JCI insight

Characterization of candidate genes in inflammatory bowel disease–associated risk loci

Joanna M. Peloquin,[1,2] Gautam Goel,[1,2] Lingjia Kong,[1,2] Hailiang Huang,[3,4] Talin Haritunians,[5] R. Balfour Sartor,[6] Mark J. Daly,[3,4] Rodney D. Newberry,[7] Dermot P. McGovern,[5] Vijay Yajnik,[1] Sergio A. Lira,[8] and Ramnik J. Xavier[1,2,4]

[1]Gastrointestinal Unit and Center for the Study of Inflammatory Bowel Disease, [2]Center for Computational and Integrative Biology, [3]Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. [4]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [5]F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. [6]Department of Medicine, Center for Gastrointestinal Biology and Disease, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [7]Department of Internal Medicine, Washington University School of Medicine, St. Louis, Missouri, USA. [8]Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

GWAS have linked SNPs to risk of inflammatory bowel disease (IBD), but a systematic characterization of disease-associated genes has been lacking. Prior studies utilized microarrays that did not capture many genes encoded within risk loci or defined expression quantitative trait loci (eQTLs) using peripheral blood, which is not the target tissue in IBD. To address these gaps, we sought to characterize the expression of IBD-associated risk genes in disease-relevant tissues and in the setting of active IBD. Terminal ileal (TI) and colonic mucosal tissues were obtained from patients with Crohn's disease or ulcerative colitis and from healthy controls. We developed a NanoString code set to profile 678 genes within IBD risk loci. A subset of patients and controls were genotyped for IBD-associated risk SNPs. Analyses included differential expression and variance analysis, weighted gene coexpression network analysis, and eQTL analysis. We identified 116 genes that discriminate between healthy TI and colon samples and uncovered patterns in variance of gene expression that highlight heterogeneity of disease. We identified 107 coexpressed gene pairs for which transcriptional regulation is either conserved or reversed in an inflammation-independent or -dependent manner. We demonstrate that on average approximately 60% of disease-associated genes are differentially expressed in inflamed tissue. Last, we identified eQTLs with either genotype-only effects on expression or an interaction effect between genotype and inflammation. Our data reinforce tissue specificity of expression in disease-associated candidate genes, highlight genes and gene pairs that are regulated in disease-relevant tissue and inflammation, and provide a foundation to advance the understanding of IBD pathogenesis.

Introduction

Inflammatory bowel diseases (IBD), encompassing Crohn's disease (CD) and ulcerative colitis (UC), are chronic, inflammatory conditions of the bowel with pathogenesis driven by both genetic and environmental risk factors. Physical manifestations of disease range from discontinuous patterns of inflammation along the entire gastrointestinal tract in CD to localized but continuous inflammation of colon in UC. GWAS have uncovered more than 200 SNP-tagged regions as associated with IBD, implicating several hundred genes as being possibly associated with disease. (1). However, there has been a critical lack of characterization of expression of these disease-associated genes in the tissue that is directly targeted in IBD.

Two approaches to determine candidate genes within risk loci include the study of tissue- and disease-specific differential gene expression and the characterization of expression quantitative trait loci (eQTLs) effects. To date, the largest gene expression studies in IBD-relevant tissues were analyzed on microarray platforms, but approximately 23% of the genes now of interest in IBD were not captured by these methods (2, 3). eQTL analyses associate SNPs with variable RNA expression of nearby genes (cis-
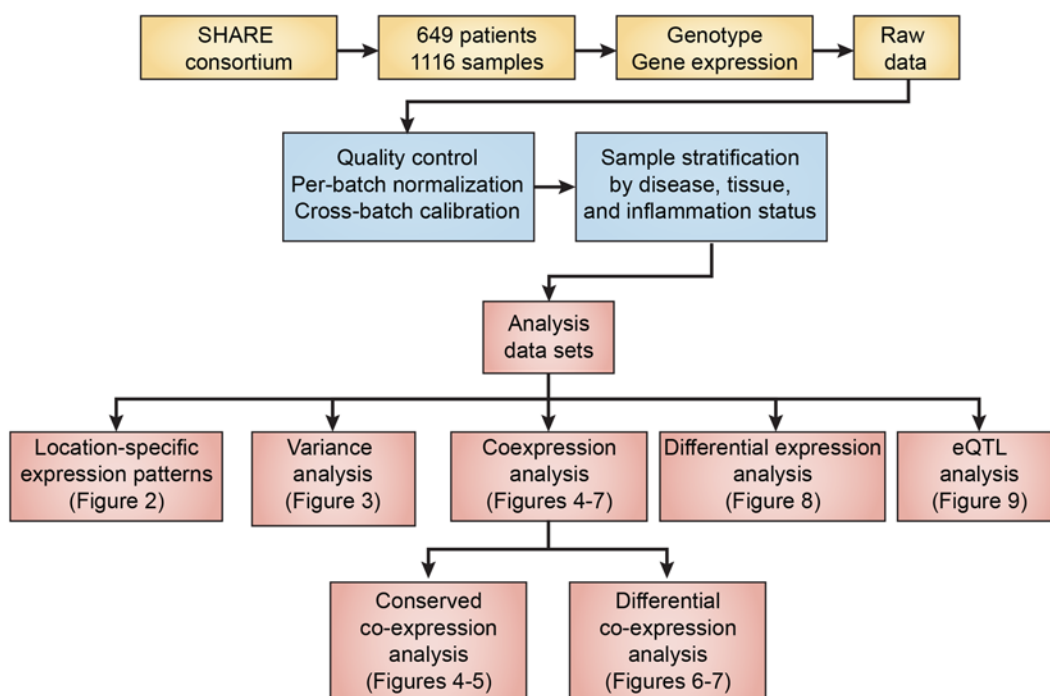
Figure 1. Study design and analysis. The Sinai-Helmsley Alliance for Research Excellence (SHARE) consortium facilitated a multi-institutional collaboration for patient recruitment and sample collection, with centralized RNA expression profiling using NanoString technology, genotyping on the Immunochip array, and bioinformatics analysis. Samples were divided into 3 analysis groups stratified by tissue type, disease type, and inflammation status, as shown in Supplemental Table 1. The analysis pipeline for RNA expression data included quality control, per-batch normalization, and cross-batch calibration. One hundred twenty-seven samples (11%) failed quality control. Eight hundred eighty-six samples, from 590 subjects who were genotyped, were selected for analysis. The final analysis cohort included 37 samples from healthy controls, 564 samples from patients with Crohn's disease, and 285 samples from patients with ulcerative colitis. Five distinct analyses were carried out with genotype and gene expression data. Location-specific expression patterns identified a gene signature that can discriminate between healthy colon and terminal ileum. Variance analysis examined the change in coefficient of variation as inflammation increases from healthy to disease state. Coexpression analysis identified both conserved and differentially coexpressed modules of genes using weighted gene coexpression network analysis. Differential expression and expression quantitative trait loci (eQTL) analysis used linear mixed-effects models to identify genes with significant regulatory patterns in inflammation of disease-specific tissue samples.

eQTL) or distant genes (trans-eQTL) (4), deriving power from sample size and allele frequency of the risk SNP (5). The largest eQTL analyses in IBD have focused on profiling peripheral blood (6). Studies of cell type– and tissue-specific eQTL effects are just beginning to be reported for other complex immune-mediated diseases (6–10). The first studies characterizing eQTLs in intestinal tissues from patients with IBD support disease and tissue specificity (11–13) but have been limited by sample size and exclusion of patients with active IBD.

The goal of this project was to characterize the differences in expression of IBD-associated disease genes in uninflamed and inflamed regions of the terminal ileum (TI) and colon from CD and UC patients (Figure 1). To permit study of a large number of mucosal tissues on a single transcriptomic platform, especially for the genes not covered by microarray platforms, we designed a NanoString code set with 678 disease-associated genes (14) and generated expression profiles for more than 1,100 mucosal tissue samples from approximately 600 patients (Supplemental Tables 1 and 2; supplemental material available online with this article; doi:10.1172/jci.insight.87899DS1). A subset of patients and controls were genotyped on the Illumina Immunochip array (15) with focused analysis of IBD-specific risk loci.

Herein we present a quantitative analysis of mRNA expression of IBD-associated genes in disease-relevant tissue. We identified a signature set of 116 genes that robustly discriminates between healthy TI and colon samples and validated that finding in an independent data set of whole-genome transcriptomic profiles from normal mucosa. We observed that genes with the least variance in expression in healthy controls showed the greatest variance across uninflamed or inflamed TI and colon in patients with CD. We present

an analysis supporting this finding that corrected for differences in sample size and confirmed this pattern to be true in age-matched patient samples. Using two distinct approaches to coexpression analysis, we identified gene modules in which the pairwise correlation structure was either conserved or differentially regulated from healthy to disease states. A detailed meta-analysis of these modules led to identification of gene pairs for which either the correlation structure was conserved independent of inflammation or the pattern of paired regulation was inverted from healthy to disease states. A systematic differential analysis revealed shared and/or unique genes that were upregulated or downregulated along the spectrum from healthy to uninflamed and inflamed states in the setting of active IBD. On average, we found approximately 60% of disease-associated genes to be differentially expressed in inflamed tissue relative to healthy tissue. Last, we identified eQTLs that were either characterized as having a genotype-only effect on expression independent of disease, tissue type, and inflammation status or as exhibiting an interaction effect between genotype and inflammation.

Results

IBD risk genes exhibit segmental specificity of expression between TI and colon across healthy controls. Clinically, UC is understood to be a colon-specific disease, whereas CD can cause disease anywhere along the gastrointestinal tract with a predilection for the TI in most adult patients. However, the question of whether IBD-associated risk genes demonstrate segmental differential expression between the TI and colon has not been previously addressed. Using tissues from the TI, ascending colon, and descending colon of 37 healthy individuals, 116 genes of the 678 genes on our code set showed significant differential expression (FDR-adjusted $q \leq 0.05$) according to tissue site (Figure 2A; complete heatmap, Supplemental Figure 1). The ascending and descending colon did not segregate independently. The colon-specific genes were not enriched for UC-specific genes, and the TI-specific genes were not enriched for CD-specific genes, but there were some notable exceptions, such as *FAM55D*. Encoded in a UC-specific risk locus (rs561722), *FAM55D* showed a 180- to 190-fold increase in relative expression in the colon versus the TI ($q = 0.006$); FAM55D is a member of a family of neurexophilin and may function as a neuropeptide, but its activity in the colon has not been studied.

