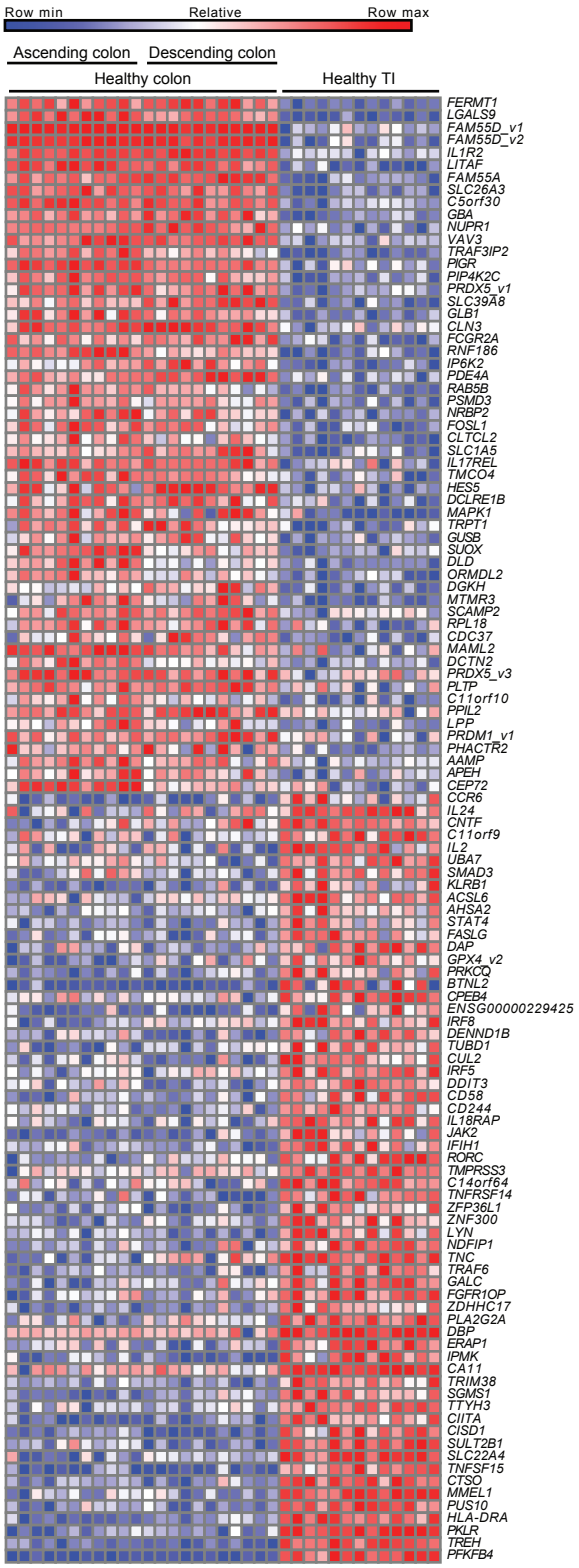
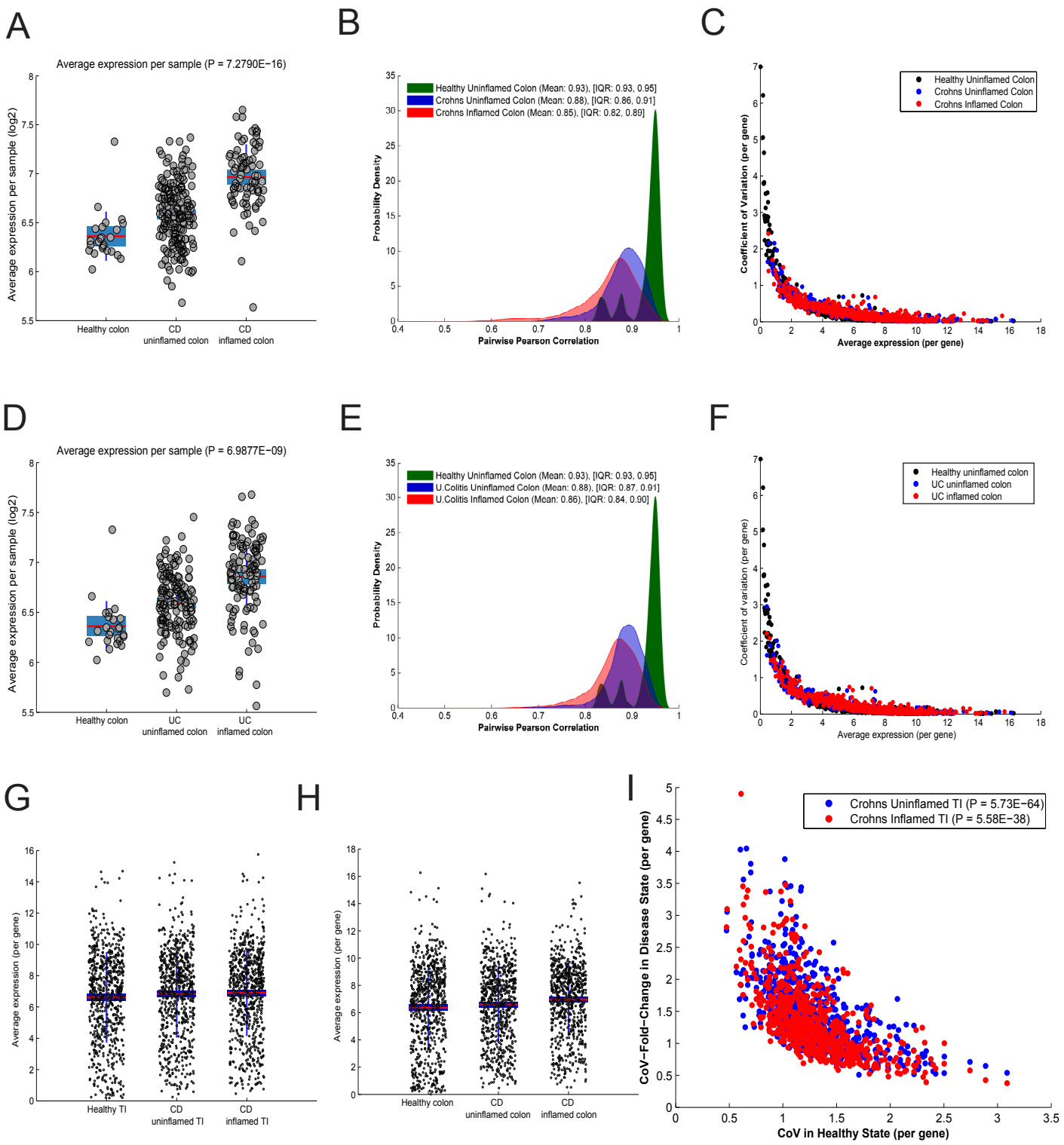
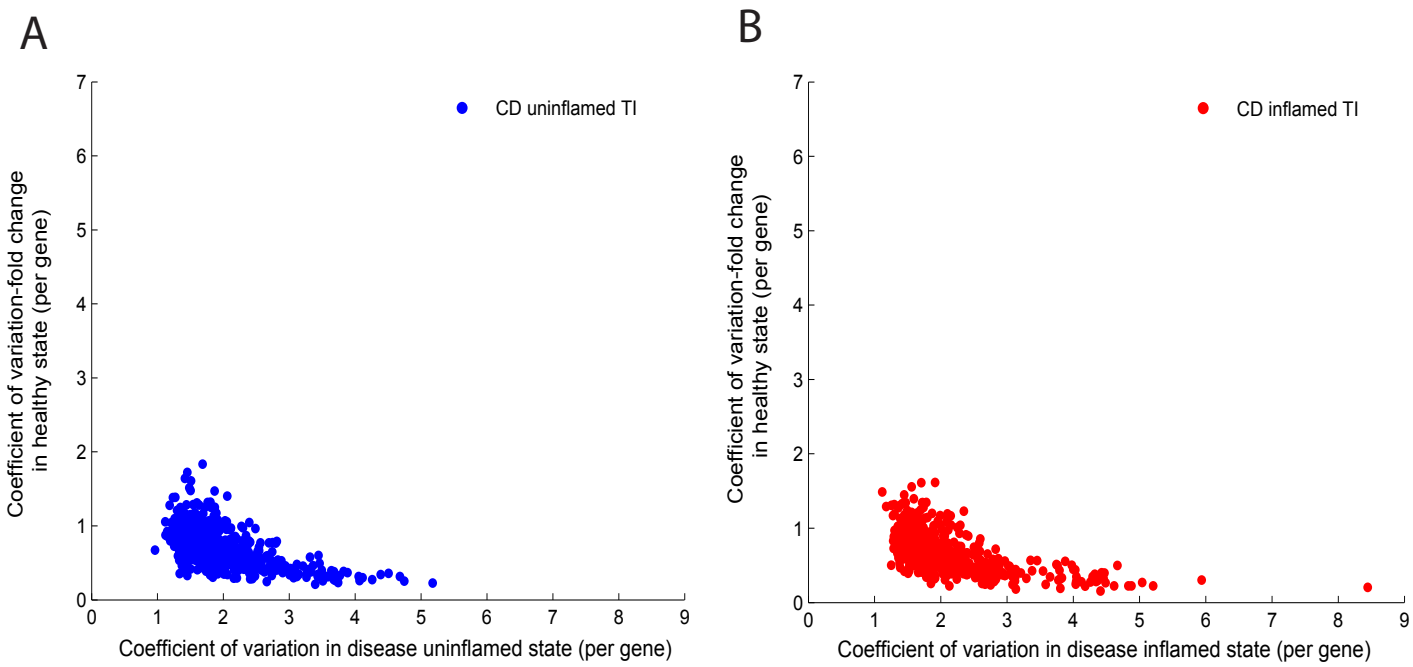


Supplemental Figure 1. Characterization of mucosal gene expression of IBD risk genes in healthy terminal ileum and colon tissues. In healthy controls, a unique set of genes are differentially expressed in the terminal ileum (TI) versus colon, but not between ascending and descending colon, suggesting tissue specificity between TI and colon, but not specific expression in the ascending or descending colon. The heatmap is colored in a row-normalized fashion, i.e., red indicates highest value for that specific gene, whereas blue indicates the lowest value for the same gene. Statistical significance of genes between TI samples (n=13) and colon samples (n=24) was estimated by computing the signal-to-noise ratio statistic. Genes with FDR adjusted p-value ≤ 0.05 and signal-to-noise ratio > 0.9 (absolute value) were selected as significant.

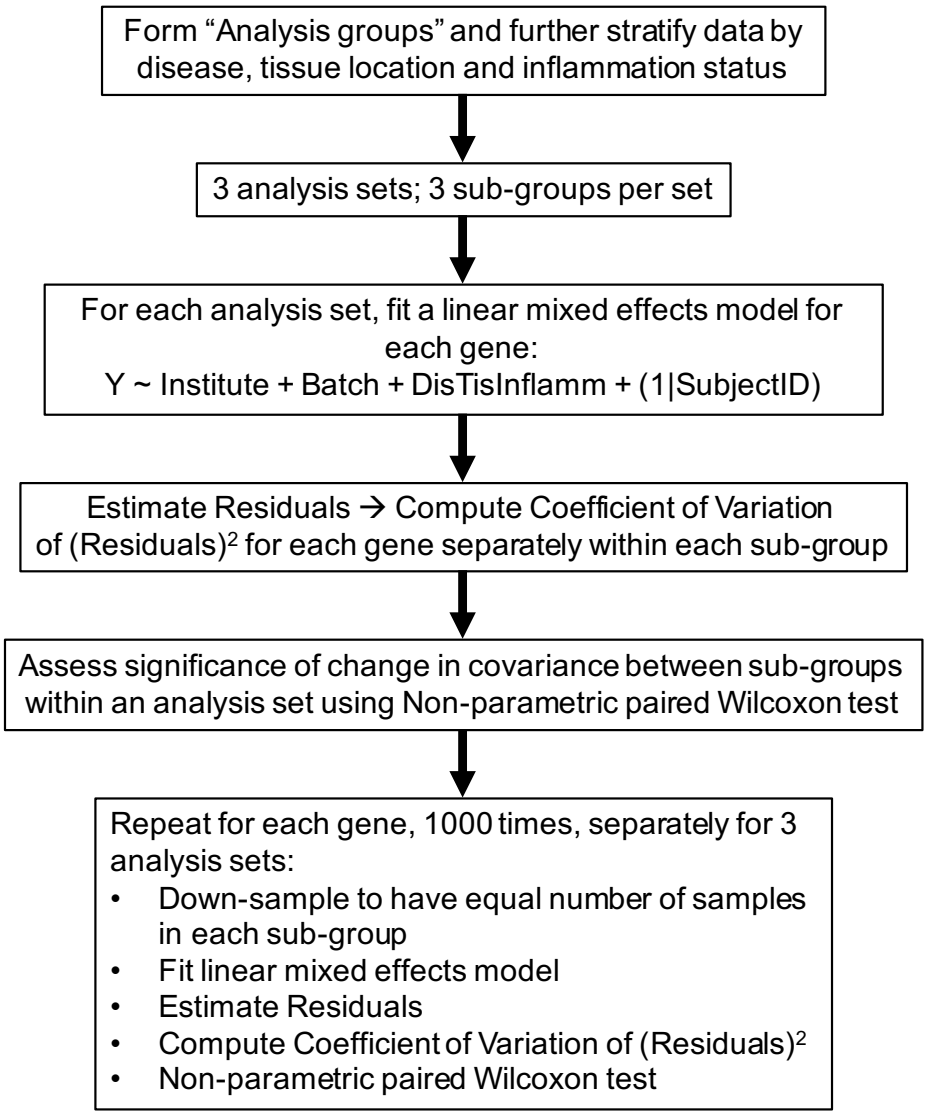


Supplemental Figure 2. Analysis of dispersion in gene expression with increasing disease inflammation. (A, D) Average expression (across genes) per sample in healthy uninflamed, CD/UC uninflamed, and CD/UC inflamed colon samples. (P-value assessed by a linear mixed effects model) (B, E) Pairwise Pearson correlation decreases from healthy controls to CD/UC uninflamed colon to CD/UC inflamed colon (mean, IQR) with associated decrease in pairwise correlation supporting increased variance in disease groups. (C, F) Within groups, genes with the highest average expression show the lowest coefficient of variation (per gene) in colon. (G-H) Global distribution of all genes within healthy or IBD patients is the same. (I) Genes with the lowest coefficient of variation in TI in the healthy state demonstrate the greatest fold-change coefficient of variation in the uninflamed or inflamed state in age-matched CD patients.



