The American Journal of Human Genetics, Volume *89*

Supplemental Data

Integrating Autoimmune Risk Loci with Gene-Expression Data Identifies Specific Pathogenic Immune Cell Subsets

Xinli Hu, Hyun Kim, Eli Stahl, Robert Plenge, Mark Daly, and Soumya Raychaudhuri

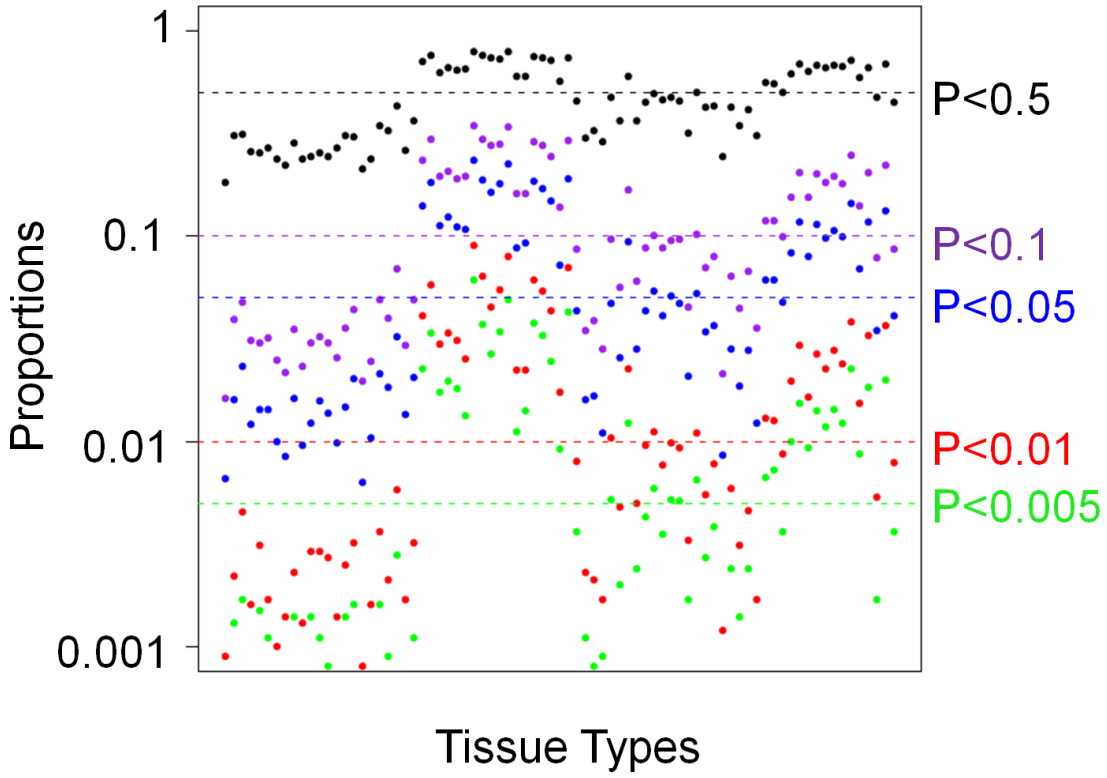

Figure S1. Analytical p-Values, Calculated without Permutations Show Inflation in Neural Tissues

We evaluated 10,000 sets of 20 random independent SNPs across the genome for tissue specific gene enrichment using the GNF human tissue expression data set. The subset of tissues enriched in the middle demonstrating marked inflation are all central or peripheral nervous system tissues. Here we present analytical p-values using the gamma distribution (see **Methods**).