To validate these tissue-specific profiles in an independent microarray analysis of TI and colon tissues, we used a publicly available transcriptomic profile of normal mucosa from 12 control subjects (6 colon and 6 TI; ref. 16). Using gene set enrichment analysis, we found a marked enrichment (FDR $q = 0.008$) of our colon upregulated gene set in the independent colon samples (Figure 2B). Likewise, the TI-specific genes were enriched in TI samples (FDR $q = 0.005$; Figure 2C). To further confirm this enrichment, we directly compared the number of significantly upregulated or downregulated genes in both data sets and found that 51 of our signature genes (26 upregulated and 25 downregulated) were also significantly regulated in the independent set ($P = 0.0045$ and $P = 0.0013$, respectively; hypergeometric test). This differential expression analysis of IBD-associated genes between TI and colon identifies tissue-specific marker genes and can inform hypotheses to further characterize distinctions observed between the TI and colon (17, 18).

IBD risk genes show low variance in expression across healthy controls with loss of expression constraint in CD. We next examined the interindividual variation in gene expression within patient and control groups and asked how this changed in the setting of disease. Conceptually, we hypothesized that if these disease-associated genes were important to maintaining intestinal homeostasis, gene expression variance would be constrained in healthy controls, whereas variance in patients would be increased, reflecting a loss in homeostatic balance.

To examine these questions, we first looked at the average expression per sample (mean over all 678 genes) and found that the average expression increased from healthy to disease uninflamed and inflamed samples (Figure 3A and Supplemental Figure 2, A and D; *P* values assessed by a linear mixed-effects model). Intragroup pairwise correlation between samples was consistently high within the healthy control group but dropped from as high as 0.96 to 0.45 between disease inflamed samples (Figure 3B and Supplemental Figure 2, B and E). Within a group, the more highly expressed genes were associated with decreasing coefficient of variation (Figure 3C and Supplemental Figure 2, C and F). The global distribution of average gene expression per gene was the same across all three groups (Supplemental Figure 2, G–H). To test for differences in variance independent of confounding effects of batch number or source institutions, we fit a linear mixed-effects model for each gene and examined the coefficient of variation of residuals for each gene and the fold change in coefficient of variation in disease uninflamed or inflamed setting with respect to the baseline healthy state.
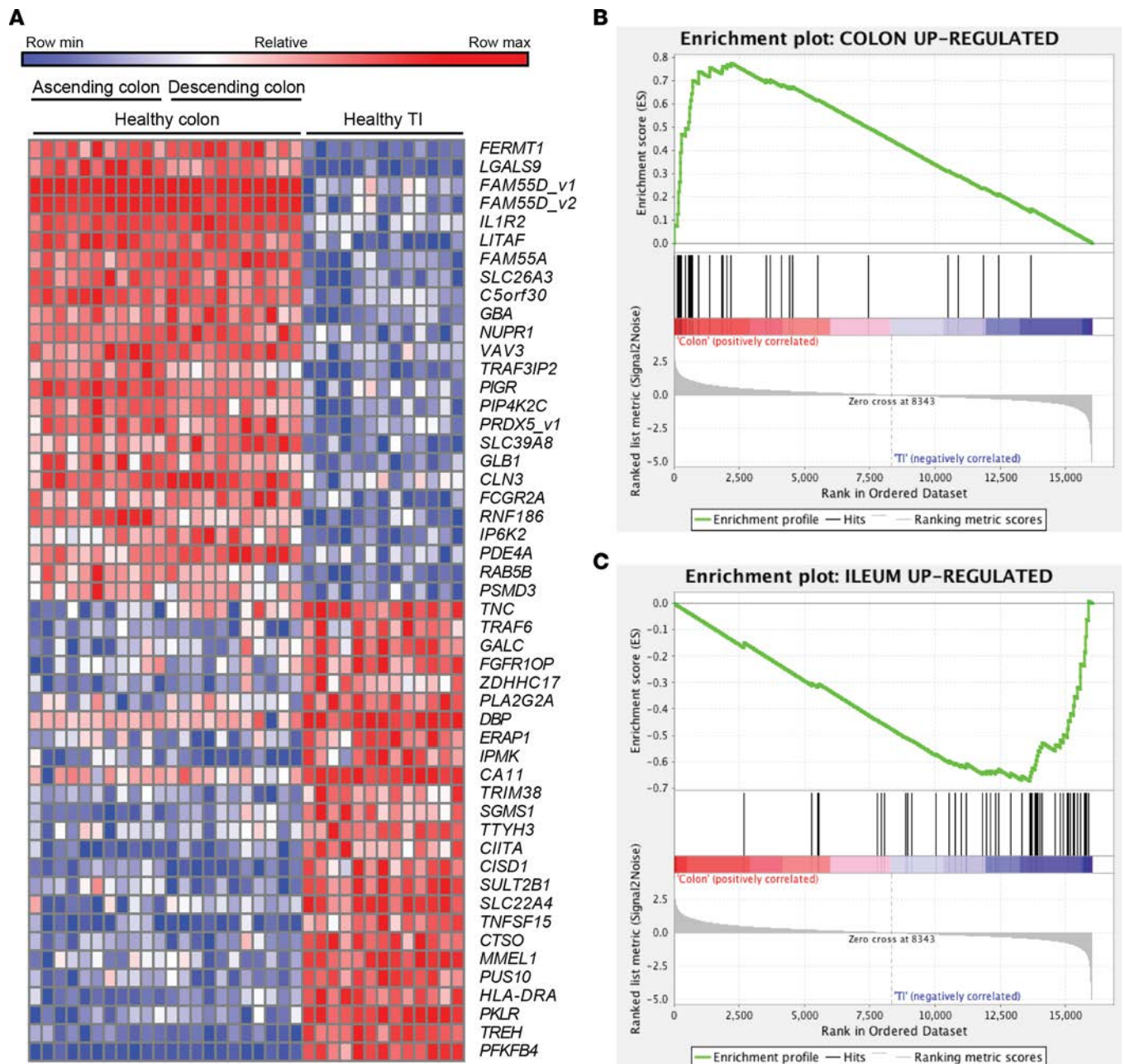
Figure 2. Characterization of mucosal gene expression of inflammatory bowel disease risk genes in healthy terminal ileum and colon tissues. (**A**) In healthy controls, a unique set of genes is differentially expressed in the terminal ileum (TI) versus colon, but not between ascending and descending colon, suggesting tissue specificity between TI and colon, but not specific expression in the ascending or descending colon. The heat map is colored in a row-normalized fashion, i.e., red indicates the highest value for that specific gene, and 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 values of ≤0.05 and signal-to-noise ratios >0.9 (absolute value) were selected as significant. The top 25 upregulated and downregulated significant genes are shown. (**B** and **C**) Using a publicly available microarray data set (GSE16879) for gene expression from 6 healthy TI and colon samples, the differential signature of 116 genes between healthy colon and TI was tested for significance as independent validation. Colon-specific upregulated genes were enriched and upregulated in the colon samples (FDR q = 0.008; gene set enrichment analysis [GSEA] P = 0.0045, hypergeometric test), whereas the TI-specific upregulated genes were enriched in the downregulated genes from the colon samples (FDR q = 0.005; GSEA P = 0.0013, hypergeometric test). The GSEA-based enrichment score (ES) reflects the degree to which a gene set is overrepresented at the top or bottom of a ranked list of genes. GSEA calculates the ES by going down the ranked list of genes, increasing a running-sum statistic when a gene is in the gene set and decreasing it when it is not. The magnitude of the increment depends on the correlation of the gene with the phenotype. The ES is the maximum deviation from 0 encountered in going down the list. A positive ES indicates gene set enrichment at the top of the ranked list; a negative ES indicates gene set enrichment at the bottom of the ranked list.
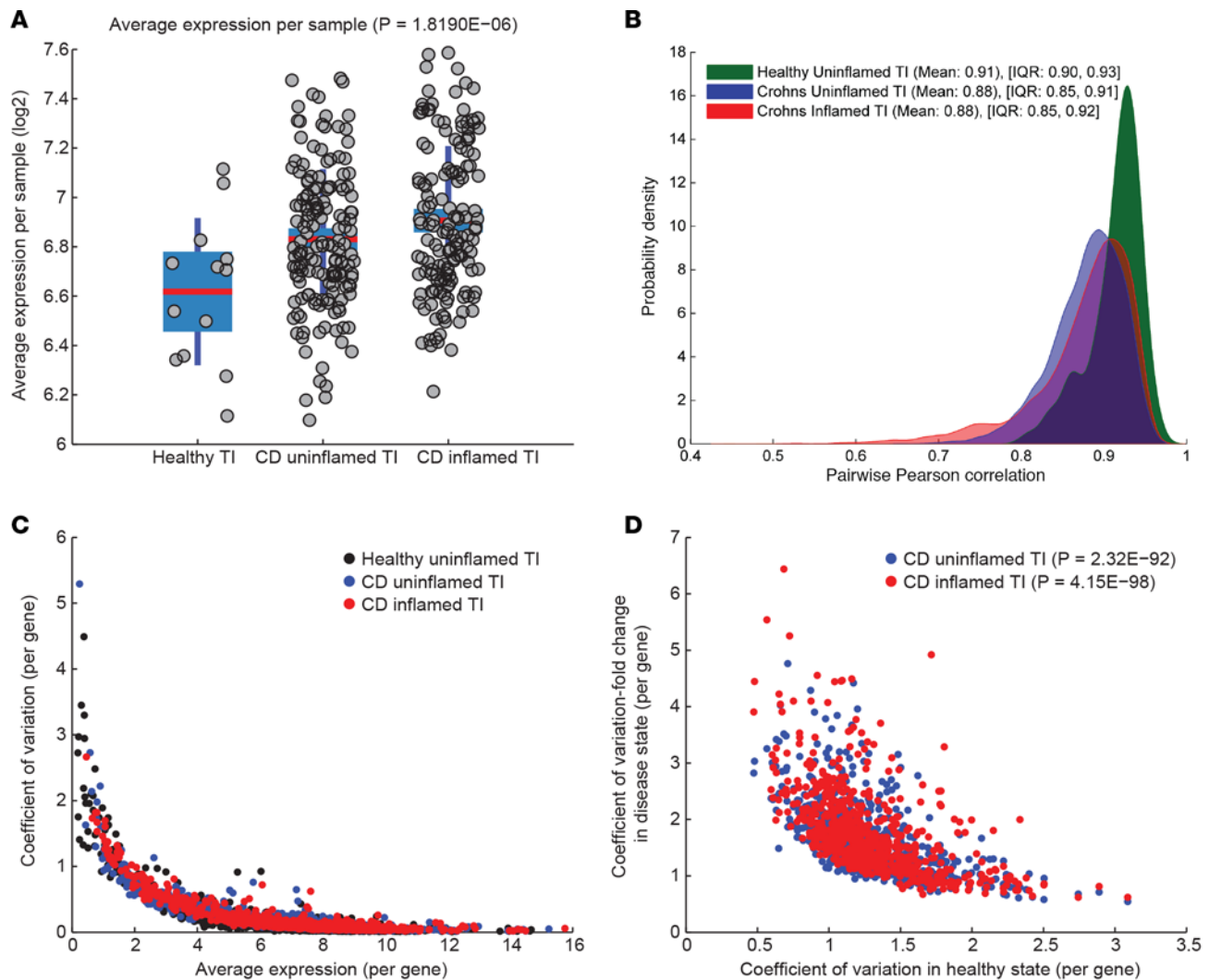
Figure 3. Variance of gene expression increases in patients with Crohn's disease. (A) Average gene expression (per sample) increases from healthy terminal ileum (TI) (*n* = 13) to Crohn's disease (CD) uninflamed TI (*n* = 156) to CD inflamed TI (*n* = 152). *P* value was assessed by a linear mixed-effects model. **(B)** Pairwise Pearson's correlation decreases from healthy controls to CD uninflamed TI to CD inflamed TI (mean, interquartile range), with associated decrease in pairwise correlation supporting increased variance in disease groups. **(C)** Within groups, genes with the highest average expression show the lowest coefficient of variation (per gene) in TI. **(D)** Of the genes with the lowest coefficient of variation in TI in the healthy state, these genes demonstrate the greatest fold change coefficient of variation in the uninflamed or inflamed state in CD; therefore, the genes with the tightest regulation across healthy controls show the greatest heterogeneity in expression in disease. *P* value was estimated by a nonparametric paired Wilcoxon test using all 678 genes.