Supplemental Figure 3. The observed pattern of dysregulation of genes does not hold true when evaluating the fold-change in coefficient of variation in healthy samples, compared to the coefficient of variation in uninflamed disease state (A) or inflamed disease state (B) as baseline.

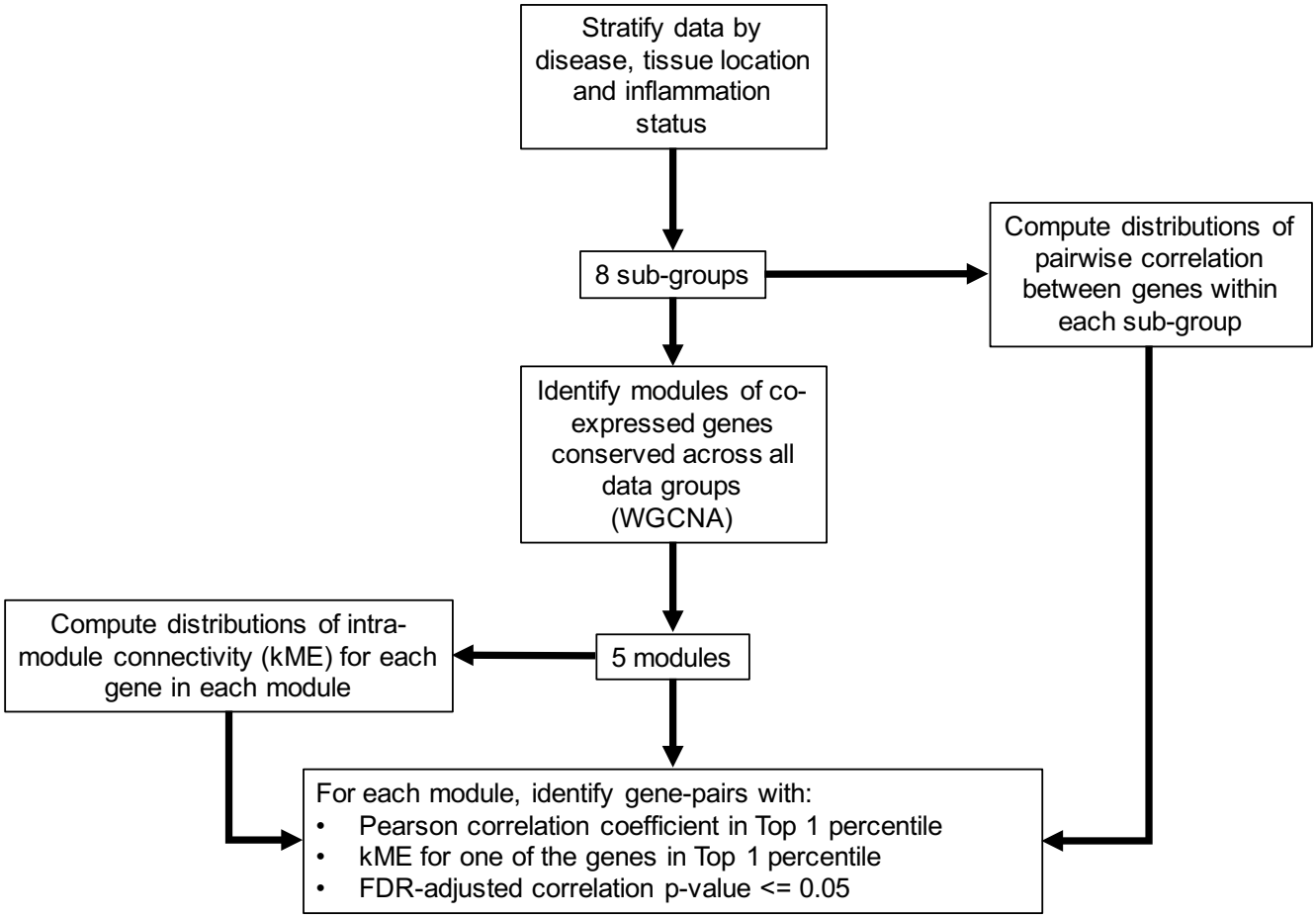
A



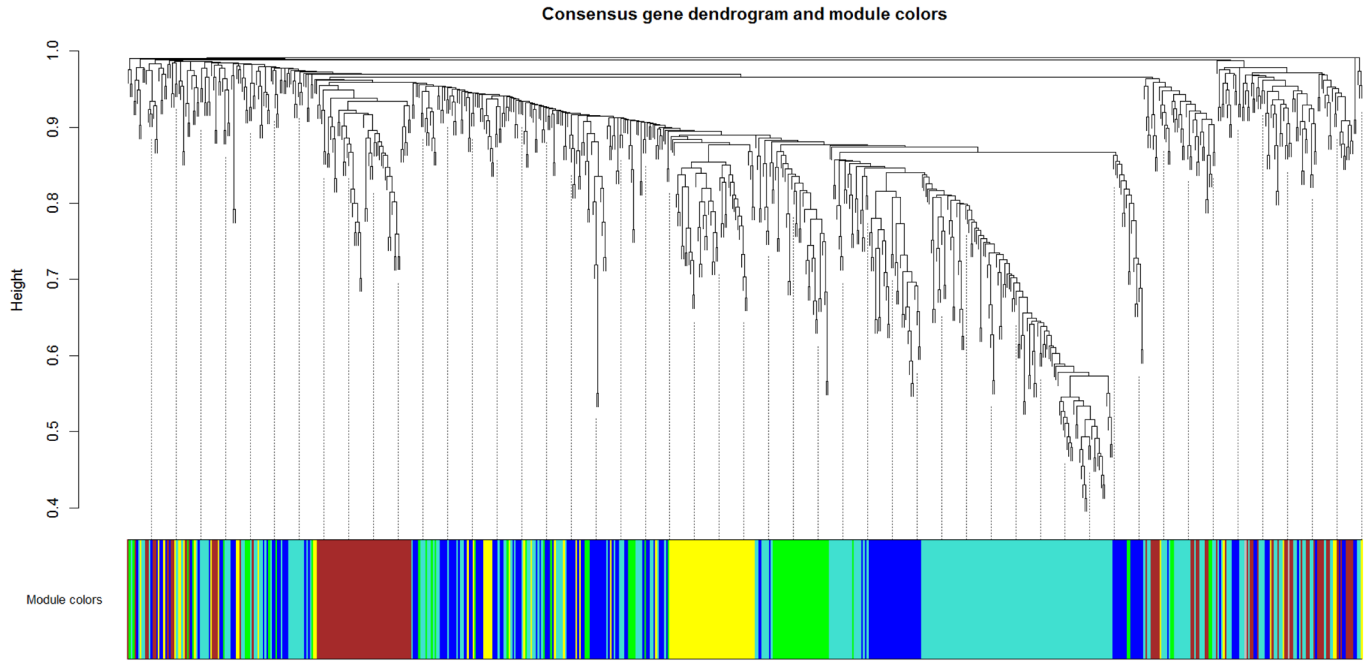
B

Groups	P-value	Robustness
CD-TI-Uninflmd vs HY-TI-Uninflmd	2.32E-92	100%
CD-TI-Inflamed vs HY-TI-Uninflmd	4.15E-98	100%
CD-Colon-Uninflmd vs HY-Colon-Uninflmd	1.05E-44	74.80%
CD-Colon-Inflamed vs HY-Colon-Uninflmd	2.36E-30	95.20%
UC-Colon-Uninflmd vs HY-Colon-Uninflmd	1.55E-21	3.30%
UC-Colon-Inflamed vs HY-Colon-Uninflmd	2.52E-04	0%

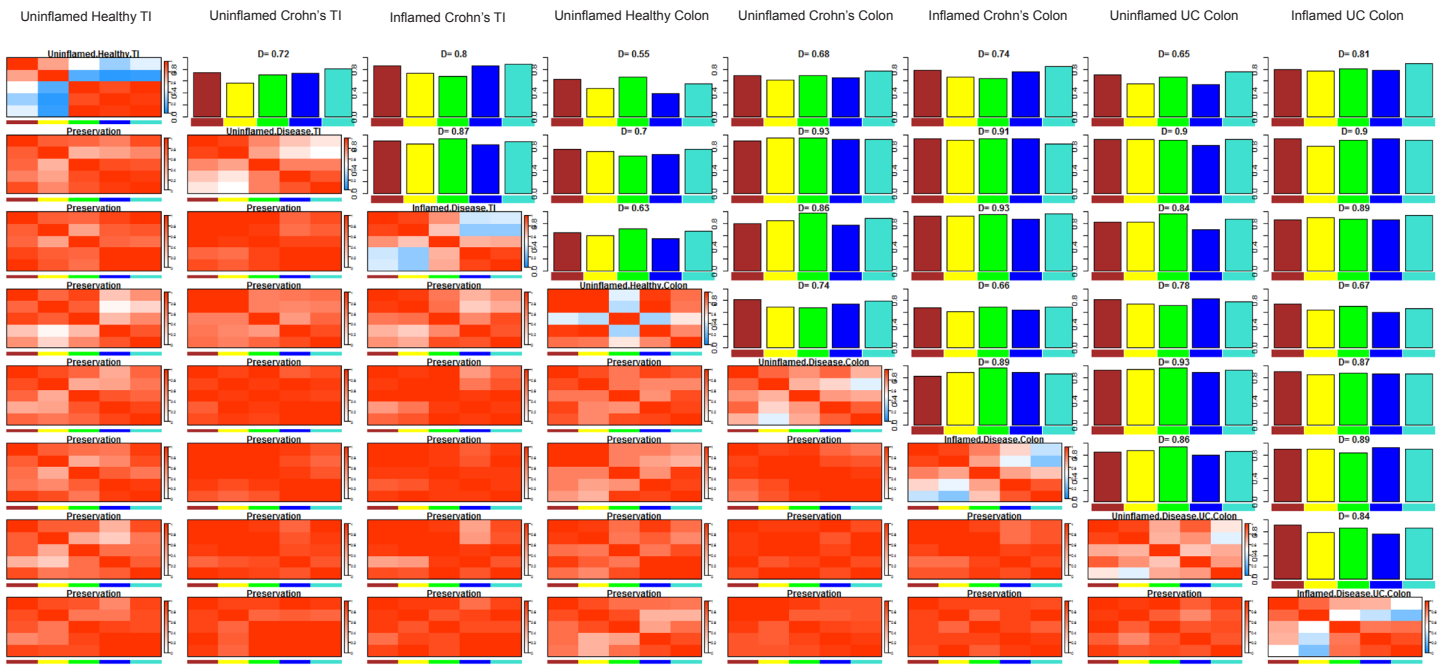
Supplemental Figure 4. Flow chart and robustness analysis of dispersion in gene expression with increasing disease inflammation. (A) Steps taken to assess differences in variance of gene expression and robustness analysis by random subsampling (1000 times). (B) Results of robustness analysis.



