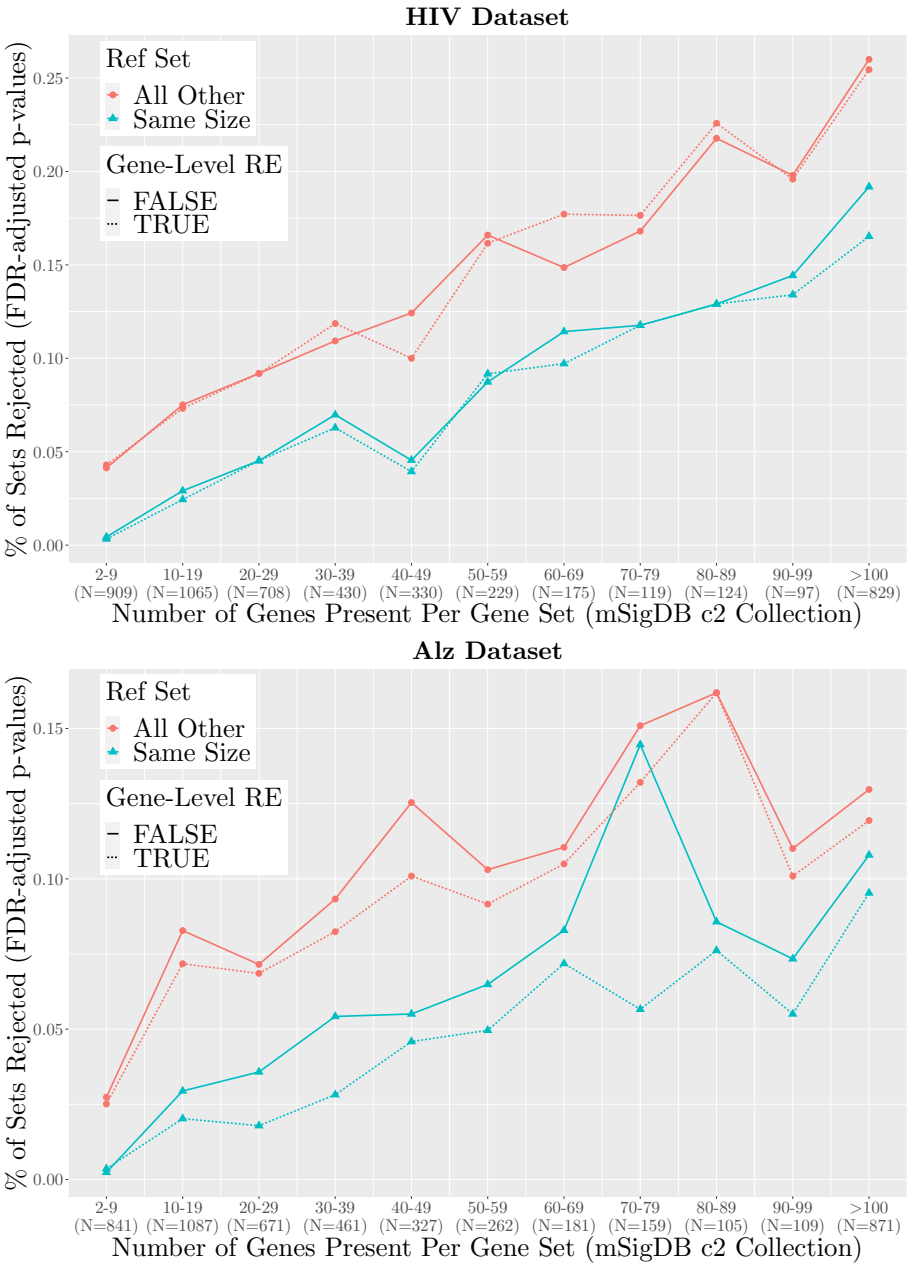


Supplementary Figure S18: **Cell-type specific variation in gene-level significance for genes in the KEGG_OXIDATIVE_PHOSPHORYLATION pathway comparing AD patients (early and late stage) to controls**. Gene names that are bolded are signifcant over all cell types after FDR-adjustment of the Fisher's method p-value.