The key finding of this analysis was the demonstration that the genes with the lowest variance in the healthy control TI and colon samples exhibited the greatest increase in variance in CD patient TI (Figure 3D) and colon samples. This pattern was confirmed when we selected only age-matched patients across three subgroups (Supplemental Figure 2I). We examined the complementary question of the relationship between the genes with the lowest variance in CD and the fold change in variance in healthy controls and did not find the same pattern (Supplemental Figure 3, A and B).This trend was significant (P = 2.32E-92 for uninflamed CD TI vs uninflamed healthy samples, and P = 4.15E-98 for inflamed CD TI vs uninflamed healthy samples) when paired changes in genes were collectively tested with a nonparametric paired Wilcoxon rank test. To validate that this observation was not an artifact due to differences in sample sizes between the three groups, we randomly downsampled an equal number of samples from the three groups 1,000 times. For each random subsampling, we recomputed the linear mixed-effects models and subsequent residuals to examine whether our hypothesis of increase in variance from healthy to disease states remained robust (Supplemental Figure 4A). Under these strict statistical parameters, we found that our hypothesis remained robust for the TI and colon samples from CD patients but did not meet statistical sig-

nificance in the UC patient samples (Supplemental Figure 4B). This finding suggests that mechanisms exist to tightly regulate the expression of these disease-related genes and maintain them at homeostatic levels. The increased variance of expression in disease states may reflect a loss of such regulatory mechanisms.

Coexpression network analysis uncovers modules of genes that are conserved or differentially expressed across tissue and inflammation phenotypes in CD and UC. We next investigated the pairwise correlation structure among all genes and how those relationships are perturbed in disease. We hypothesized that if we could find a set of genes that were correlated and coclustered in all the samples, then this group of genes could be under a common transcriptional regulation program. Likewise, pairs of genes that were differentially coexpressed and for which the correlation structure changed from the healthy state to disease uninflamed or inflamed states would be equally informative. Importantly, members of these coexpression networks could suggest testable hypotheses for novel interactions and functions of disease genes for which limited functional characteristics are known.

We applied weighted gene coexpression network analysis (WGCNA) (19) to 8 data groups stratified by disease, tissue type, and inflammation status (Supplemental Figure 5). We used all 678 genes for consensus eigengene network detection. Briefly, the method uses correlation patterns between genes to construct networks of independent data sets and then uses consensus dissimilarity clustering to identify preserved modules. As shown in Supplemental Figure 6, transcripts were clustered into distinct groups, herein referred to as modules, using dynamic tree cutting. We identified 5 distinct gene modules, which we have depicted with arbitrary colors (blue, brown, green, turquoise, and yellow; Supplemental Figure 6). Gene expression per module was further condensed into eigengene expression (first principal component), and the correlation of each gene to its eigengene expression was quantified (kME). The closer kME is to 1 or –1, the stronger the evidence that the gene is part of that module. The number of genes included in the modules ranged from 50 (green) to 180 (turquoise), and their mean kME values ranged from 0.2349 (blue) to 0.4698 (green). As shown in Supplemental Figure 7, preservation of consensus module eigengenes networks between any two data groups ranged from 0.55 to 0.93. To further identify pairs of hub genes for which the correlation structure was conserved across all data groups, we identified gene pairs for which (a) the correlation scored in the top 1 percentile compared with the null distribution of all pairwise correlation coefficients (Supplemental Figure 8), (b) the kME value for at least one of the genes in the pair scored in the top 1 percentile compared with the null distribution (Supplemental Figure 9), and (c) the FDR-adjusted *P* value of correlation was ≤ 0.05 in all 8 data groups. This analysis resulted in a network of 33 genes, with 73 paired gene interactions (Figure 4A and Supplemental Table 3).

This conserved network showed the greatest enrichment for the T and B cell receptor signaling pathway (Figure 4B; *q* = 7.8E-12 and 3E-4) by gene set overlap analysis and enrichment for ETS2 transcription factor–binding sites (Figure 4C; *q* = 2.54E-9). Upregulation of genes with ETS2-binding sites has been reported in UC patients (20). Among all 33 genes in this network, the top 3 genes with highest number of conserved correlations were *SP140, ARHGAP30*, and *ARHGAP9*. The two isoforms of *SP140* themselves were most highly correlated (average correlation 0.91; turquoise module; Supplemental Table 3), followed by the gene pair *SP140* and *IKZF3* (average correlation 0.88; Figure 5 and Supplemental Figure 10). SP140 contains a bromodomain and a plant homeodomain that recognize and bind acetylated and methylated histones, respectively, and are found in multiple epigenetic "readers" (21). SP140 is hypothesized to act in immune tolerance or acute viral immunity (22). IKZF3 is important in lymphocyte differentiation and proliferation, particularly in promoting Th17 differentiation (23). *SP140* and *IKZF3* share similar patterns of tissue-specific expression, with the highest expression in the spleen, small intestine, and whole blood (24) as well as B cell subsets (25, 26). In addition to IBD, SNPs in both genes have been associated with risk for multiple sclerosis (27, 28).

The initial WGCNA analysis sought to relate genes through conserved correlation of expression, regardless of inflammation or tissue. We then asked the complementary question of which correlation pairs may be changed in the setting of inflammation compared with the healthy state. We applied a method called DiffCoEx (29) that builds on the WGCNA framework to identify differentially coexpressed gene modules separately in our three analysis sets (Supplemental Figures 11 and 12). We identified 35 differentially coexpressed modules (12 in CD TI data sets, 12 in CD colon data sets, and 11 in UC colon data sets; Supplemental Table 4). The number of genes in each module ranged from 24 to 355. We did not find any differentially coexpressed modules when comparing UC inflamed TI samples to UC uninflamed TI samples.
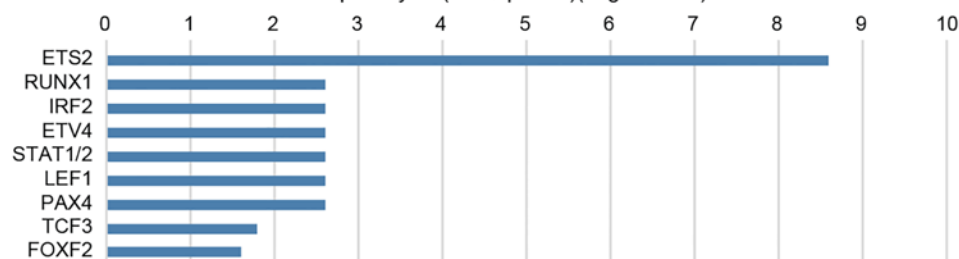
A
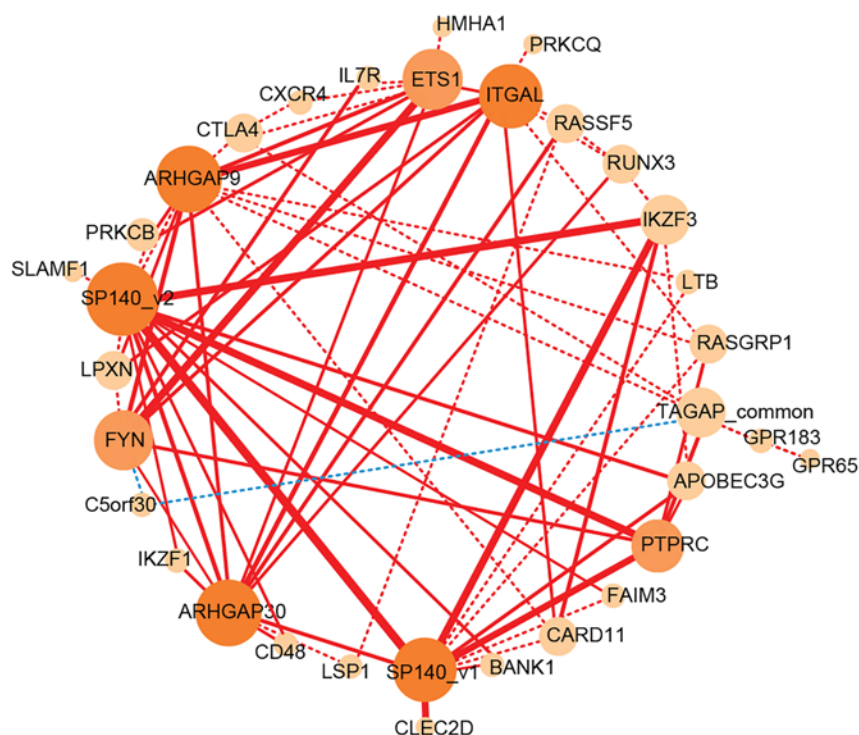
B