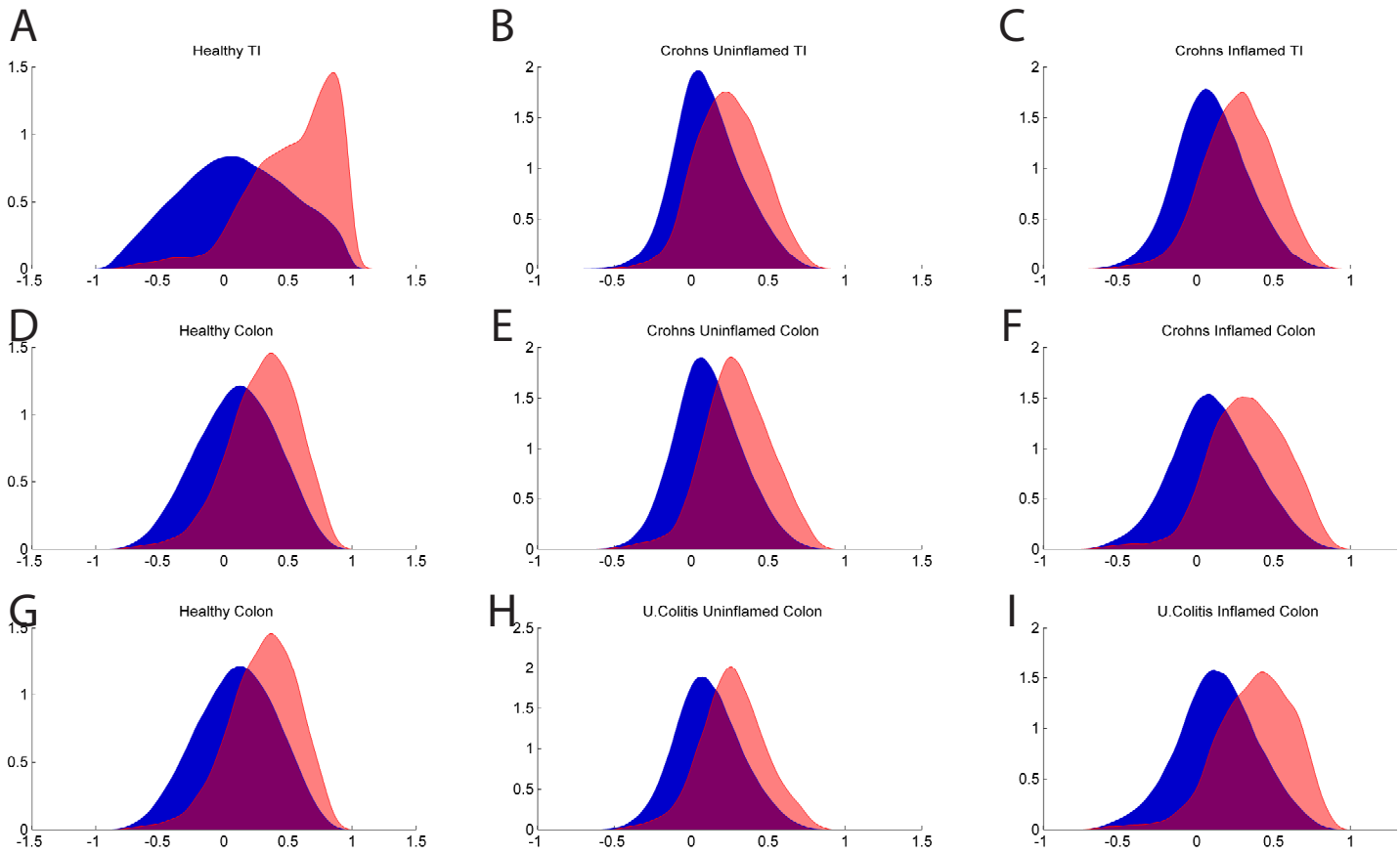
Supplemental Figure 5. Flow chart for identification of consensus eigengene network and hub gene pairs with conserved correlation independent of disease, tissue type, and inflammation status. Analysis subgroups refer to the data stratification shown in Supplemental Table 1.



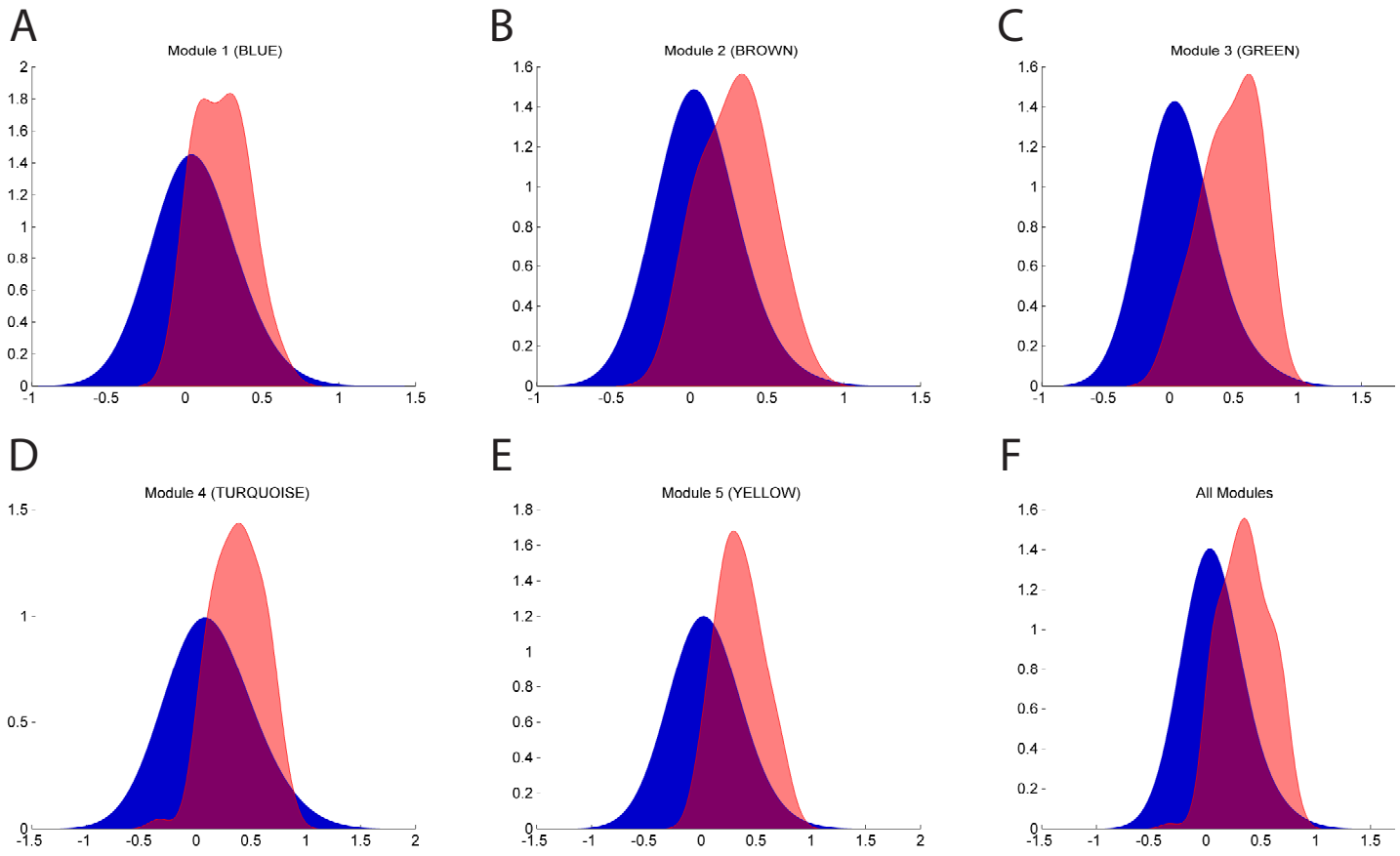
Supplemental Figure 6. WGCNA consensus eigengene network analysis across eight sample groups: Uninflamed.Healthy.TI, Uninflamed.CD.TI, Inflamed.CD.TI, Uninflamed.Healthy.Colon, Uninflamed.CD.Colon, Inflamed.CD.Colon, Uninflamed.Disease.UC.Colon, and Inflamed.Disease.UC.Colon. Hierarchical clustering dendrogram of genes for identifying five consensus modules. Genes in each module are assigned the same color, shown in the color band below the dendrogram.



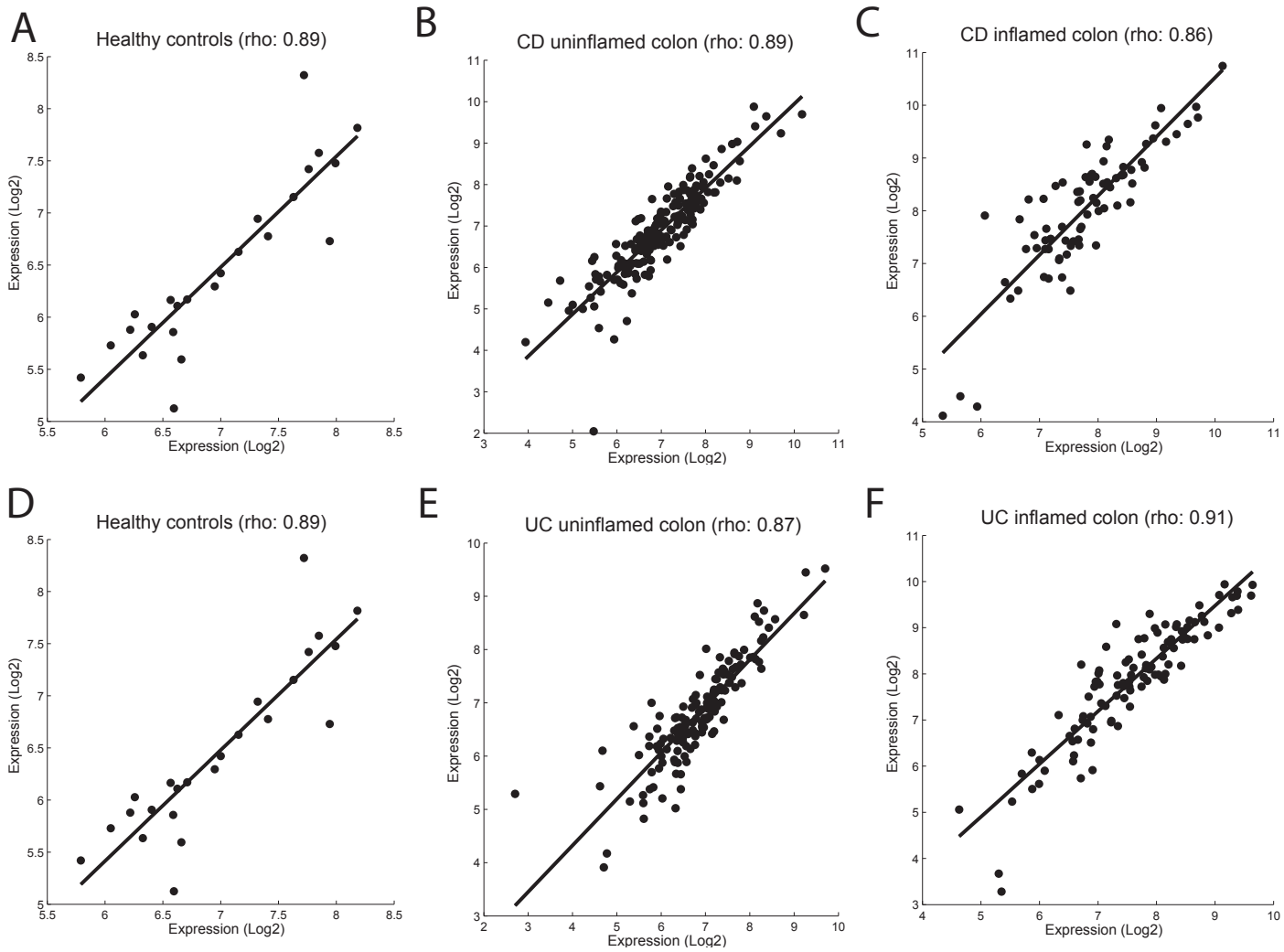
Supplemental Figure 7. WGCNA consensus eigengene network analysis across eight sample groups: Uninflamed.Healthy.TI, Uninflamed.CD.TI, Inflamed.CD.TI, Uninflamed.Healthy.Colon, Uninflamed. CD.Colon, Inflamed.CD.Colon, Uninflamed.Disease.UC.Colon, and Inflamed.Disease.UC.Colon. Matrix of plots showing the consensus eigengene networks in the eight groups. The diagonal plots of the matrix show the heatmaps of eigengene adjacencies in each eigengene network. Colors are from low adjacency (blue) to high adjacency (red). Each bar plot shows the preservation of consensus module eigengenes between any two groups. The preservation of the overall network measure D for these two groups is shown as well. Each heatmap shows the adjacency for the pairwise preservation networks ranging from low adjacency (white) to high adjacency (red).