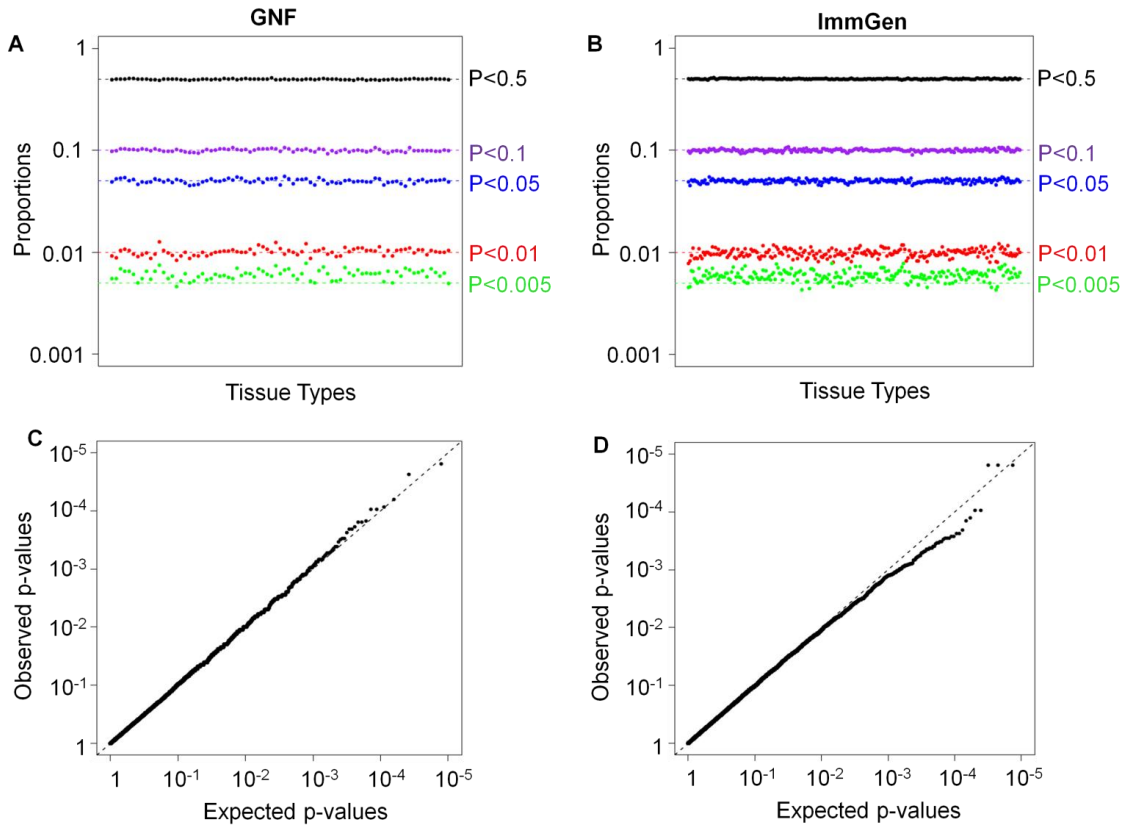

Figure S2. Type I Error Estimates for Statistical Approach

(A) To test the statistical properties of our approach, we selected 10,000 random SNP sets of 20 independent SNPs across the genome. We tested 79 tissues expression profiles from the GNF dataset for enrichment of specifically expressed genes. For each tissue type, we plotted the proportion of sets that obtained specific p-value thresholds. (B) Similar results for the 223 tissue profiles in the ImmGen dataset. (C) We tested 500 random SNP sets and tested 79 tissues expression profiles from the GNF dataset for enrichment. After aggregating p-values for all tissues, we plotted observed p-values as a function of expected p-values in a Q-Q plot. D) Similar results for the ImmGen dataset.