Pathway gene set overlap analysis (FDR q values) (-log10 scale)

C
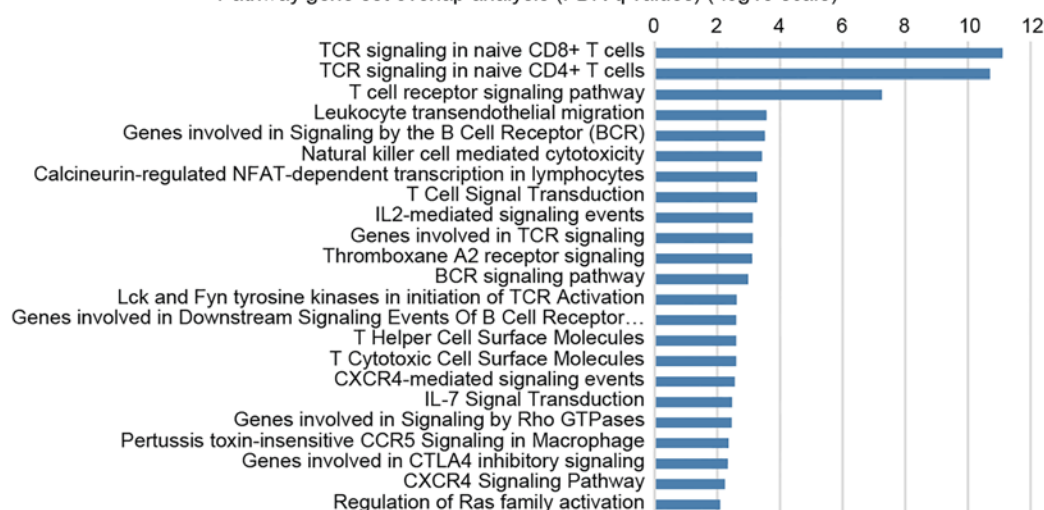
Gene set overlap analysis (FDR q value)(-log10 scale)

Figure 4. Network of correlated hub genes conserved across distinct data groups. (**A**) Gene pairs were selected based on significance of pairwise correlation across all data groups and high correlation of gene expression with eigengene profile (kME value). Genes are shown as nodes, and pairwise correlations are displayed as edges (red, positive correlation; blue, negative correlation). Larger nodes denote higher kME values, and thicker edges represent stronger absolute correlations. (**B** and **C**) Gene set overlap analysis of conserved network genes, as evaluated with Molecular Signatures Database (MSigDB) database. P values were estimated using a hypergeometric test. FDR q values are plotted on a $-\log_{10}$ scale. Q value > 1.3 is significant. Enrichment of pathways is shown in **B**. Enrichment for transcription factor binding sites is shown in **C**.
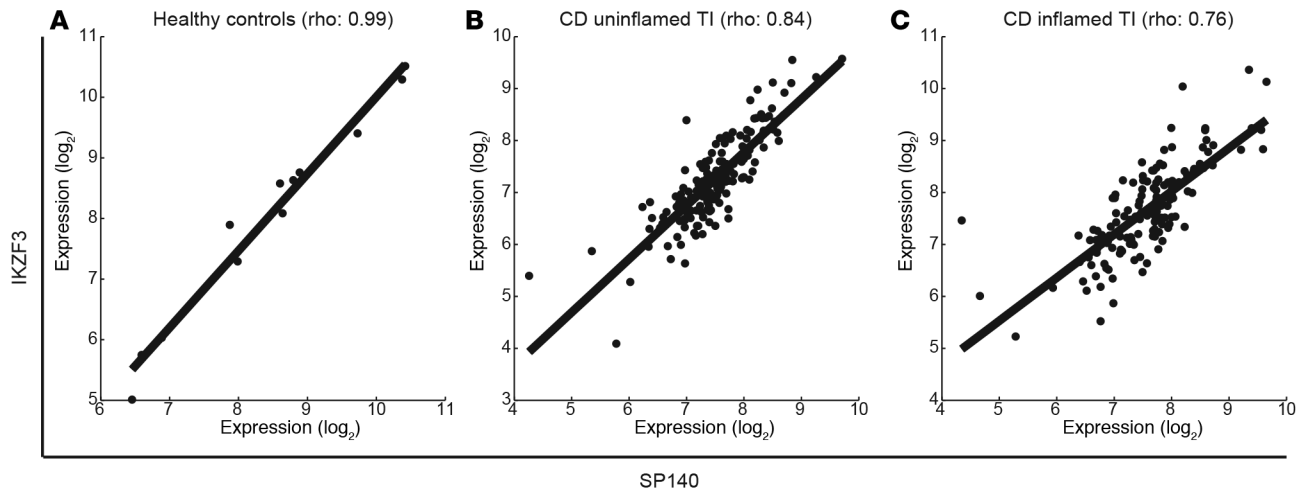
Figure 5. Pairwise correlation between *SP140* and *IKZF3* conserved across data groups. Pairwise correlation between *SP140* and *IKZF3* is conserved across disease and tissue groups, independent of inflammation status. Expression correlation in terminal ileum (TI) tissue samples from (**A**) healthy controls, (**B**) Crohn's disease (CD) uninflamed, and (**C**) CD inflamed groups, respectively. Linear Pearson's correlation coefficient was estimated using 13 healthy TI samples, 156 uninflamed CD TI samples, and 152 inflamed CD TI samples.

We further evaluated each of these gene modules and identified transcription factors that were significantly enriched and bind to the promoter regions of genes (within –2 kb to +2 kb of the transcription start site) in these modules (Supplemental Table 5). Immune-regulated transcription factors were enriched across all modules, including NF-κB, REL, NFAT, STAT1, STAT3–STAT6, TCF1, TCF3, TCF8, IRF1, IRF2, IRF7, and IRF8. Notably, ELF1, ETS, NF-κB, and RUNX1, found to be enriched in one or more differentially coexpressed modules in our data, have been reported to be enriched with autoimmune SNP variants that disrupt the binding site motif for these transcription factors (27). Additionally, nuclear receptors with known links to intestinal immunity and epithelial integrity, such as vitamin D receptor (VDR) (30–32), sterol regulatory element binding transcription factor 1 (SREBF1) (33, 34), and aryl hydrocarbon receptor nuclear translocator (ARNT) (35, 36), were also enriched in one or more modules. We observed the greatest change in number of differentially coexpressed modules between healthy and disease states rather than between disease uninflamed and inflamed states.

To identify gene pairs within these modules with the most significant changes in pairwise correlation coefficient across groups, we estimated (a) the empirical *P* value of change in correlation with respect to null distribution of all possible pairwise values of change in correlation coefficient (Supplemental Figure 13–15) and (b) FDR-adjusted *P* values for all pairs of genes. We identified 34 gene pairs (36 unique genes) with significant changes in correlation from the healthy state to disease uninflamed and inflamed states (Supplemental Table 6). Recent studies offer insights into potential roles in IBD-relevant pathways for the proteins encoded by these genes, such as C1ORF106 in autophagy-dependent intracellular pathogen defense (37) and C5ORF30 in negative regulation of immune and inflammatory pathways (38), but the activities of many of the genes remain to be studied in the context of IBD or other immune-mediated diseases.

In general, these networks represent clusters of genes within which transcriptional regulation is simultaneously reprogrammed in the inflammatory state (example networks, Figure 6; example plots, Figure 7; and Supplemental Figure 16). As an example, *SULT1A1* and *SMAD3* were negatively correlated in healthy TI with higher expression of *SULT1A1* correlating with lower *SMAD3* (Figure 7A); yet in CD, both in uninflamed and inflamed TI, the gene pair was positively correlated (Figure 7, B and C). SULT1A1 is a sulfotransferase important in xenobiosis, shown to be induced by TGF-β (and TNF-α) in a colon cancer cell line (39). This pair of genes may be connected through TGF-β, given that *SMAD3* is directly downstream of TGF-β signaling, a pathway tied to fibrostenotic disease in CD, or another shared ligand. CD-associated risk alleles in *SMAD3* have been associated with increased risk of recurrent surgeries in CD (40), underscoring its association with more severe disease (40). Further study will be required to understand the opposing correlation phenotypes between healthy and disease states.
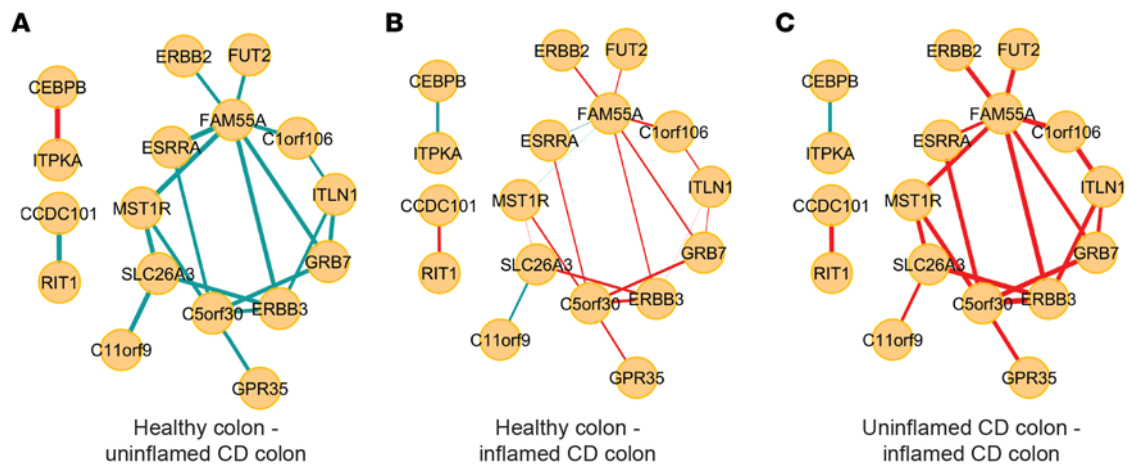
Figure 6. Network visualization of differentially coexpressed module changes in colon tissue of healthy and Crohn's disease patients. (**A**) Differential coexpression module from healthy state to uninflamed Crohn's disease (CD) state. (**B**) Differential coexpression module from healthy state to inflamed CD state. (**C**) Differential coexpression module from uninflamed CD state to inflamed CD state. Gene pairs were selected based on whether the Pearson's correlation coefficient exhibited the most significant change in all 3 pairwise comparisons between 3 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. Linear Pearson's correlation coefficient was estimated using 24 healthy colon samples, 177 uninflamed CD colon samples, and 79 inflamed CD colon samples.