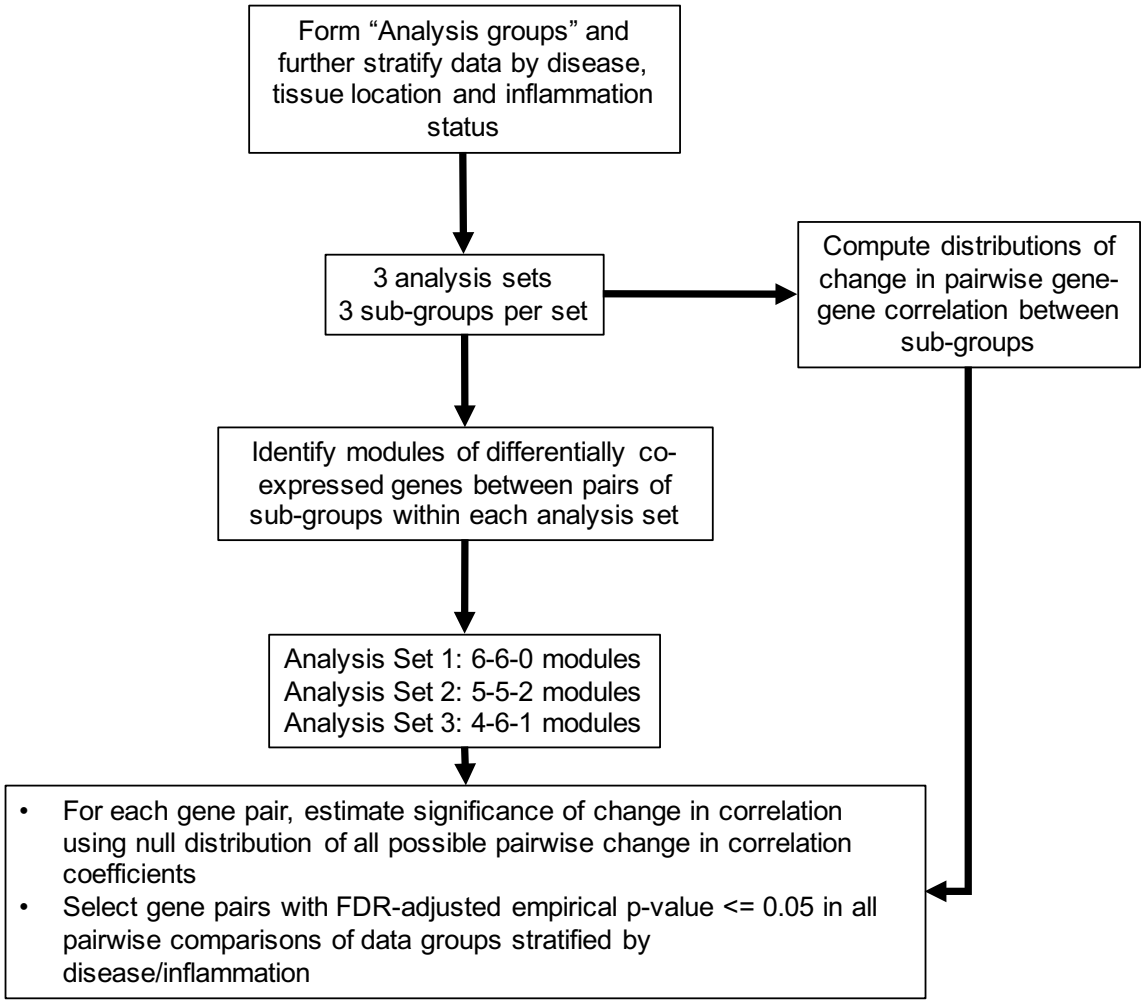
Supplemental Figure 8. Distributions of pairwise Pearson correlation coefficients. Blue density plots show distribution of all pairwise Pearson correlation coefficients observed with the full panel of 678 genes. Red density plots show the distribution of select pairs of genes identified by WGCNA as part of the consensus eigengene network that is conserved across all data subgroups. Only pairs of genes identified in the largest module (TURQUOISE) are used here. Panels A-I represent different analysis sets and data subgroups as stratified in Supplemental Table 1.



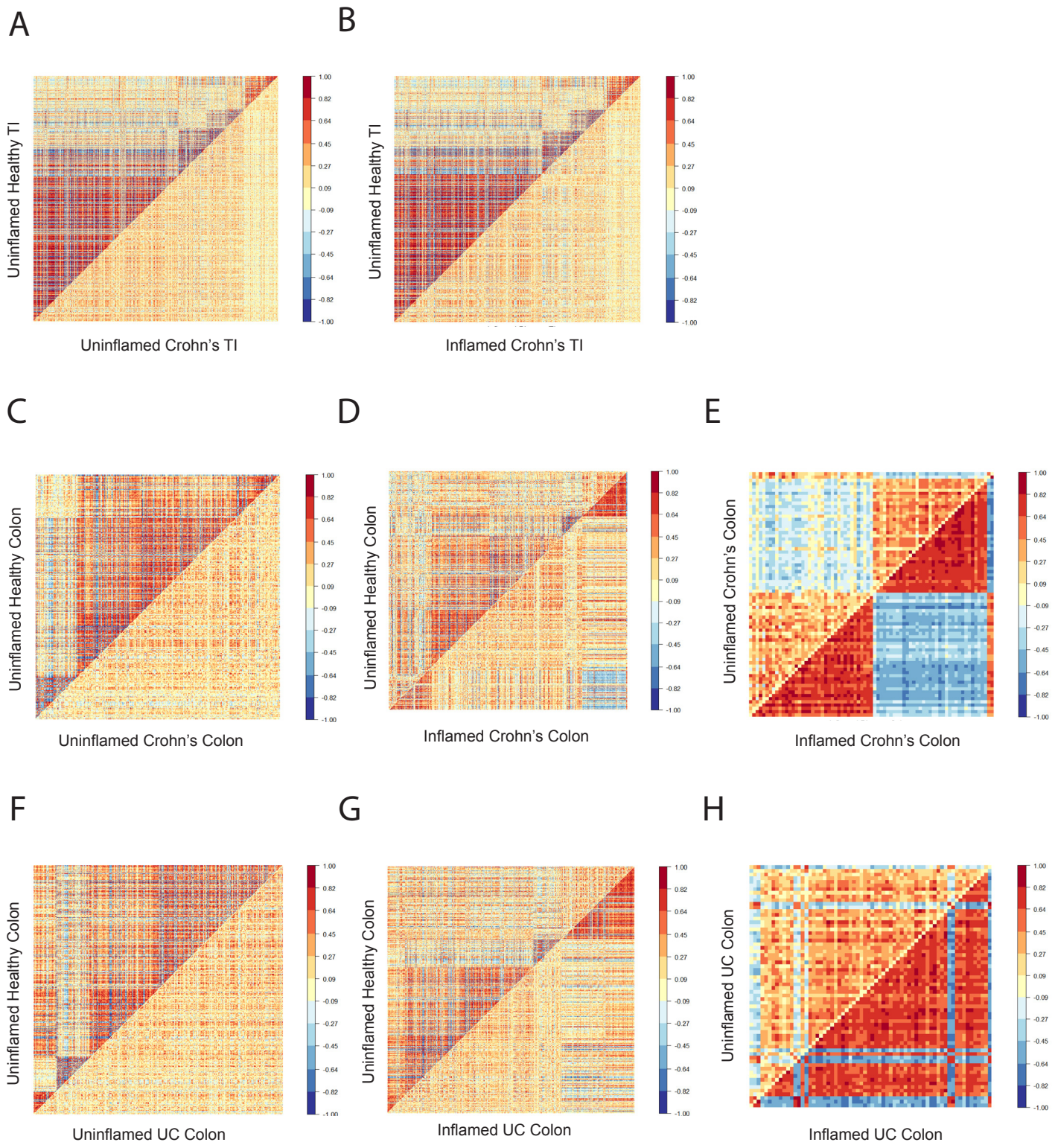
Supplemental Figure 9. Distribution of kME values. WGCNA condenses gene expression per module into an eigengene expression (first principal component) and the correlation of each gene to its eigengene expression is quantified (kME). The closer kME is to 1 or -1, the stronger the evidence that the gene is part of that module. The distribution of these kME values is depicted for all 678 genes in blue and module-specific genes in red. (A-E) Module-specific distribution of kME values of five consensus modules identified by WGNCA. (F) Distribution after pooling data from all modules.



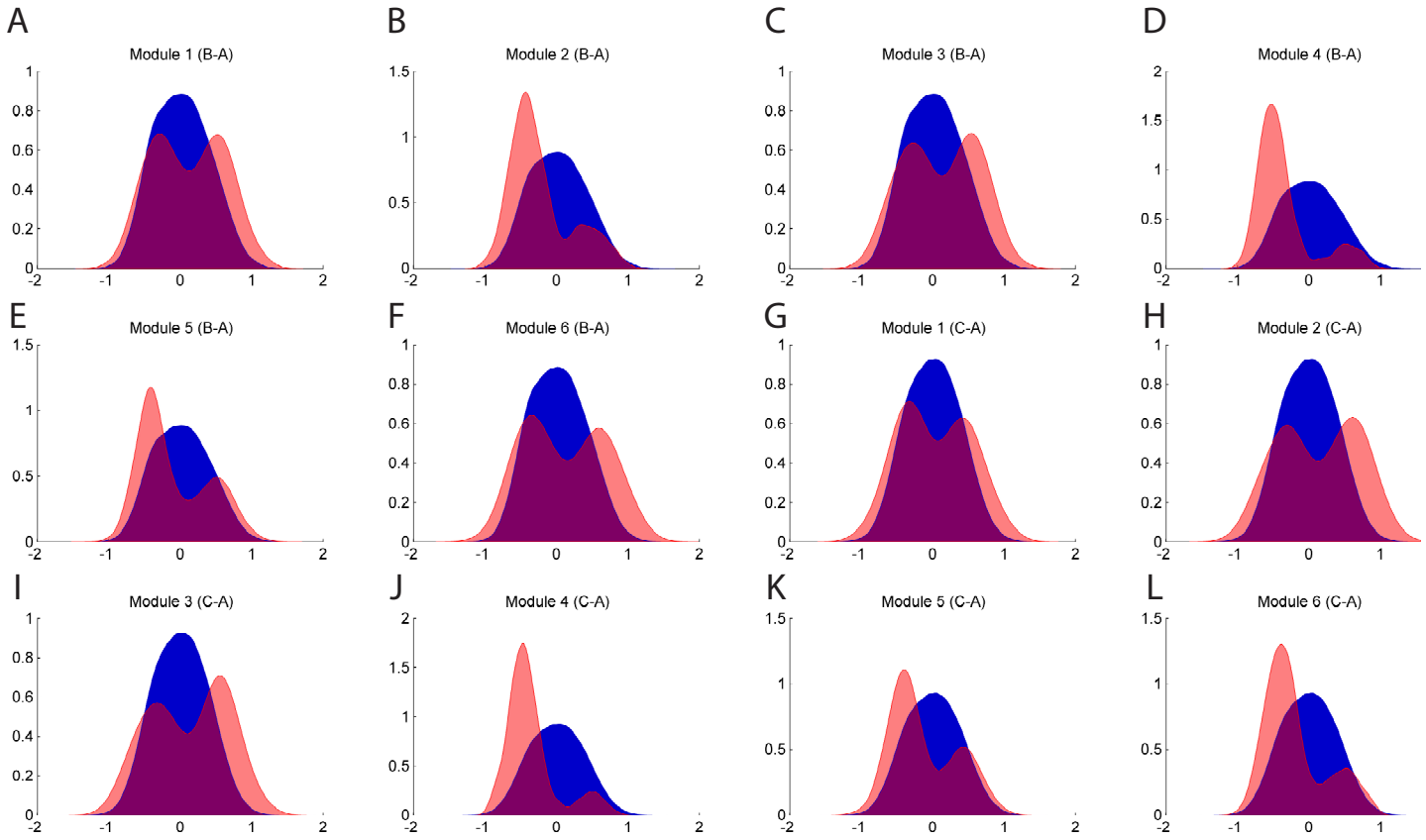
Supplemental Figure 10. Pairwise correlation between *SP140* (x-axis) and *IKZF3* (y-axis) conserved across data groups. Pairwise correlation between *SP140* and *IKZF3* is conserved across disease and tissue groups, independent of inflammation status. (A-C) Expression correlation in colon tissue samples from healthy controls, CD uninflamed, and CD inflamed groups respectively. (D-F) Expression correlation in colon tissue samples from healthy controls, UC uninflamed, and UC inflamed groups respectively.