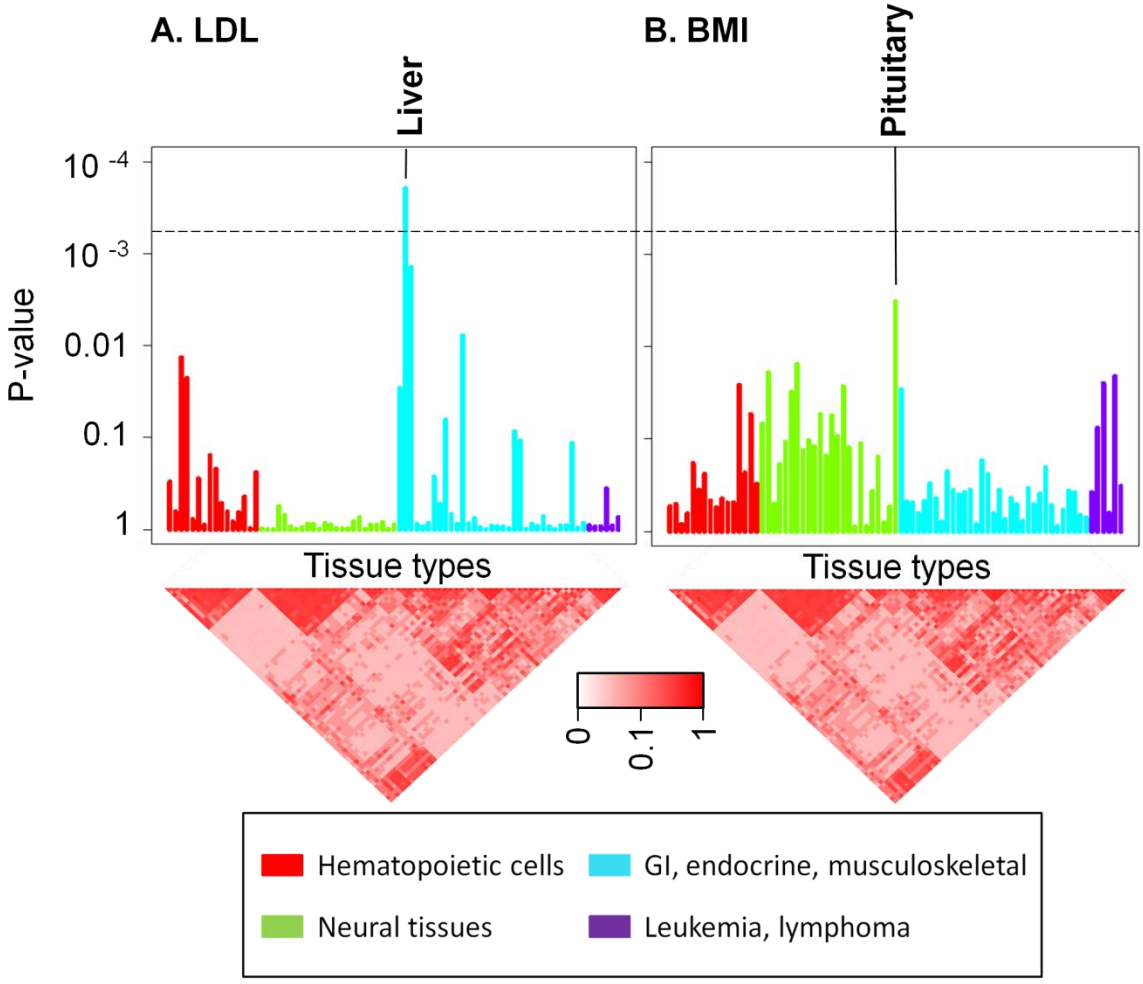

Figure S3. Application to Metabolic Traits

(A) 37 SNPs associated with LDL metabolism were evaluated for gene enrichment in 79 human tissue types. The Bonferroni-corrected *p*-value is shown by a dotted line. Only the liver showed statistically significant specific expression of genes in LD with cholesterol metabolism SNPs (p= 1.95x10^{-4}). We have plotted a heat map along the bottom to depict the *p*-value correlation between tissue types among random SNP sets. (B) 32 SNPs associated with obesity were evaluated for gene enrichment in 79 human tissue types. The pituitary achieved the most significant *p*-value (p=3.25x10^{-3}).