Differential gene expression across patient groups compared with controls demonstrates disease- and tissue-specific patterns. We next evaluated the patterns of distinct and overlapping expression of risk genes, comparing CD and UC, to further characterize IBD-associated risk genes in disease-relevant tissues (Figure 8). We used linear mixed models to identify IBD-associated risk genes that were differentially expressed between the healthy state and disease uninflamed or inflamed states (Supplemental Figure 17). In this model, the uninflamed tissues from patients served as an intermediate state between healthy and inflamed tissues, in line with prior data supporting baseline expression differences in intestinal mucosal tissues devoid of active inflammation in patients with IBD (41).

We found 385 differentially expressed genes in CD TI (example plots, Figure 8, A–C, and Supplemental Table 7), 431 in CD colon (example plots, Figure 8, D–F, and Supplemental Table 8), and 439 in UC colon (example plots, Figure 8, G–I, and, Supplemental Table 9) compared with the healthy state. In the colon, CD and UC shared a large set of differentially expressed genes (~88%, 376 genes in common). The most significantly upregulated gene between healthy TI and CD TI was Fc γ receptor 3A (*FCGR3A*; FDR 3.18E-25; Figure 8A); additional upregulated gene family members included *FCGR3B* (FDR 4.30E-24) and *FCGR2A* (FDR 1.99E-22), suggesting a potential common regulator in the TI. *FCGR3A* and *FCGR3B* are among the top 20 most significantly upregulated genes in a treatment-naive pediatric ileal CD population (41). *FCGR3A* (FDR 1.10E-17) and *FCGR3B* (FDR 7.56E-21; Figure 8D) were also upregulated in inflamed CD colon. Upregulation of the decoy receptor for Fas ligand *TNFRSF6B* (FDR 8.06E-26, Figure 8E) was found only in CD colon, not in CD TI. Exploring these unique patterns may shed light on similarities and differences between small intestine and colon inflammation.

Other findings included downregulation of *VDR* in both UC (FDR 8.17E-21) and CD (FDR 5.18E-18; Figure 8F). In patients, serum vitamin D concentrations and UC disease activity are inversely related (42). The role of vitamin D in intestinal homeostasis has been shown through regulation of autophagy (43) and tight junctions (44). Genetic polymorphisms in *VDR* affect variations in serum vitamin D concentrations (45), and further studies have shown that VDR mediates microbe-host interactions (32). Consistent with a prior report (46), the cation/carnitine transporter *SLC22A5* showed universal downregulation across CD TI (FDR 1.42E-12), CD colon (FDR 2.31E-16), and UC colon (FDR 1.37E-27; Figure 8H). Supplementation with carnitine, proposed to function as an antioxidant through fatty acid oxidation and immunosuppressive properties (47), has shown efficacy in treating UC (48).

In contrast to the large number of differentially expressed genes found by comparing healthy and dis-
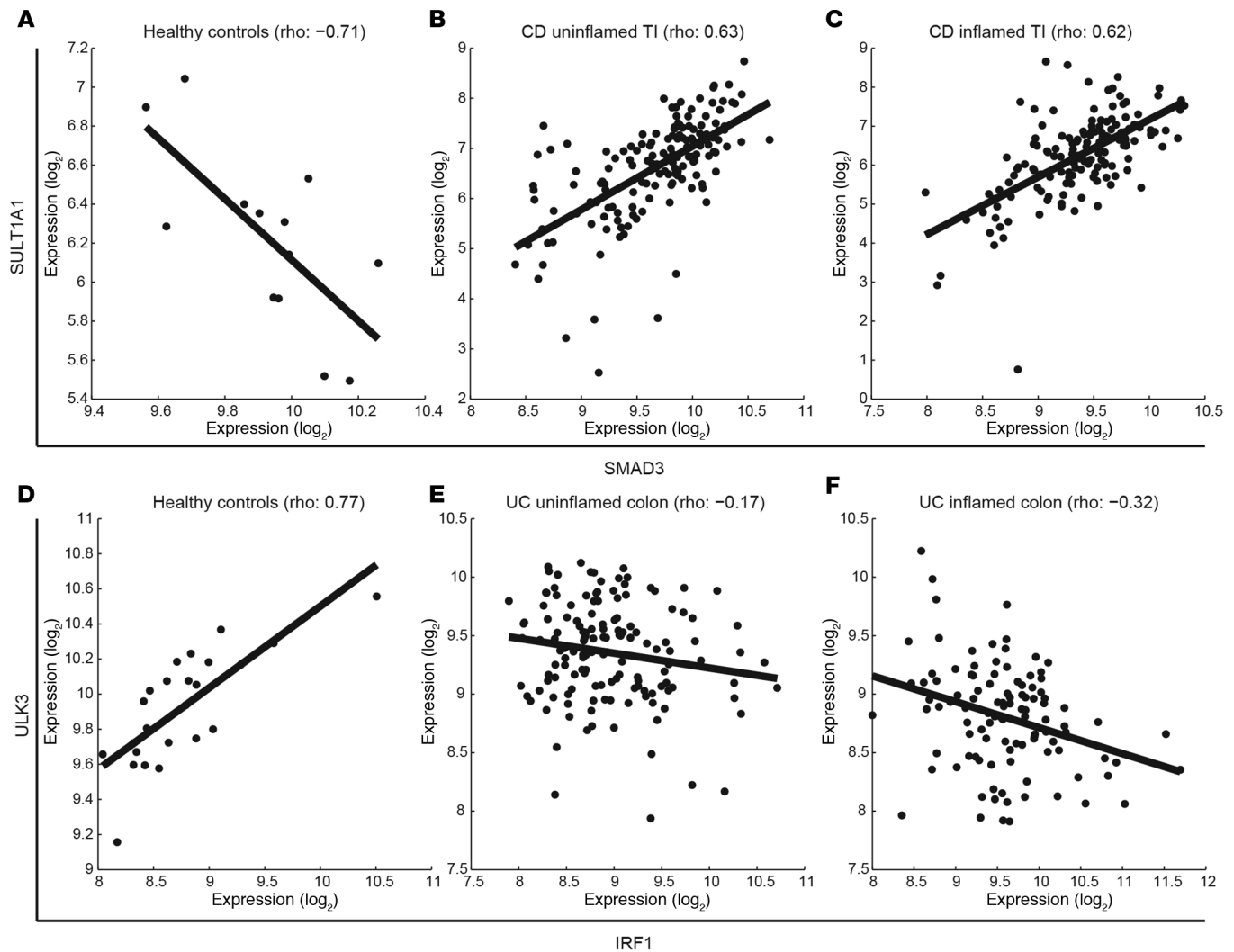
Figure 7. Pairwise correlations show differential directionality of correlation between controls and disease groups. *SULT1A1* and *SMAD3* show a correlation coefficient of −0.71 in (**A**) healthy terminal ileum (TI) but are positively correlated in (**B**) uninflamed Crohn's disease (CD) TI (rho 0.63) and (**C**) inflamed CD TI (rho 0.62). Linear Pearson's correlation coefficient was estimated using 13 healthy TI samples, 156 uninflamed CD TI samples, and 152 inflamed CD TI samples. *IRF1* and *ULK3* show a correlation coefficient of 0.77 in (**D**) healthy colon but a negative correlation in (**E**) uninflamed ulcerative colitis (UC) colon (rho −0.17) and (**F**) inflamed UC colon (−0.32). Linear Pearson's correlation coefficient was estimated using 24 healthy colon samples, 142 uninflamed UC colon samples, and 107 inflamed UC colon samples.

ease states, differential analysis of uninflamed TI/colon samples from CD and UC patients yielded only 14 significant genes (6 in TI and 8 in colon; Supplemental Table 10), suggestive of CD versus UC specificity (Supplemental Figure 18). The gene with the greatest differential expression between the three groups was IFN regulatory factor 4 (*IRF4*), with the highest expression in uninflamed CD colon. IRF4 is a master transcription factor for Th17 cells (49) and CD11b+ dendritic cell differentiation that further drives Th17 responses (50). Th17 cells play complex roles in IBD, with contributions to both inflammatory and immunoregulatory responses (51, 52). IRF4 is induced by NOD2 and negatively regulates the innate immune response by inhibiting the polyubiquitination of RICK and TRAF6 to downregulate NF-κB, protecting mice from colitis in experimental models (53).