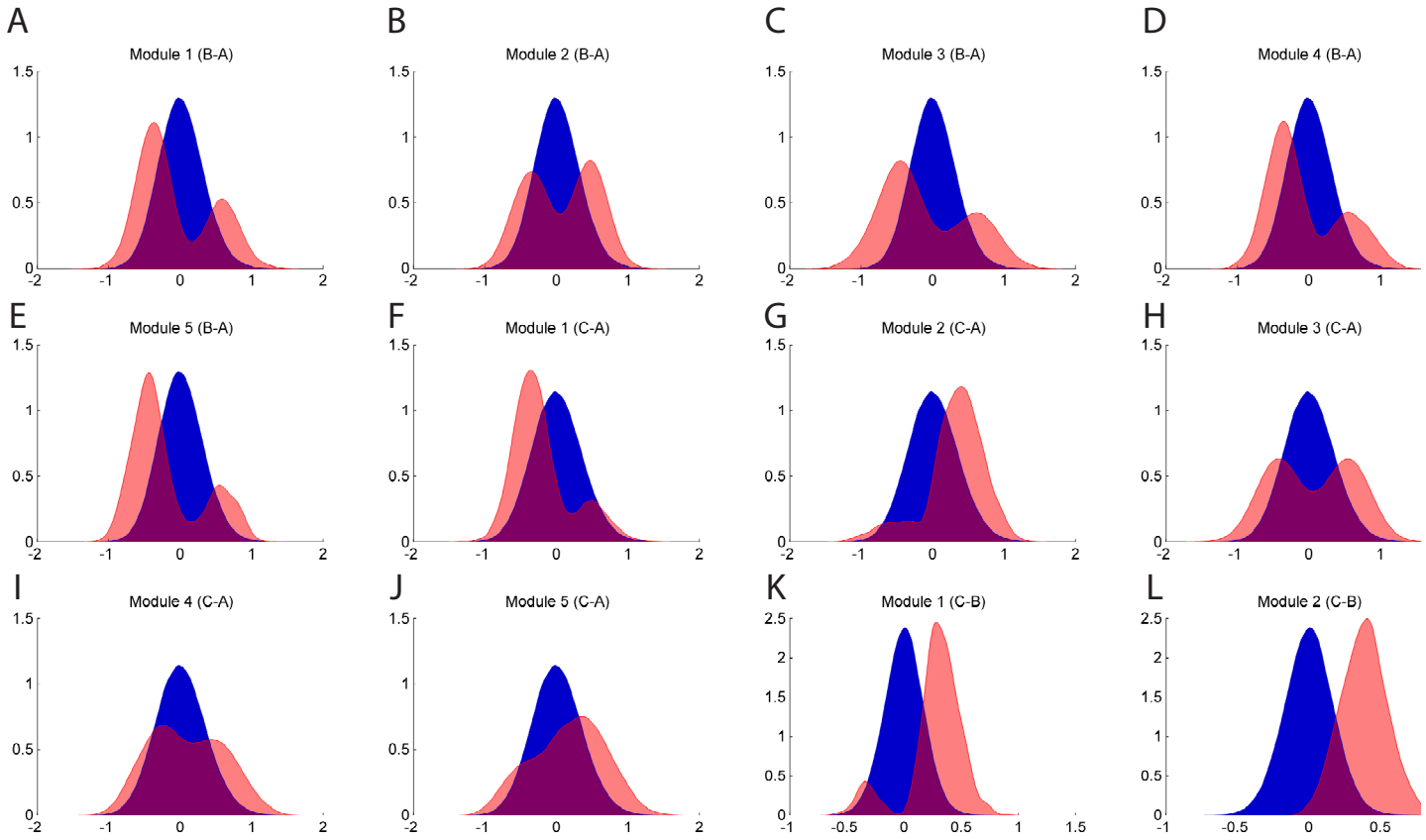
Supplemental Figure 11. Flow chart for identification of differentially co-expressed modules using Diff-CoEx and gene pairs for which the correlation structure is significantly altered from healthy to disease inflamed state. Analysis sets 1, 2, and 3 refer to data stratification as shown in Supplemental Table 1.



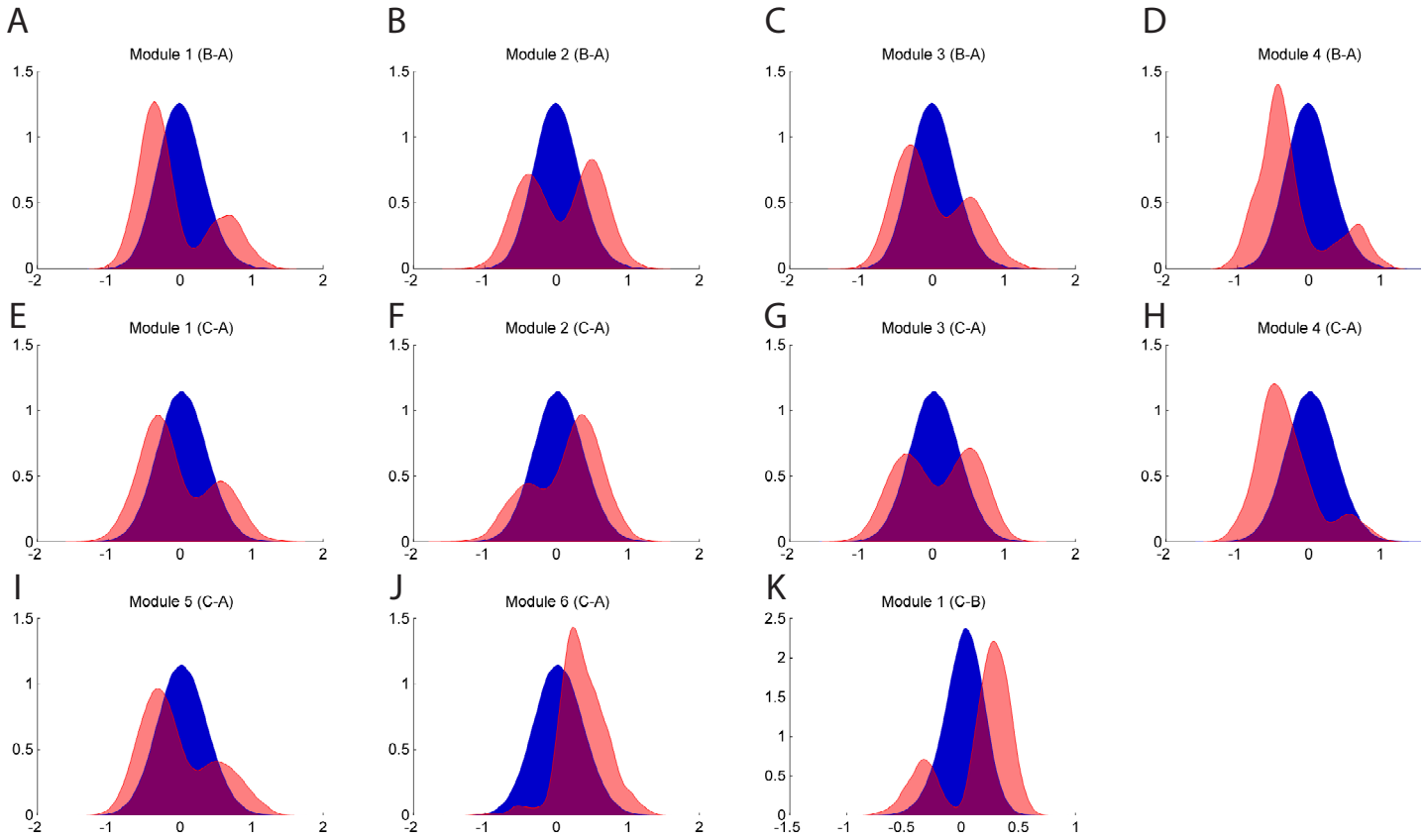
Supplementary Figure 12. Differentially coexpressed gene module analysis. Comparative correlation heatmap shows correlation between pairs of genes ranging from negative correlation (blue) to positive correlation (red) from each comparison. Plots A, C, and F show the comparative correlation heatmaps from uninflamed healthy state to uninflamed disease state; plots B, D, and G show the comparative correlation heatmaps from uninflamed healthy state to inflamed disease state; plots E and H show the comparative correlation heatmaps from uninflamed disease state to inflamed disease state.



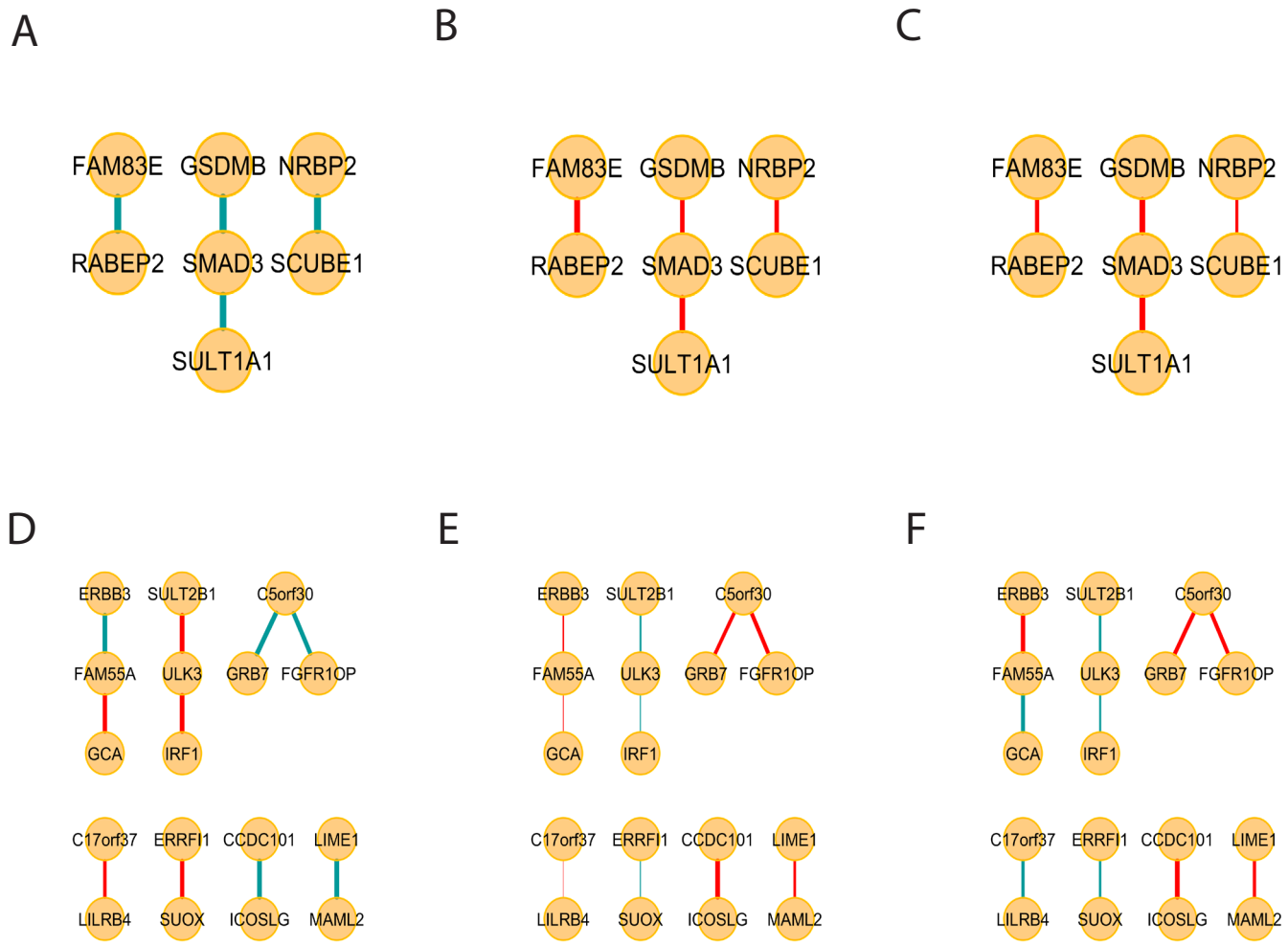
Supplementary Figure 13. Distributions of change in pairwise Pearson correlation coefficients. DiffCoEx identified 12 modules that were differentially co-expressed between healthy TI and CD uninflamed and inflamed TI samples (Supplemental Table 5). These plots show the distribution of change in pairwise Pearson correlation coefficient between two data subgroups. Blue density plots refer to null distribution of change in Pearson correlation coefficients observed with all pairs of 678 genes. Red density plots show distribution of select pairs of genes identified by DiffCoEx as part of the differentially expressed modules. Panels A-L represent different modules.