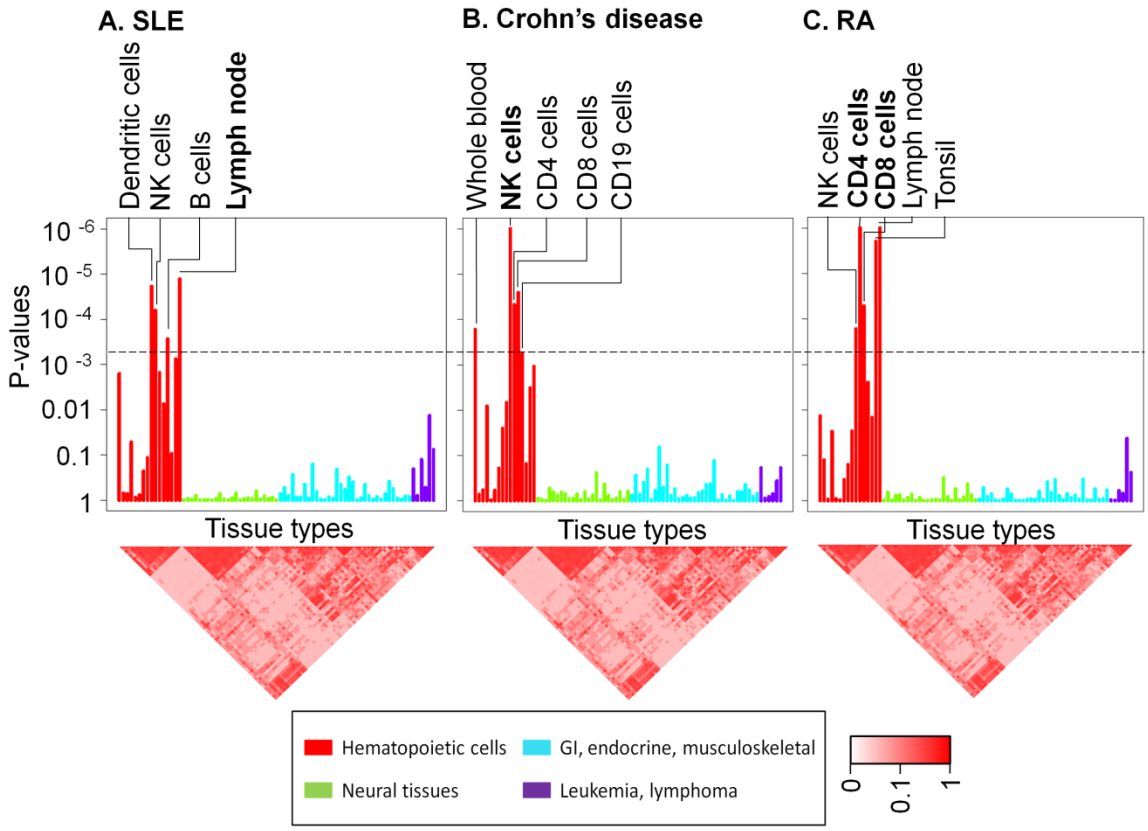

Figure S4. Application to Autoimmune Diseases

SNPs associated to SLE, Crohn's Disease, and RA evaluated for cell-specific gene enrichment in 79 human tissue types from the Novartis expression data. Subsets of hematopoietic cells showed statistically significant enrichment in each of the three diseases, while none of the other tissues showed significant enrichment. A) In SLE, B-cells, dendritic cells, NK cells, and lymph node tissue showed significant enrichment p-values after adjusting for multiple hypothesis testing. B) NK cells, CD8+ and CD4+ T-cells, whole blood, as well as CD19+ cells achieved statistical significance in Crohn's Disease. C) In RA CD4+ T-cells, CD8+ T-cells, tonsil, lymph node, and NK cells showed significant enrichment p-values after adjusting for multiple hypothesis testing.

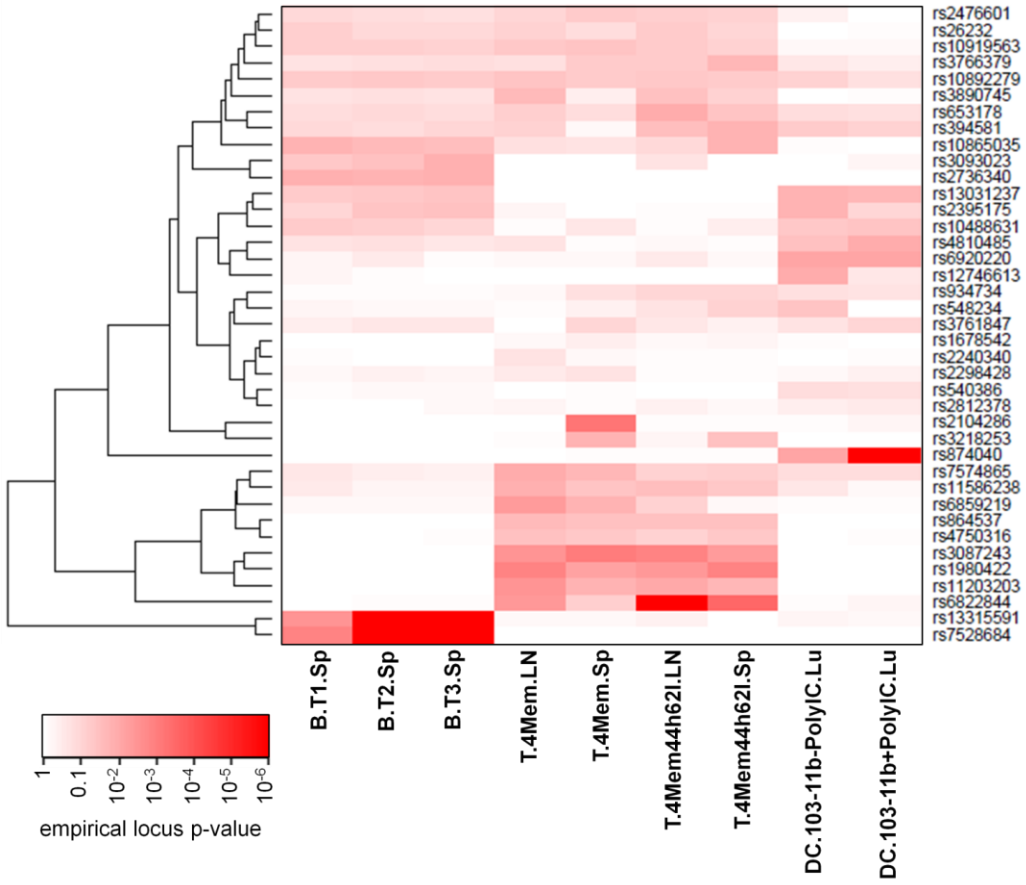

Figure S5. Patterns of Cell-Specific Expression of RA Loci

Here we plot the association between specific SNPs associated with RA (right) and selected tissues (bottom). Redness in each box correlates with the significance as measured by the empirical locus p-value; red boxes indicate that the SNP is in LD that is highly expressed in the tissue based on ImmGen, after accounting for the number of genes within the locus. SNPs are hierarchically clusters (left). Some of the SNPs toward the top are uninformative either because they lack a gene that is specifically expressed, or have genes with specific expression in multiple displayed cells. SNPs toward the bottom have the most informative expression patterns. A large number of the most informative SNPs have signal for the CD4+ effector memory cell subsets, but a small number have specific expression for transitional B-cells as well. These sets are mutually exclusive.