eQTL analysis highlights effects of disease-associated genetic variation on gene expression. We conducted an eQTL analysis to examine how disease-associated variants regulate gene expression in gut tissue. We extended the linear mixed models from our differential expression analysis to include and test the effect of genotype and the interaction between genotype and the independent variables of disease, tissue type, and inflammation status (Supplemental Figure 19) using samples from healthy controls and the uninflamed and inflamed samples from the CD and UC patient groups. The model merged patients and controls heterozygous or homozygous
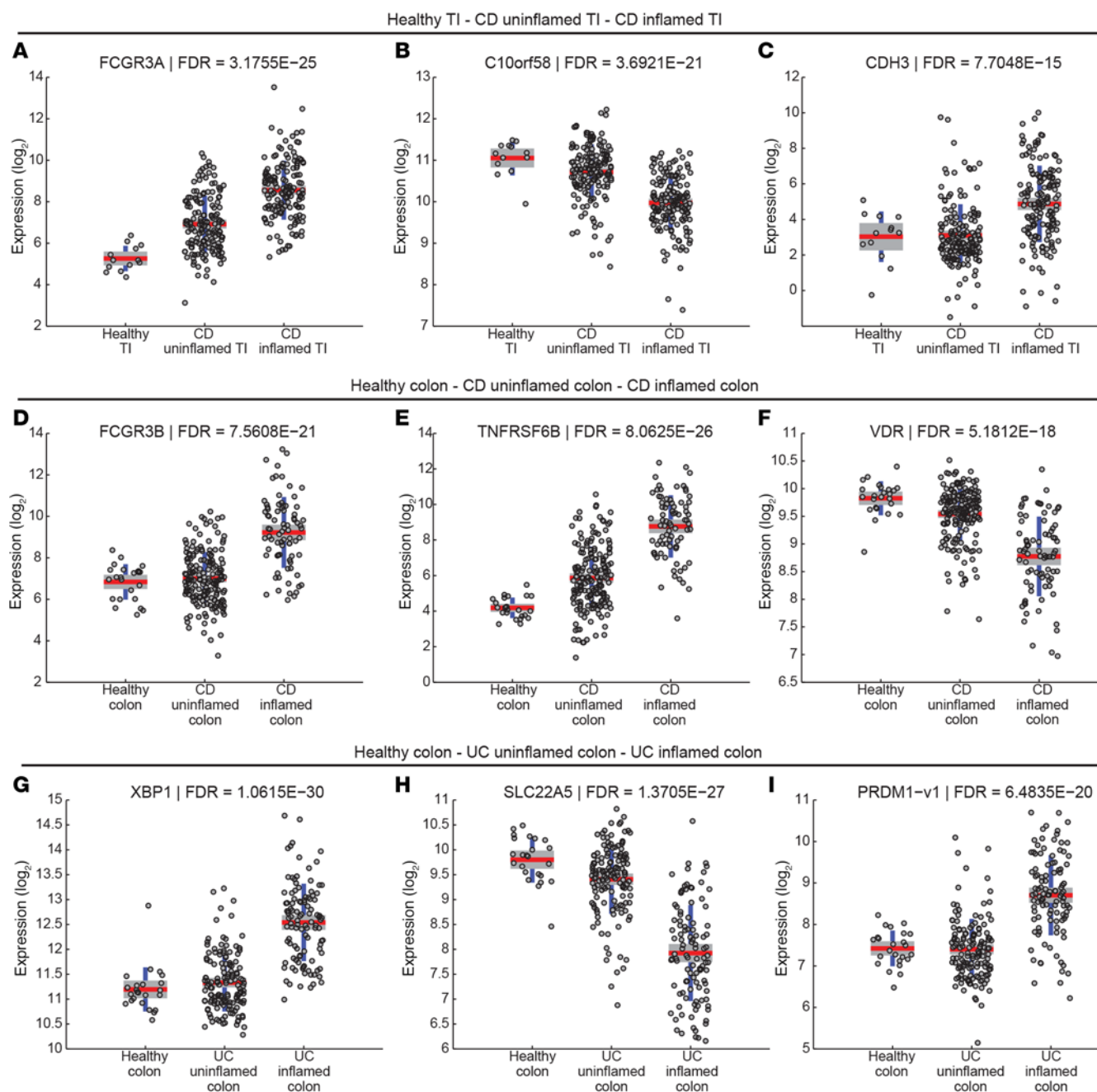
Figure 8. Top differentially expressed genes across healthy controls and uninflamed and inflamed disease states. (**A**) *FCGR3A* (FDR = 3.1755E-25), (**B**) *C10orf58* (FDR = 3.621E-21), and (**C**) *CDH3* (FDR = 7.7048E-15) show the most statistically significant differential expression across healthy terminal ileum (TI), Crohn's disease (CD) uninflamed TI, and CD inflamed TI. A linear mixed-effects model was used to estimate significance of difference in gene expression. Genes with FDR adjusted *P* values of ≤0.05 were considered significant. 13 healthy TI samples, 156 uninflamed CD TI samples, and 152 inflamed CD TI samples were used for this analysis. (**D**) *FCGR3B* (FDR = 7.5608E-21), (**E**) *TNFRSF6B* (FDR = 8.0625E-26), and (**F**) *VDR* (FDR = 5.18E-18) show the most statistically significant differential expression across healthy colon, CD uninflamed colon, and CD inflamed colon. A linear mixed-effects model was used to estimate significance of difference in gene expression. Genes with FDR adjusted *P* values of ≤0.05 were considered significant. 24 healthy colon samples, 177 uninflamed CD colon samples, and 79 inflamed CD colon samples were used for this analysis. (**G**) *XBP1* (FDR = 1.0615E-30), (**H**) *SLC22A5* (FDR = 1.3705E-27), and (I) *PRDM1* (FDR = 6.4835E-20) show the most statistically significant differential expression across healthy colon, ulcerative colitis (UC) uninflamed colon, and UC inflamed colon. A linear mixed-effects model was used to estimate significance of difference in gene expression. Genes with FDR-adjusted *P* values of ≤0.05 were considered significant. 24 healthy colon samples, 142 uninflamed UC colon samples, and 107 inflamed UC colon samples were used for this analysis.

for the IBD-associated risk allele compared with those homozygous for the non-risk allele. Candidate eQTLs were selected based on *P* values of coefficients for select variable terms (genotype and interaction terms) in the linear model and further validated by independently permuting genotype and gene expression 1,000 times (Supplemental Figure 20). We identified 34 eQTLs that were categorized to have a genotype-only effect, an interaction-only effect, or both. Key significant eQTLs, with tissue and disease specificity and associated *P* values, are shown in Figure 9; all eQTLs are provided in Supplemental Table 11.

The identified eQTLs include examples of differential gene expression driven only by genotype (genotype effect) and others in which the eQTL effect emerges in the setting of inflammation (interaction effect). The eQTL effect between rs1363907 and endoplasmic reticulum aminopeptidase 2 (*ERAP2*), which plays a central role for peptide trimming and antigen presentation on major histocompatibility complex I, is one example of genotype-only effect across CD TI (Figure 9A), CD colon (Figure 9D), and UC colon (Figure 9G). This finding replicates an eQTL previously reported in ileal (11) and colonic tissue (12, 13). Our data extend these previous findings by describing how the inflammatory state changes *ERAP2* expression (increased in CD colon and UC colon with inflammation but unchanged by inflammation in the CD TI). Illustrating how this analysis builds on prior eQTL studies and incorporates the effect of inflammation, the genotype-only eQTLs can be correlated with the interaction-only effect of rs10758669 and B cell scaffold protein with ankyrin repeats 1 (*BANK1*), in which *BANK1* expression is increased in the setting of inflammation in CD TI for individuals heterozygous or homozygous for the risk allele, whereas expression is decreased in those homozygous for the non-risk allele (Figure 9B). This eQTL analysis serves to illustrate how disease-associated genetic variants effect gene expression in the gut and in the setting of active disease.

Discussion

An important challenge in the GWAS era is the harnessing of knowledge gained by understanding the genetic risk of disease to inform studies of intestinal homeostasis and IBD pathophysiology. Using tissue-, disease-, and inflammation-specific differential expression, coexpression, and eQTL analysis, we sought to characterize the expression of candidate genes in IBD-associated risk loci.

Our approach addressed feasibility challenges inherent to studying a large number of genotyped patients with tissue samples, making it possible to examine differences between CD and UC, uninflamed and inflamed tissues, and TI and colon, all on the same transcriptomic platform. Our data highlight the segmental expression of IBD-associated risk genes across the TI and colon. The coexpression analysis demonstrates how disease-associated genes are interrelated and how these gene-gene relationships are changed in the setting of IBD. Multiple studies have suggested that coexpressed genes share similar biological function (54) with biological drivers of coexpression (and tissue-specific patterns), including transcription factors, spatial configuration of chromosomes, mRNA degradation, miRNAs, and epigenetic modulation (55, 56); we too found pathway and transcription factor enrichment (Figure 4) within our network of correlated hub genes conserved across tissue types.

We expanded traditional differential expression analyses by examining variance of expression between healthy controls and patients with IBD. The inclusion of healthy controls permitted exploration of how the genes are expressed in health, representing a state of homeostasis. The age of the healthy control population was older than that in the disease group, yet in a subgroup analysis of age-matched patients only, our significant observations, including the change in variance, followed the same trends (Supplemental Figures 2I and 21–24). One model in the literature suggests that disease-associated genes demonstrate greater stochastic variation in health (57), and alterations in variance can define disease phenotypes (58); yet these questions have not been examined in the context of the gut or IBD. After accounting for batch and institutional effects, we considered cellular heterogeneity as a potential driver of the increased variance in CD patients; however, two observations argue that heterogeneity in cellular composition is unlikely to explain this finding. First, we found that the observation was also true in the uninflamed tissues, in which the cellular composition more closely aligns with healthy tissue. Second, the trend of increased variance did not hold in the UC patients, in which cellular heterogeneity would be present in the inflamed samples. These data, coupled with the differential expression data, particularly in the uninflamed samples compared with the healthy controls, suggest that the IBD-associated risk genes are important in intestinal homeostasis and exhibit perturbed expression, even in the uninflamed state in patients with IBD. The variance in gene expression suggests that these genes are otherwise held under tight regulatory constraints that are lost in the setting of CD and may represent a novel phenotypic marker of disease. Although consensus clustering
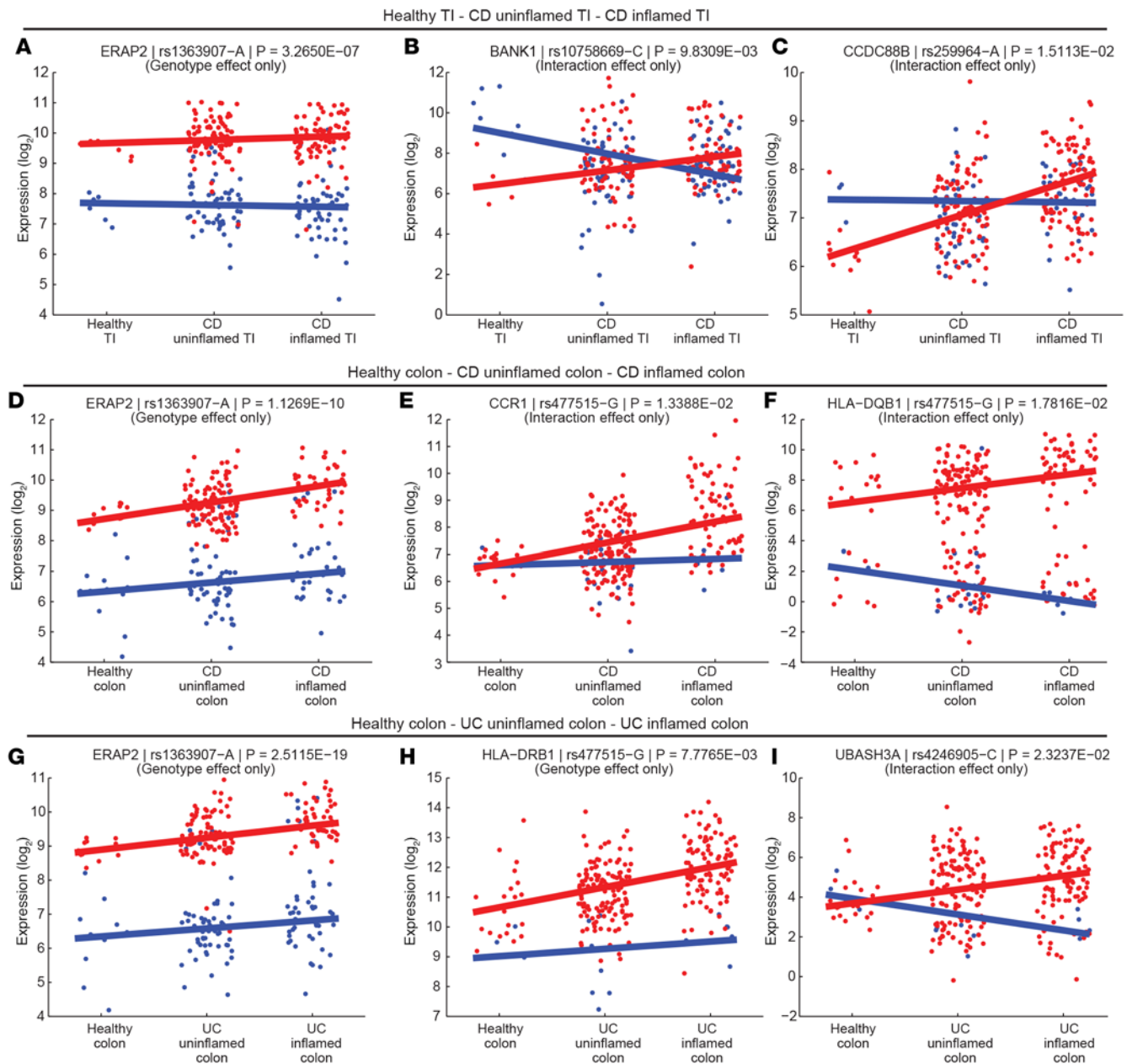
Figure 9. Top expression quantitative trait loci (eQTL) identified with genes demonstrating differential expression across healthy controls and disease groups. Log$_2$ gene expression depicted in blue corresponds to healthy controls and patients homozygous for the inflammatory bowel disease–associated (IBD-associated) non-risk allele; log$_2$ gene expression depicted in red corresponds to healthy controls and patients heterozygous or homozygous for the IBD-associated risk allele. A linear mixed-effects model was used to estimate significance of effect of genotype only and interaction of genotype with tissue inflammation status on gene expression. Genes with P values of ≤ 0.05 for one of the three terms (genotype, tissue inflammation status, or interaction term) were considered significant. Significance was further validated by permuting genotype and gene expression independently 1,000 times each. 13 healthy terminal ileum (TI) samples, 156 uninflamed Crohn's disease (CD) TI samples, 152 inflamed CD TI samples, 142 uninflamed ulcerative colitis (UC) colon samples, and 107 inflamed UC colon samples were used for this analysis. Genotype-only effects are shown in **A**, **D**, **G**, and **H**; interaction-only effects are shown in **B**, **C**, **E**, **F**, and **I**. rs1363907 and *ERAP2* show the strongest cis-eQTL effect across tissue and disease groups (TI shown in **A**, colon shown in **D** and **G**).