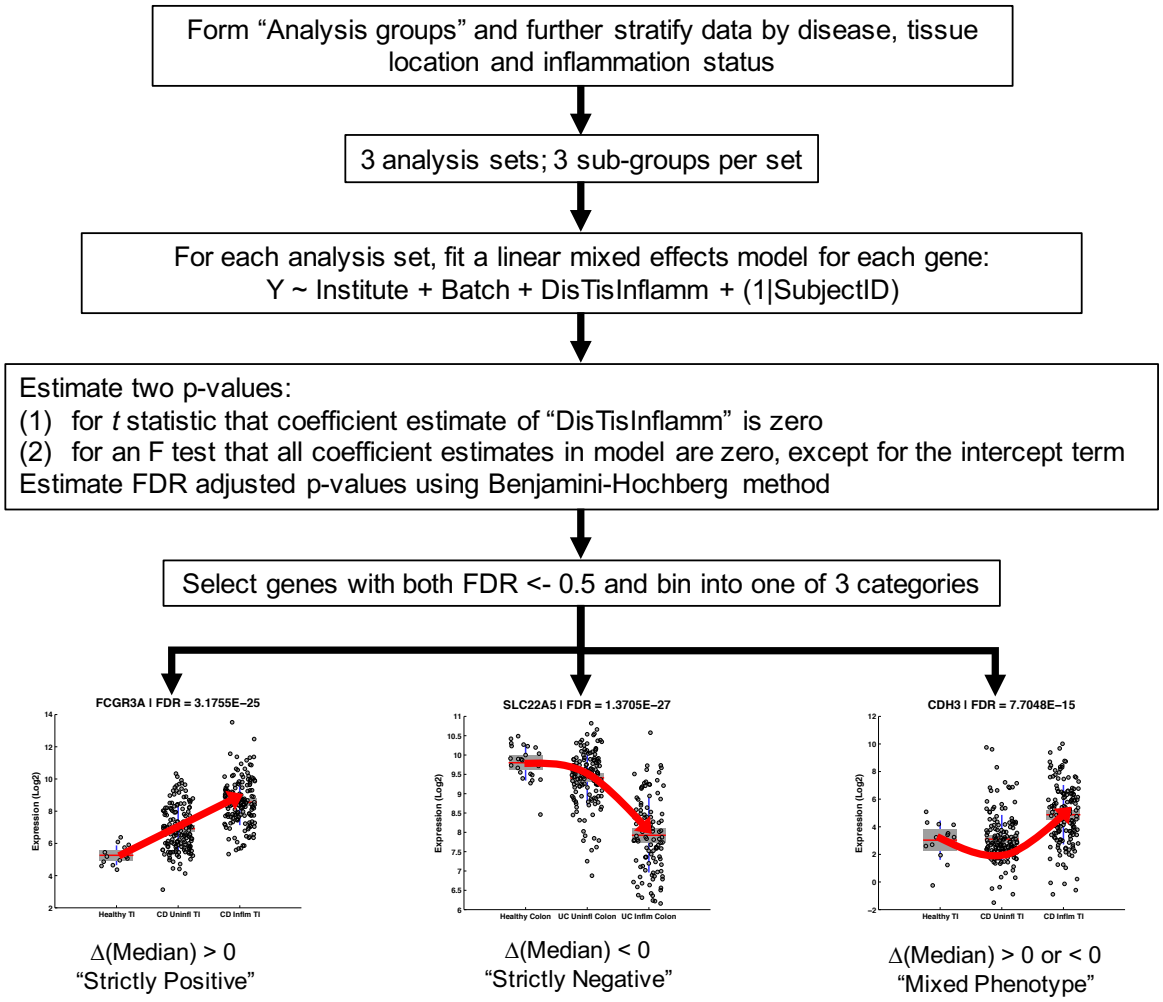
Supplemental Figure 14. Distributions of change in pairwise Pearson correlation coefficients. DiffCoEx identified 12 modules that were differentially co-expressed between healthy colon and CD uninflamed and inflamed colon samples (Supplemental Table 5). These plots show the distribution of change in pairwise Pearson correlation coefficient between two data subgroups. Blue density plots refer to null distribution of change in Pearson correlation coefficients observed with all pairs of 678 genes. Red density plots show distribution of select pairs of genes identified by DiffCoEx as part of the differentially expressed modules. Panels A-L represent different modules.



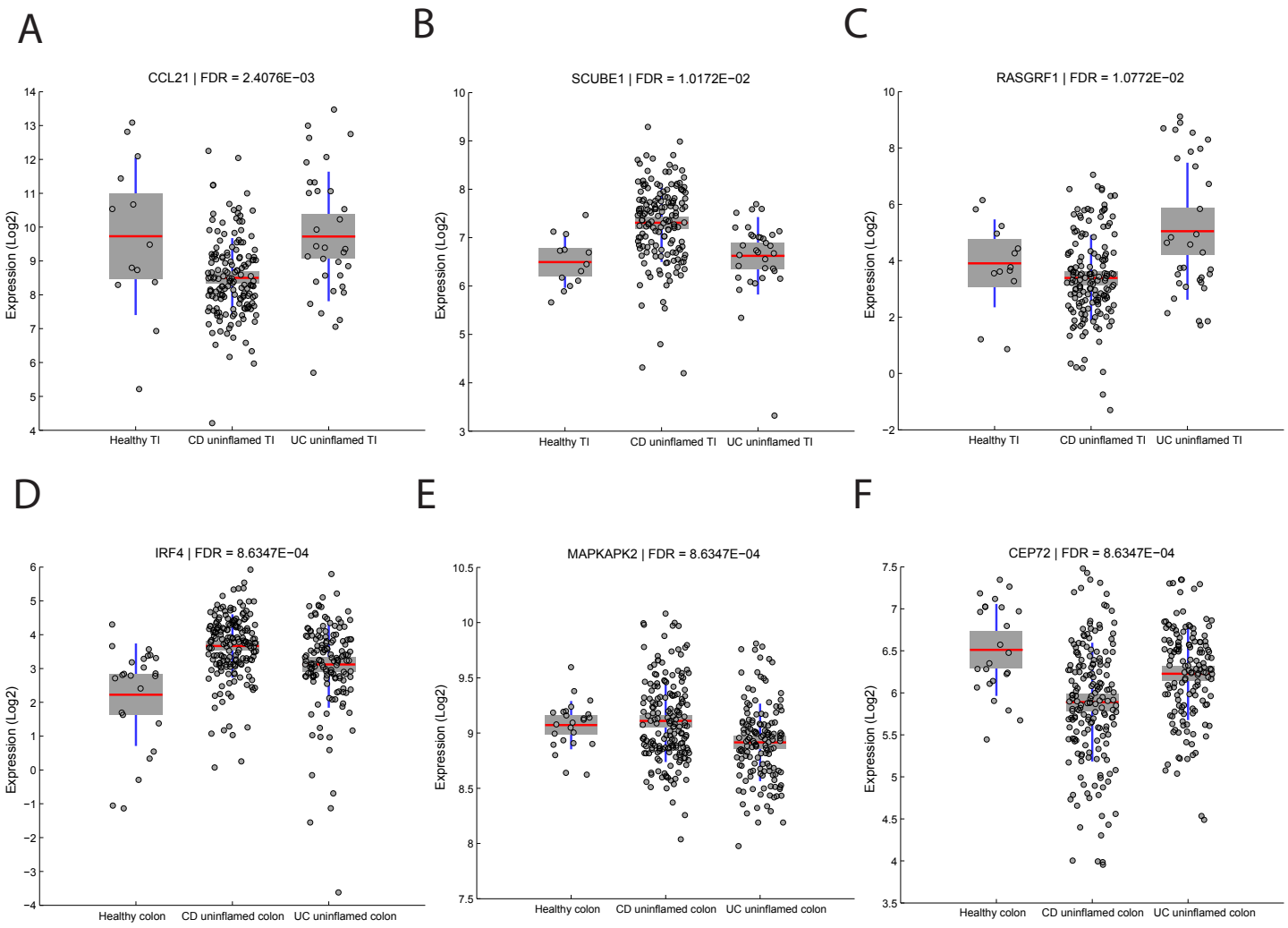
Supplemental Figure 15. Distributions of change in pairwise Pearson correlation coefficients. DiffCoEx identified 11 modules that were differentially co-expressed between healthy colon and UC uninflamed and inflamed colon samples (Supplemental Table 5). These plots show the distribution of change in pairwise Pearson correlation coefficient between two data subgroups. Blue density plots refer to null distribution of change in Pearson correlation coefficients observed with all pairs of 678 genes. Red density plots show distribution of select pairs of genes identified by DiffCoEx as part of the differentially expressed modules. Panels A-K represent different modules.



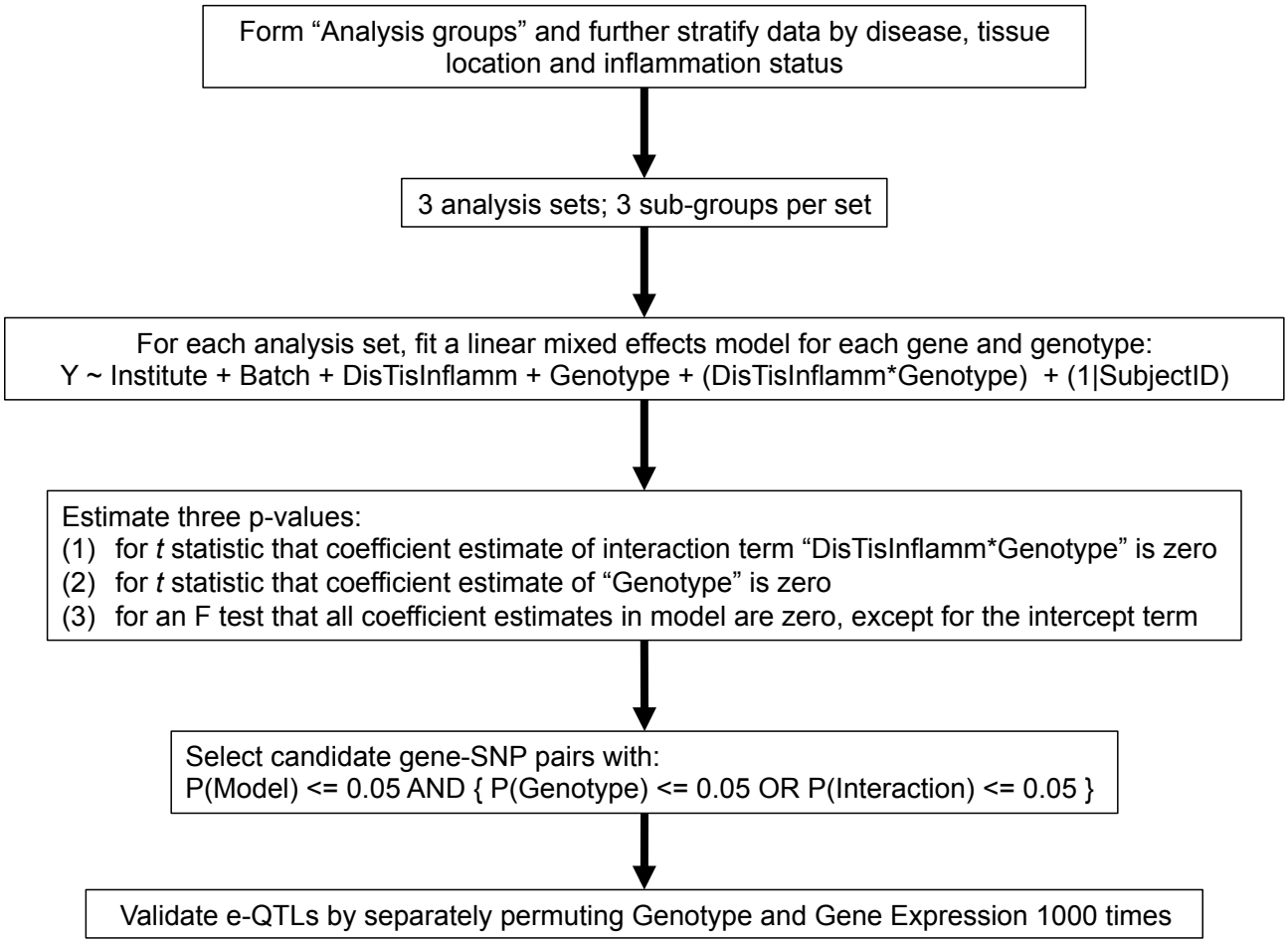
Supplemental Figure 16. Network visualization of differentially coexpressed modules. (A-C) Differentially co-expressed gene pairs in TI samples of healthy (A), uninflamed CD (B), and inflamed CD (C) samples. (D-F) Differentially co-expressed gene pairs in colon samples of healthy (D), uninflamed UC (E), and inflamed UC (F) samples. Gene pairs were selected based on whether the Pearson's correlation coefficient exhibited the most significant change in all three pairwise comparisons between three data groups. For each module, genes are shown as nodes and pairwise correlations are displayed as edges (red, positive correlation; blue, negative correlation). Thicker edges represent stronger absolute correlations.