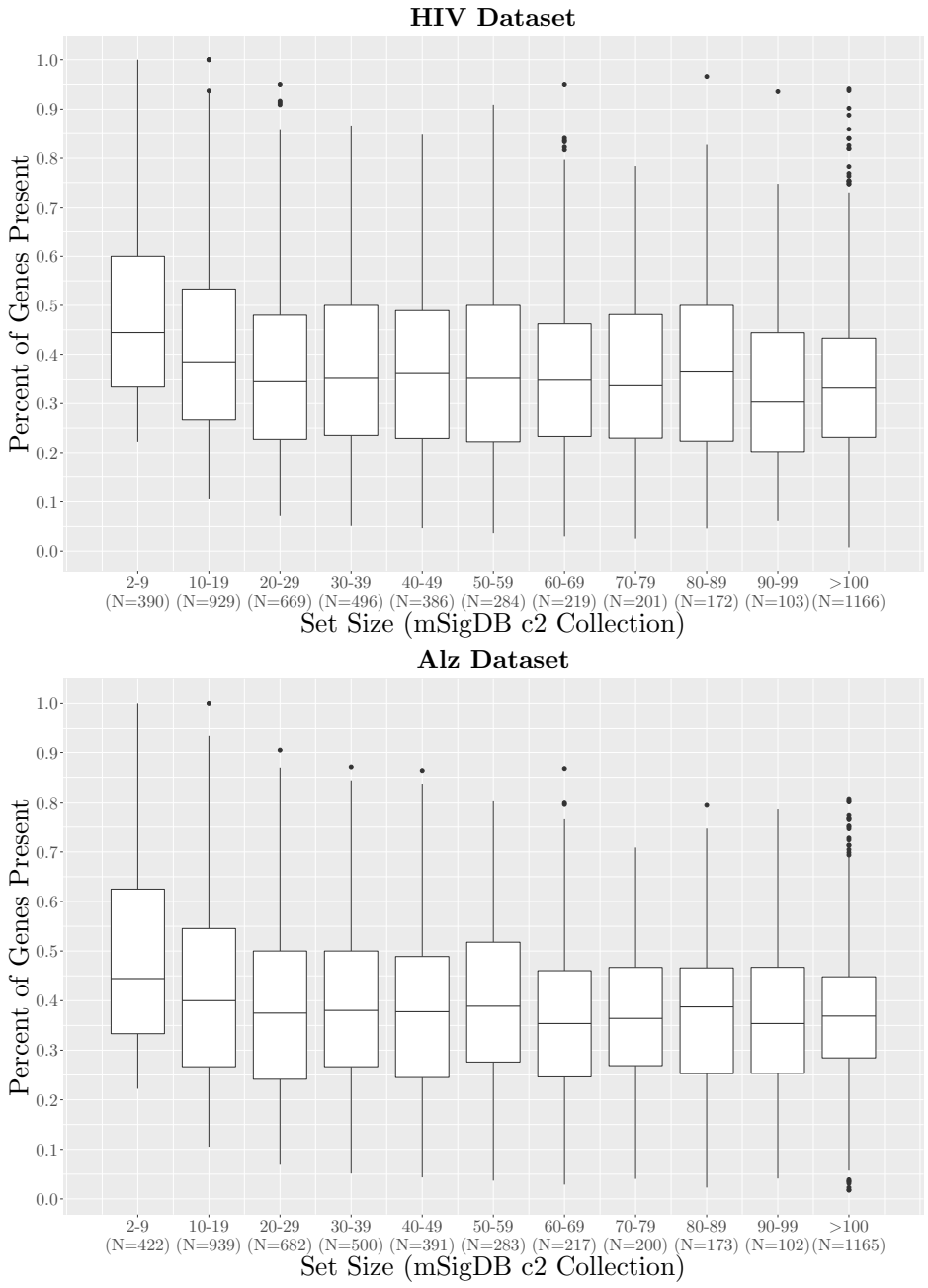


Supplementary Figure S19: **Cell-type specific variation in gene-level significance for genes in the MOOTHA_VOXPHOS pathway comparing AD patients (early and late stage) to controls**. Gene names that are bolded are signifcant over all cell types after FDR-adjustment of the Fisher's method p-value.

S6 Additional Real Data Figures



Supplementary Figure S20: **Percentage of sets rejected in the HIV and Alzheimer's datasets**. Fisher's-method p-values adjusted for FDR were used for testing in four settings varying the choice of reference set between the complement set of genes ("All Other") or a random reference of the same size as the test set ("Same Size"), and with and without random effects present at the gene-level. The presence of gene-level random effects in the model does not greatly affect the percentage of sets rejected in either the HIV dataset (top) or the Alzheimer's dataset (bottom).



Supplementary Figure S21: **Percentage of genes present by set size in the HIV dataset (top) and the Alzheimer's dataset (bottom).**