analysis in this data set failed to elucidate subgroups, we envision that whole-genome transcriptomic profiles will be better equipped to identify subgroups of patients.

Demonstrating the value of a tissue- and disease-based approach to study the effects of disease-associated genetic variation, this work has relevance to other complex, multigenic autoimmune and inflam-

matory diseases. Refining candidate genes in IBD and autoimmune/autoinflammatory disease-associated risk loci has advanced by exome-sequencing studies, fine mapping of high confidence risk loci, and eQTL analysis (27, 59–61). However, the major integrative analyses with eQTLs, such as those for rheumatoid arthritis (61), systemic lupus erythematosus (62), and multiple sclerosis (63), have used eQTLs defined in peripheral blood (64) or immune cell subsets (63, 65). eQTL mapping in neutrophils (66, 67), monocytes (10), and dendritic cells (68) has expanded by comparing resting cells with cells after immune stimulation. These "induced" or "response" eQTLs are enriched for GWAS-identified loci, supporting the hypothesis that these loci have context-specific (including inflammation-specific) effects. The approach employed in our work allowed for a demonstration of how genetic variation affects differential gene expression under in vivo inflammatory conditions in disease-relevant tissues.

Taken together, these data reinforce the tissue specificity of gene expression in IBD-associated risk loci and the value of studying genes in the context of intestinal tissue and IBD-associated inflammation. We offer these data, grounded in our understanding of the genetic susceptibility to disease, tissue-based differential expression, and eQTL effects, to prioritize candidate genes and to inform future functional studies that will use genetics to advance the understanding of IBD pathogenesis (Supplemental Table 12).

Methods

Patient recruitment and sample collection. Prospectively recruited patients and archived patient samples were collected at 5 academic institutions throughout the United States. Patients were recruited after a known diagnosis of CD or UC. Healthy controls were recruited at time of routine colonoscopy and were considered healthy if they reported no history of gastrointestinal symptoms (abdominal pain or diarrhea) for >1 day per week in past 6 months, no history of IBD or irritable bowel syndrome, no history of asthma or autoimmune disease, and no personal or family history of colon cancer. At the time of biopsy, healthy controls had a median age of 59 years, within a range of 42–69 years of age. Median age for patients at the time of biopsy was 39–40 years, but patients had a much larger age range that spanned from 15 to 79 years and included individuals in most decades in this range. Blood was banked for DNA extraction and genotyping. At the time of endoscopy, mucosal biopsies were collected in RNAlater (Qiagen) and stored at –80°C ahead of downstream processing. If samples were collected at the time of surgery, mucosal tissue samples obtained from bowel resection specimens with colonoscopic biopsy forceps were processed immediately or stored at 4°C for processing within 12 hours. Intestinal source location and inflammation status was indicated as "uninflamed" or "inflamed" based on the endoscopic or gross appearance of the tissue by the endoscopist or investigator at time of surgical specimen collection. All biospecimens were gifted with patient consent (see *Study approval*) for this study by coinvestigators (R. Balfour Sartor, Rodney D. Newberry, Dermot P. McGovern, Vijay Yajnik, Sergio A. Lira, and Ramnik J. Xavier).

Genotyping. Genomic DNA was extracted from buffy coats, with genotyping performed at the Broad Institute or Cedars-Sinai Medical Center on the Illumina Immunochip array containing 196,806 SNPs and 718 small insertions/deletions associated with 11 autoimmune or inflammatory diseases (15). Five hundred and ninety-three patients were genotyped on this platform. For this study, we restricted our analysis to the 163 SNPs identified as associated with CD, UC, or both diseases as previously published (1) and focused transcriptional analysis as described below. Of the 163 SNPs, 159 SNPs passed QC (genotyping call rate >99%, minor allele frequency >5%) on both genotyping platforms and were included in downstream analysis.

Tissue total RNA extraction or cell lysis. More than 1,100 tissue samples were collected. Using a combined TRIzol (Life Technologies) and chloroform extraction and the RNeasy Midi Kit (Qiagen) for RNA extraction, total RNA was extracted by the manufacturer's protocol, and RNA purity and quantity was measured by NanoDrop spectrophotometer (Promega).

NanoString code set construction. The custom NanoString probe set targeted 678 unique genes, including 15 housekeeping genes (Supplemental Table 2). Of these genes, 156 (23%) were not captured on the Agilent microarray platform that was used in the largest published microarray study examining gene expression in IBD (2, 3). Four-hundred and sixty-three of the genes fall within a 500-kb window flanking a tagged risk SNP as previously published (1). We prioritized these genes from the >1,400 genes encoded within the loci based on known expression patterns in cell types of interest in IBD (e.g., intestinal epithelium, immune cell types), functional annotation where available, and genes previously implicated in eQTL analyses in previous studies. The core set of IBD-associated risk genes on the code set was taken directly from the

prioritized genes in the Jostins et al. IBD GWAS meta-analysis (1). The additional genes in the probe set were selected for activity in the immune response and in inflammation, with emphasis on genes implicated in type 1 diabetes and celiac disease, given the genetic overlap in susceptibility loci for these diseases (1, 69) (Supplemental Table 2, for a few select genes, we added probes for multiple isoforms of the same genes and these are named as *_v1 or *_v2).

NanoString gene expression profiling. Extracted RNA was used as input for expression profiling by NanoString. All samples were prepared according to the manufacturer's specifications (NanoString Technologies). Briefly, input RNA (100 ng) was hybridized with reporter probes to specific transcript tags in a multiplexed ligation reaction using reverse complementary bridge oligonucleotides. Spiked-in controls (in addition to housekeeping genes) were used in each reaction to permit normalization. After the ligation reaction, tagged transcripts were hybridized to tag-specific capture probes with attached fluorescent bar codes by incubation at 65°C for 14 to 28 hours. Excess probes were removed on the nCounter Prep Station according to the manufacturer's instructions (NanoString Technologies), and the captured, transcript-specific, bar-coded ternary complexes were immobilized on a streptavidin-coated cartridge. The individual bar codes were counted on the nCounter Digital Analyzer (NanoString Technologies) using a high density of fields of view (>600) to quantify target RNA molecules present in the sample. Data are accessible in the GEO database under accession number GSE73094.

NanoString data normalization. NanoString gene expression profiles were generated using 3 distinct versions of the probe sets over a span of 3 years. Explicit normalization was required to correct for probe version and batch-level differences. This was achieved by a 2-stage normalization process. In the first stage, each lot of data pertaining to a specific version of the probe set was normalized independently using the following 6 steps. (a) Background noise was estimated using mean expression levels of spiked-in negative controls. (b) Estimated noise value was subtracted from raw counts data for each sample. Negative count values were reset to 1. (c) Count values of spike-in positive controls were summed per sample, and the average value was used to estimate a scaling factor for each sample. The expression value of each sample was adjusted using the sample-specific normalization factor. (d) Samples with higher than acceptable background noise and positive control-based scaling factors of <0.3 or >3.0 were identified as outliers and removed from analysis. (e) Next, the geometric mean of housekeeping genes was computed for each sample, and the average value was used to compute the normalization factor. Data were adjusted in a similar fashion as was done previously with spiked-in positive controls. (f) Finally, samples with housekeeping gene–based normalization factors of <0.2 or >5.0 were removed as outliers. In the second stage of normalization, data generated from first version of probe sets (batch 1) were chosen as the common reference to which the remainder of the data (batches 2 and 3) were independently calibrated using a small set of samples that were repeatedly profiled in all 3 lots. This was achieved with the following steps. (a) The average of medians of calibration samples, for a given pair of lots, was computed and used to estimate the normalization factor for each sample. (b) Each lot was scaled with the normalization factor computed in the previous step. (c) Geometric means of counts per gene among repeated samples was computed within a lot, and a calibration factor was estimated per gene (4). Calibration for each gene was applied to the nonreference lot. All computations were performed in MATLAB. Nine hundred and eighty-nine samples passed data normalization and quality control.