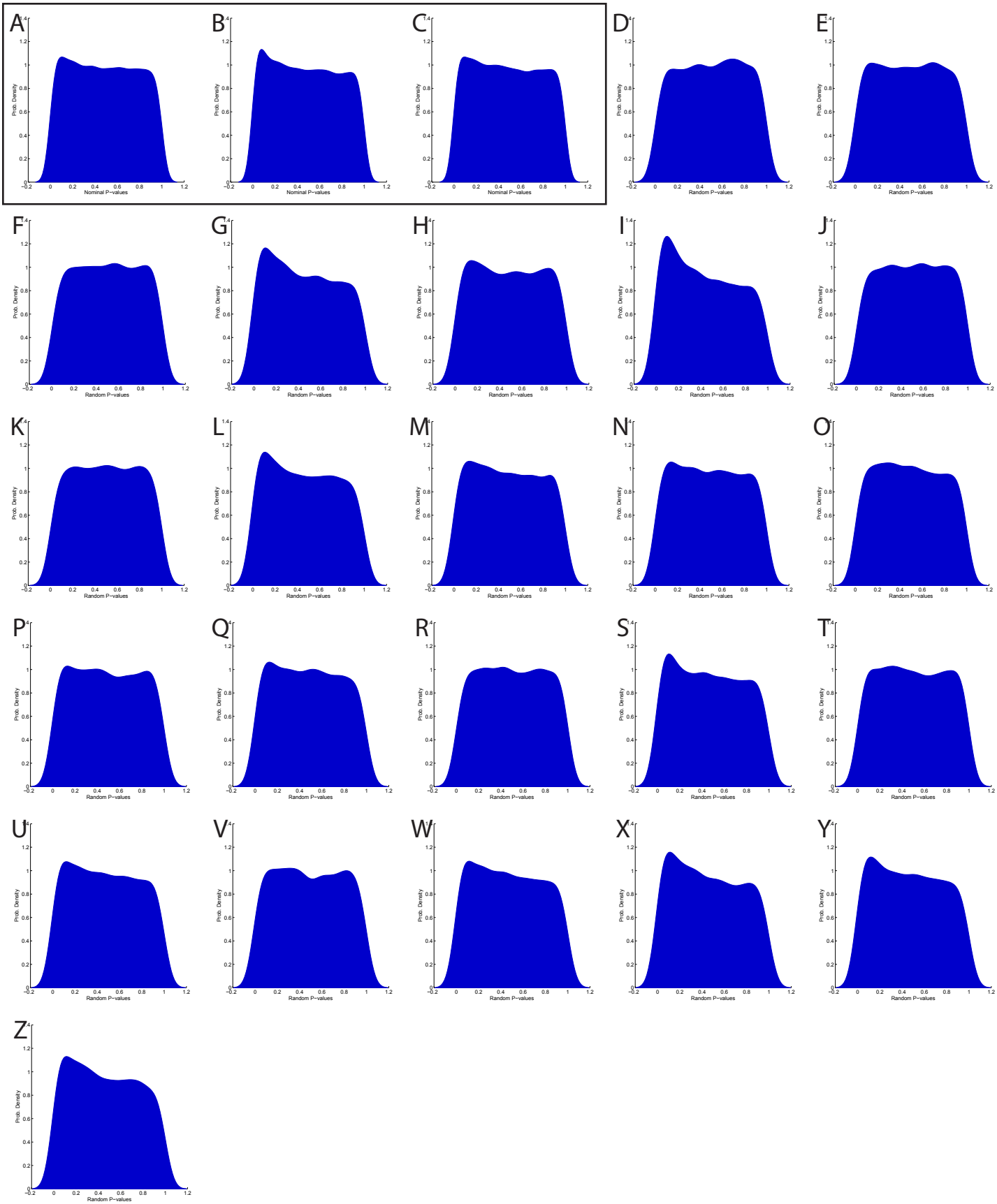
Supplemental Figure 17. Flow chart for identification of differentially expressed genes using linear mixed effects models. Analysis sets refer to data stratification as shown in Supplemental Table 1.



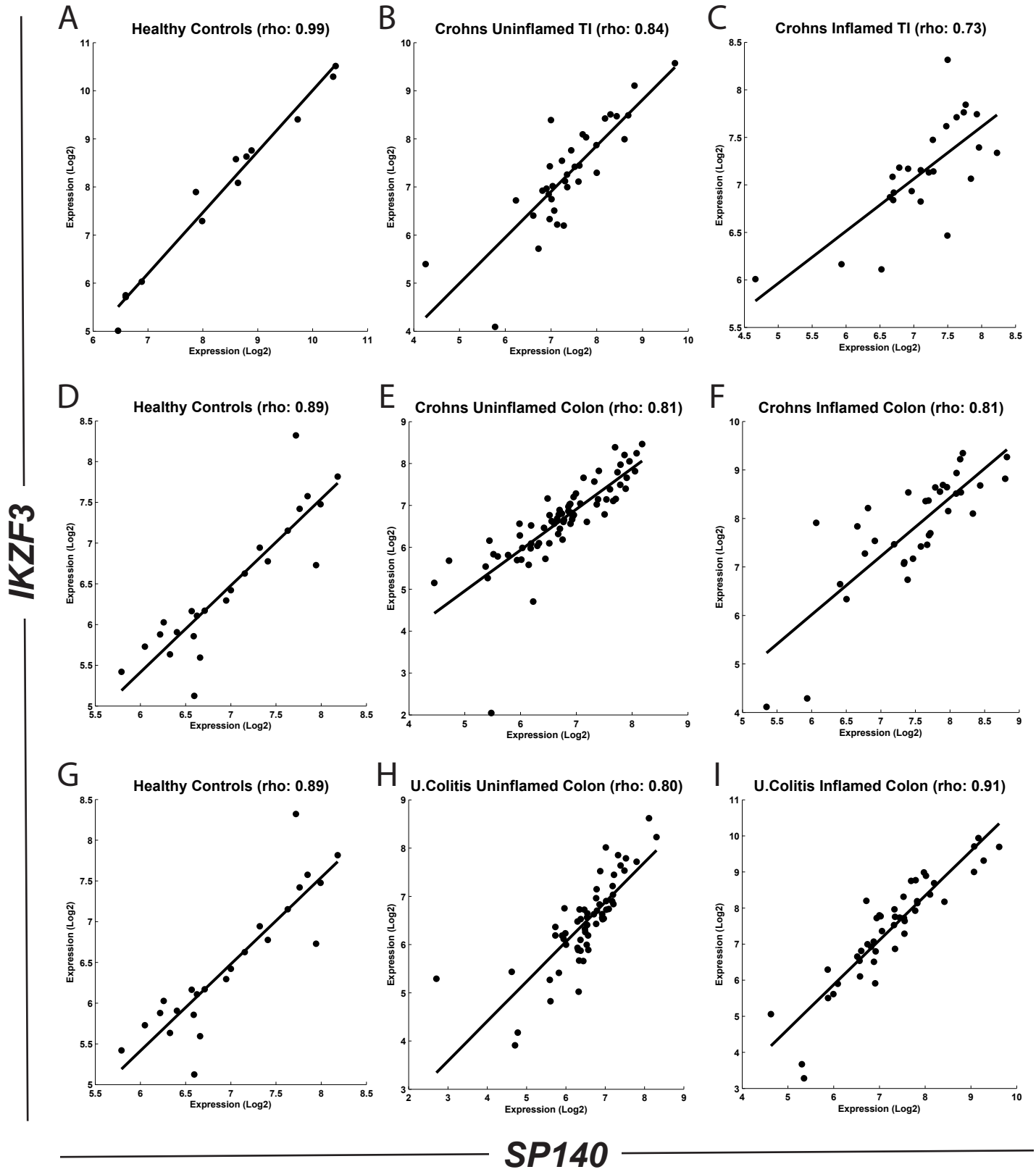
Supplemental Figure 18: Top differentially expressed genes in uninflamed TI and colon samples between CD and UC patients. (A) *CCL21* (FDR = 2.4076E−03), (B) *SCUBE1* (FDR = 1.0172E−02), and (C) *RASGRF1* (FDR = 1.0772E−02) exhibit the most statistically significant differential expression between uninflamed TI samples of CD and UC patients. (D) *IRF4* (FDR = 8.6347E−04), (E) *MAPKAP2* (FDR = 8.6347E−04), and (F) *CEP72* (FDR = 8.6347E−04) exhibit the most statistically significant differential expression between uninflamed colon samples of CD and UC patients.



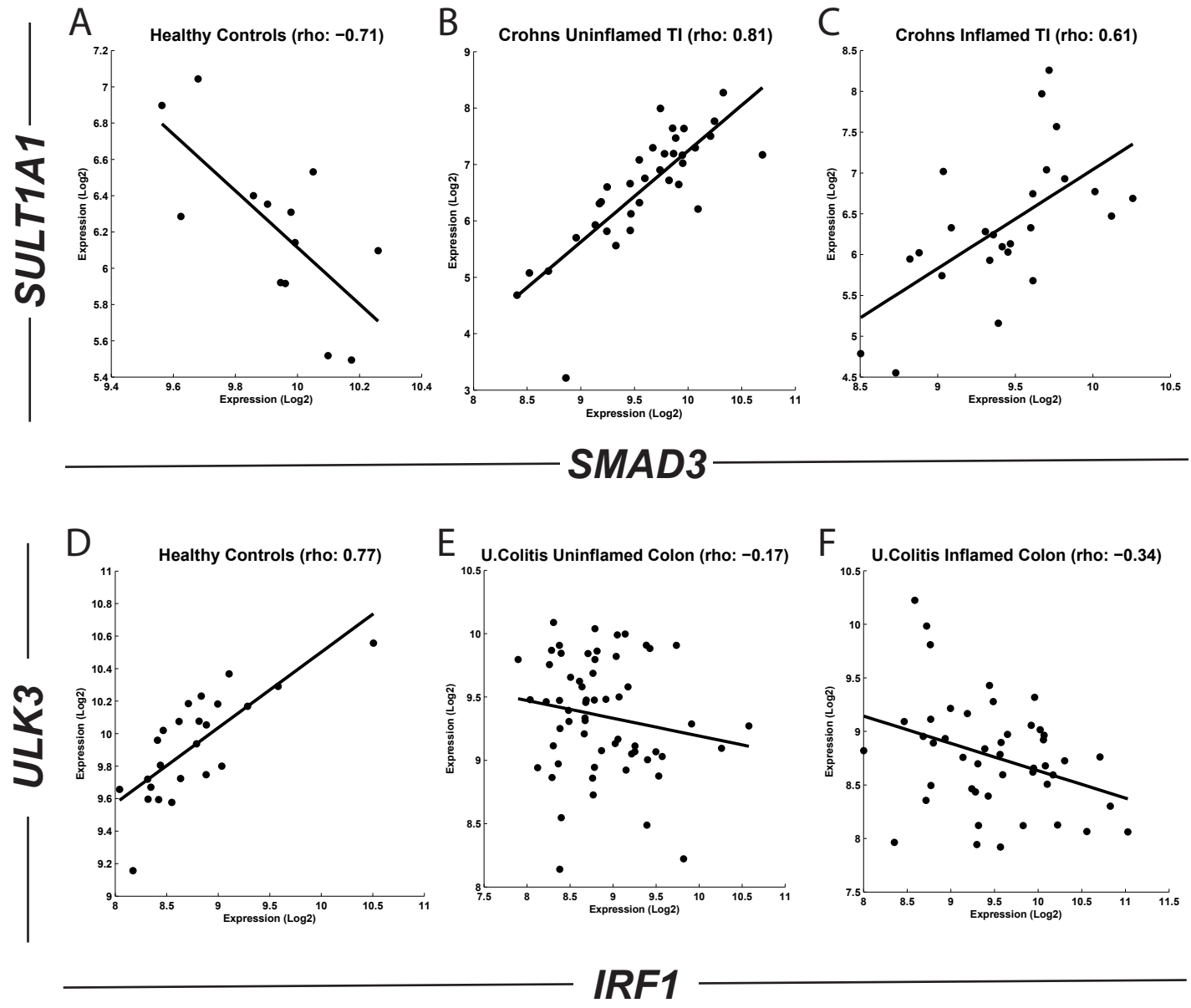
Supplemental Figure 19. Flow chart for identification of eQTLs using linear mixed effects models. Candidate eQTLs identified by p-value assessment of coefficients of the linear model were further validated by independently permuting the genotype and gene expression 1000 times. Distribution of nominal and permuted p-values are shown in Supplemental Figure 20. All p-values of significant eQTLs are listed in Supplemental Table 11.



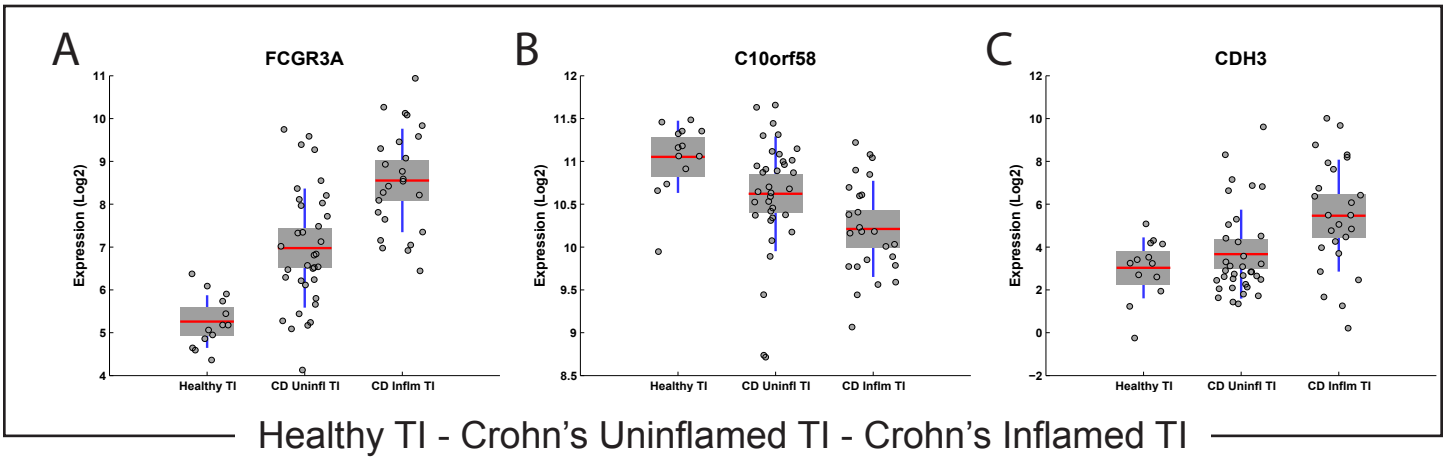
Supplemental Figure 20. Distribution of p-values. (A-C) Distribution of nominal p-values computed for the t statistic that coefficient estimate of genotype effect in linear model is zero. 34 candidate eQTLs were identified and further tested by permuting genotype and gene expression independently. (D-Z) Distribution of p-values obtained by permuting genotype and re-fitting the linear model for each of the candidate eQTLs that were tested across three groups. Each distribution is derived from 1,000 random permutations.

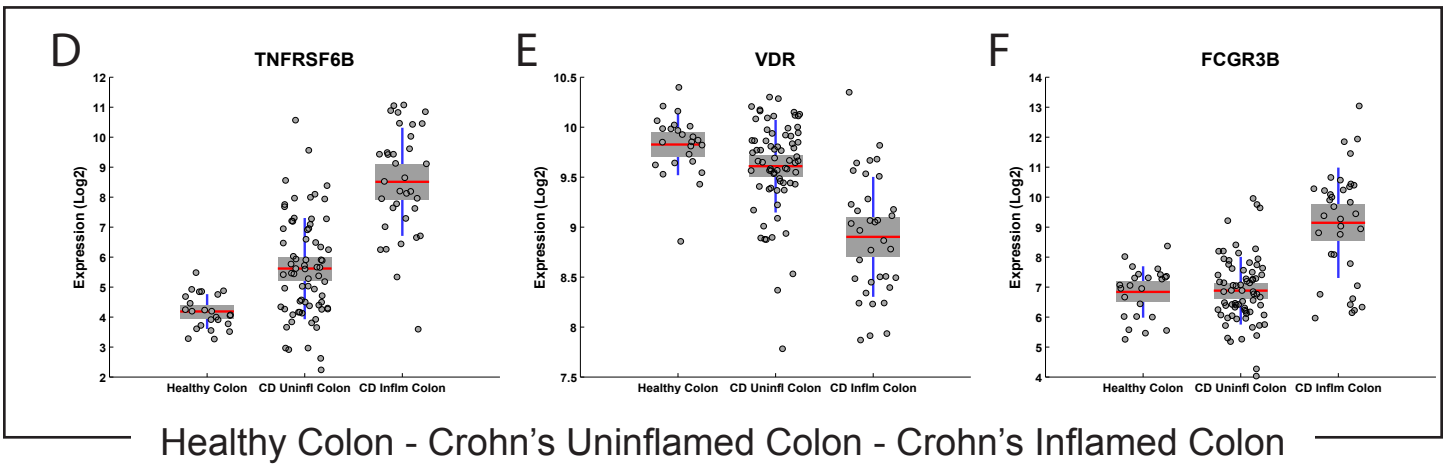


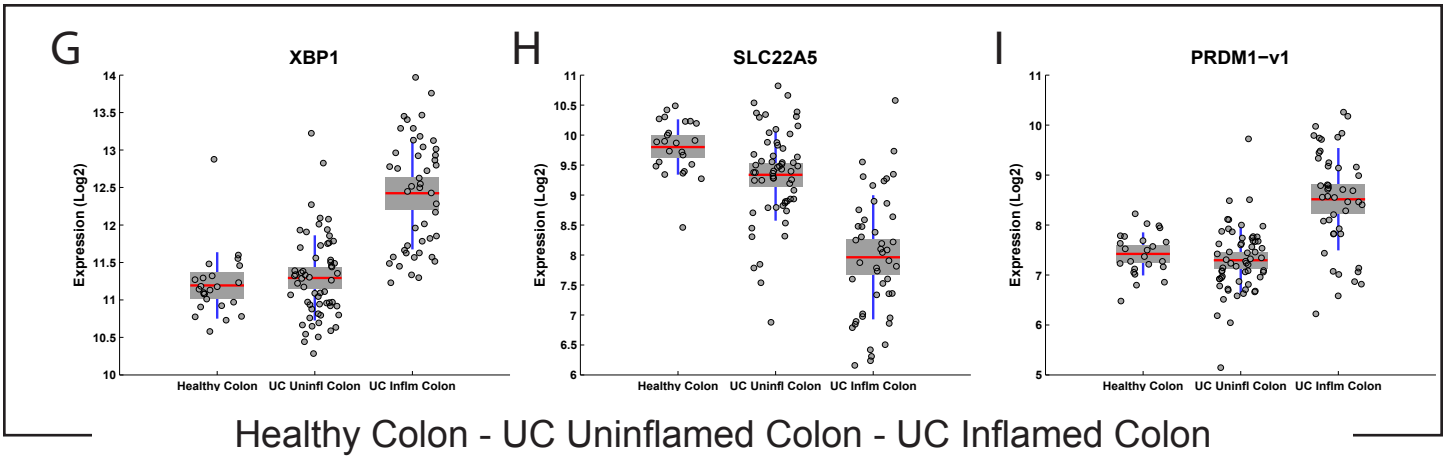
Supplemental Figure 21. Pairwise correlation between *SP140* and *IKZF3* conserved across groups with data from age-matched patients.



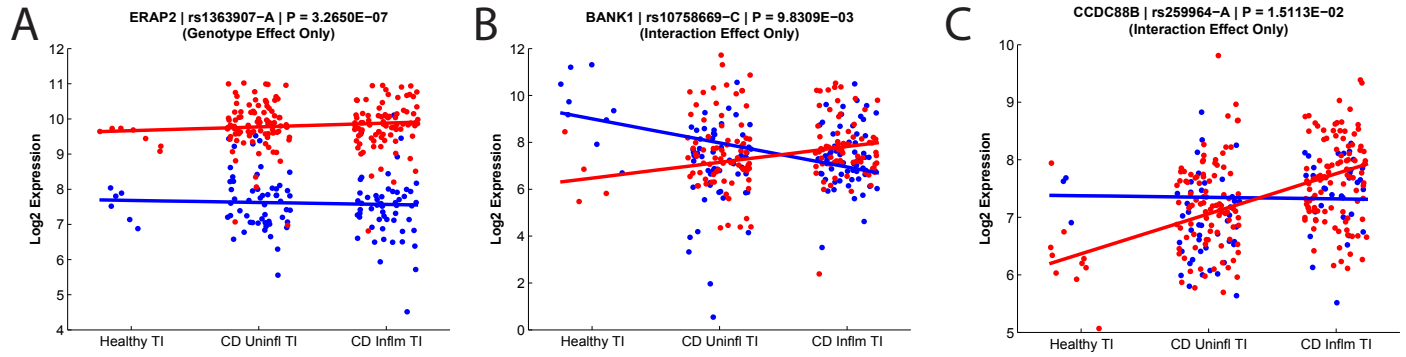
Supplemental Figure 22. Pairwise correlation between *SULT1A1* and *SMAD3* (A-C) and *IRF1* and *ULK3* (D-F) show differential directionality of correlation between controls and disease groups with data from age-matched patients.



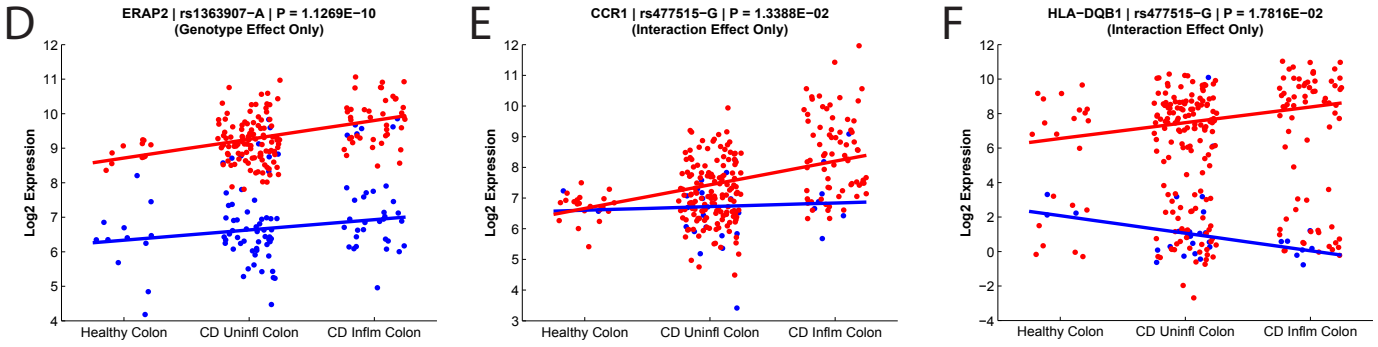




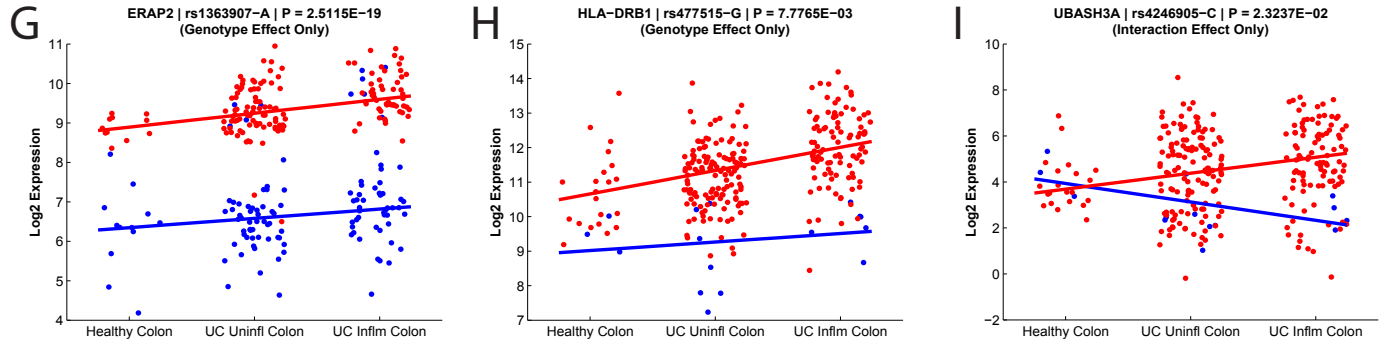
Supplemental Figure 23. Top differentially expressed genes across healthy controls, uninflamed, and inflamed disease states with data from age-matched patients.



Healthy TI - Crohn's Uninflamed TI - Crohn's Inflamed TI



Healthy Colon - Crohn's Uninflamed Colon - Crohn's Inflamed Colon



Healthy Colon - UC Uninflamed Colon - UC Inflamed Colon

Supplemental Figure 24. Top eQTLs identified with genes demonstrating differential expression across healthy controls and disease groups with data from age-matched patients.

Supplementary Table 1 (XLS): Analysis groups formed by stratifying data based on tissue type, disease type, and inflammation status.

Supplementary Table 2 (XLS): Breakdown of NanoString codeset design. For a few select genes, probes for multiple isoforms of the same genes were added and these are named as *_v1 or *_v2.

Supplementary Table 3 (XLS): Gene pairs with conserved correlation structure independent of disease, tissue type, and inflammation. Gene names with asterisks indicate genes that were identified as significantly differentially expressed, as per the linear mixed effects model analysis, in all three analysis sets identified in Supplemental Table 1. Candidate gene pairs were first identified by WGCNA consensus eigengene network analysis and then the most robust gene pairs with conserved correlation were identified as outlined in Figure S3.

Supplementary Table 4 (PDF): Table summarizes comparisons that were used in differential co-expression module analysis and total number of differentially co-expressed modules that were detected (red).

Supplementary Table 5 (XLS): Gene set overlap analysis of differentially co-expressed modules to identify enrichment for transcription factor binding sites (TFBS). Transcription factors were identified on the basis whether there was a statistically significant enrichment of genes with promoter regions [-2kb,2kb] around transcription start site containing a binding motif specific to that transcription factor. TFBS annotations were used from mSigDB database hosted at The Broad Institute.

Supplementary Table 6 (XLS): Gene pairs with significant change in correlation structure from healthy inflamed to disease uninflamed to disease inflamed state. Candidate gene pairs were first identified as members of differentially co-expressed modules by DiffCoEx and then the most robust gene pairs with altered correlation patterns were identified as outlined in Supplemental Figure 11. Analysis was carried out separately for each of the analysis sets identified in

Supplemental Table 1. Gene names with asterisks indicate genes that were identified as significantly differentially expressed in that respective analysis set.

Supplementary Table 7 (XLS): Differentially expressed genes in TI of healthy and CD uninflamed and inflamed samples. Differential analysis scheme is outlined in Supplemental Figure 17.

Supplementary Table 8 (XLS): Differentially expressed genes in colon of healthy and CD uninflamed and inflamed samples. Differential analysis scheme is outlined in Supplemental Figure 17.

Supplementary Table 9 (XLS): Differentially expressed genes in colon of healthy and UC uninflamed and inflamed samples. Differential analysis scheme is outlined in Supplemental Figure 17.

Supplementary Table 10 (XLS): Differentially expressed genes in uninflamed TI/colon samples of CD and UC patients. Differential analysis scheme is outlined in Supplemental Figure 17.

Supplementary Table 11 (XLS): Significant eQTLs detected using linear mixed effects models, as outlined in Supplemental Figure 19.

Supplementary Table 12 (XLS): Summary table of results for IBD-associated risk loci with candidate genes prioritized by scoring significantly in at least one of the five analyses: TI v. colon expression, conserved co-expression pairs, differential co-expression pairs, differential expression, and eQTL. Additional annotations included are the IBD-associated SNP, SNP location, dbSNP functional annotation, gene of interest encoded within the loci, gene position, patterns of differential expression, co-expression pairs, and eQTL effect.