Statistics. Statistical significance of genes between TI and colon samples was estimated by computing the signal-to-noise ratio statistic. Genes with FDR-adjusted *P* values of ≤0.05 and signal-to-noise ratios of >0.9 (absolute value) were selected as significant. Healthy TI and colon samples profiled on our NanoString platform were compared with microarray profiles of TI and colon samples from healthy (control) individuals used in a publicly available data set (GSE16879) (16). Normalized and log$_2$-transformed data from the selected 560 genotype patients, amounting to 886 samples, were stratified by disease, tissue type, and inflammation status (Figure 1). Three analysis groups were formed: (a) healthy versus CD uninflamed and inflamed TI samples, (b) healthy versus CD uninflamed and inflamed colon samples, and (c) healthy versus UC uninflamed and inflamed colon samples (Supplemental Table 1). Analysis of coefficient of variation, consensus coexpression, differential coexpression, differential expression, and eQTLs was conducted as outlined in Supplemental Figures 4, 5, 11, 17, and 19, respectively. FDR-adjusted *P* values, estimated by the Benjamini-Hochberg method, were used in all analysis unless stated otherwise. The threshold for significance was FDR-adjusted *P* value ≤ 0.05.

Study approval. This study was carried out with approval from the institutional review boards at Cedars-Sinai Medical Center (CR00009268), Icahn School of Medicine (HSM#11-01669), Massachusetts General Hospital (2011P000397), University of North Carolina at Chapel Hill (11-0359), and Washington University School of Medicine (201204075). Participants or their guardians provided written informed consent.

Author contributions

JMP, GG, RBS, RDN, DPM, SAL, and RJX conceived and designed the experiments. JMP performed the experiments. JMP, GG, LK, HH, TH, MJD, SAL, and RJX analyzed the data. GG, RBS, RDN, DPM, VY, SAL, and RJX contributed reagents, materials, and/or analysis tools. JMP, GG, LK, SAL, and RJX contributed to the writing of the manuscript. JMP, GG, and RJX approved the final version of the manuscript.

Address correspondence to: Ramnik J. Xavier, Massachusetts General Hospital, Center for Computational and Integrative Biology, 185 Cambridge Street – Simches 7222, Boston, Massachusetts 02114, USA. Phone: 617.643.3331. Email: xavier@molbio.mgh.harvard.edu.

JMP's present address is: Division of Gastroenterology and Hepatology, Johns Hopkins Hospital, Baltimore, Maryland, USA.

1. Jostins L, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491(7422):119–124.
2. Noble CL, et al. Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. *Gut*. 2008;57(10):1398–1405.
3. Noble CL, et al. Characterization of intestinal gene expression profiles in Crohn's disease by genome-wide microarray analysis. *Inflamm Bowel Dis*. 2010;16(10):1717–1728.
4. Westra HJ, Franke L. From genome to function by studying eQTLs. *Biochim Biophys Acta*. 2014;1842(10):1896–1902.
5. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580–585.
6. Fu J, et al. Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. *PLoS Genet*. 2012;8(1):e1002431.
7. Grundberg E, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet*. 2012;44(10):1084–1089.
8. Fairfax BP, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet*. 2012;44(5):502–510.
9. McKenzie M, Henders AK, Caracella A, Wray NR, Powell JE. Overlap of expression quantitative trait loci (eQTL) in human brain and blood. *BMC Med Genomics*. 2014;7:31.
10. Fairfax BP, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*. 2014;343(6175):1246949.
11. Kabakchiev B, Silverberg MS. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. *Gastroenterology*. 2013;144(7):1488–1496, 1496.e1–3
12. Singh T, Levine AP, Smith PJ, Smith AM, Segal AW, Barrett JC. Characterization of expression quantitative trait loci in the human colon. *Inflamm Bowel Dis*. 2015;21(2):251–256.
13. Hulur I, et al. Enrichment of inflammatory bowel disease and colorectal cancer risk variants in colon expression quantitative trait loci. *BMC Genomics*. 2015;16:138.
14. Geiss GK, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol*. 2008;26(3):317–325.
15. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet*. 2013;14(9):661–673.
16. Arijs I, et al. Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment. *PLoS One*. 2009;4(11):e7984.
17. Mowat AM, Agace WW. Regional specialization within the intestinal immune system. *Nat Rev Immunol*. 2014;14(10):667–685.
18. Mann ER, et al. Compartment-specific immunity in the human gut: properties and functions of dendritic cells in the colon versus the ileum. *Gut*. 2016;65(2):256–270.
19. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol*. 2007;1:54.
20. van der Pouw Kraan TC, Zwiers A, Mulder CJ, Kraal G, Bouma G. Acute experimental colitis and human chronic inflam-

matory diseases share expression of inflammation-related genes with conserved Ets2 binding sites. *Inflamm Bowel Dis.* 2009;15(2):224–235.

21. Zucchelli C, et al. Structure of human Sp140 PHD finger: an atypical fold interacting with Pin1. *FEBS J.* 2014;281(1):216–231.

22. Madani N, et al. Implication of the lymphocyte-specific nuclear body protein Sp140 in an innate response to human immunodeficiency virus type 1. *J Virol.* 2002;76(21):11133–11138.

23. Quintana FJ, et al. Aiolos promotes TH17 differentiation by directly silencing Il2 expression. *Nat Immunol.* 2012;13(8):770–777.

24. Melé M, et al. Human genomics. The human transcriptome across tissues and individuals. *Science.* 2015;348(6235):660–665.

25. Heng TS, Painter MW, Immunological Genome Project Consortium. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol.* 2008;9(10):1091–1094.

26. Billot K, et al. Differential aiolos expression in human hematopoietic subpopulations. *Leuk Res.* 2010;34(3):289–293.

27. Farh KK, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature.* 2015;518(7539):337–343.

28. Matesanz F, et al. A functional variant that affects exon-skipping and protein expression of SP140 as genetic mechanism predisposing to multiple sclerosis. *Hum Mol Genet.* 2015;24(19):5619–5627.

29. Tesson BM, Breitling R, Jansen RC. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinformatics.* 2010;11:497.

30. Wu S, et al. Vitamin D receptor pathway is required for probiotic protection in colitis. *Am J Physiol Gastrointest Liver Physiol.* 2015;309(5):G341–G349.

31. Sun J. VDR/vitamin D receptor regulates autophagic activity through ATG16L1. *Autophagy.* 2016;12(6):1057–1058.

32. Jin D, et al. Lack of vitamin D receptor causes dysbiosis and changes the functions of the murine intestinal microbiome. *Clin Ther.* 2015;37(5):996–1009.e7.

33. Heimerl S, et al. Alterations in intestinal fatty acid metabolism in inflammatory bowel disease. *Biochim Biophys Acta.* 2006;1762(3):341–350.

34. Wei X, et al. Fatty acid synthase modulates intestinal barrier function through palmitoylation of mucin 2. *Cell Host Microbe.* 2012;11(2):140–152.

35. Nakajima K, et al. The ARNT-STAT3 axis regulates the differentiation of intestinal intraepithelial TCRαβ+CD8αα+ cells. *Nat Commun.* 2013;4:2112.

36. Lamas B, et al. CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. *Nat Med.* 2016;22(6):598–605.

37. Lassen KG, et al. Genetic Coding Variant in GPR65 Alters Lysosomal pH and Links Lysosomal Dysfunction with Colitis Risk. *Immunity.* 2016;44(6):1392–1405.

38. Muthana M, et al. C5orf30 is a negative regulator of tissue damage in rheumatoid arthritis. *Proc Natl Acad Sci U S A.* 2015;112(37):11618–11623.

39. Davies E, Tsang CW, Ghazali AR, Harris RM, Waring RH. Effects of culture with TNF-alpha, TGF-beta and insulin on sulphotransferase (SULT 1A1 and 1A3) activity in human colon and neuronal cell lines. *Toxicol In Vitro.* 2004;18(6):749–754.

40. Fowler SA, et al. SMAD3 gene variant is a risk factor for recurrent surgery in patients with Crohn's disease. *J Crohns Colitis.* 2014;8(8):845–851.

41. Haberman Y, et al. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J Clin Invest.* 2014;124(8):3617–3633.

42. Meckel K, et al. Serum 25-hydroxyvitamin D concentration is inversely associated with mucosal inflammation in patients with ulcerative colitis. *Am J Clin Nutr.* 2016;104(1):113–120.

43. Wu S, et al. Intestinal epithelial vitamin D receptor deletion leads to defective autophagy in colitis. *Gut.* 2015;64(7):1082–1094.

44. Du J, et al. 1,25-Dihydroxyvitamin D protects intestinal epithelial barrier by regulating the myosin light chain kinase signaling pathway. *Inflamm Bowel Dis.* 2015;21(11):2495–2506.

45. Carvalho AY, et al. The role of Vitamin D level and related single nucleotide polymorphisms in Crohn's disease. *Nutrients.* 2013;5(10):3898–3909.

46. Yamamoto-Furusho JK, Mendivil EJ, Mendivil-Rangel EJ, Villeda-Ramírez MA, Fonseca-Camarillo G, Barreto-Zuñiga R. Gene expression of carnitine organic cation transporters 1 and 2 (OCTN) is downregulated in patients with ulcerative colitis. *Inflamm Bowel Dis.* 2011;17(10):2205–2206.

47. Fortin G, et al. L-carnitine, a diet component and organic cation transporter OCTN ligand, displays immunosuppressive properties and abrogates intestinal inflammation. *Clin Exp Immunol.* 2009;156(1):161–171.

48. Scioli MG, et al. Propionyl-L-carnitine is efficacious in ulcerative colitis through its action on the immune function and microvasculature. *Clin Transl Gastroenterol.* 2014;5:e55.

49. Gaublomme JT, et al. Single-cell genomics unveils critical regulators of Th17 cell pathogenicity. *Cell.* 2015;163(6):1400–1412.

50. Schlitzer A, et al. IRF4 transcription factor-dependent CD11b+ dendritic cells in human and mouse control mucosal IL-17 cytokine responses. *Immunity.* 2013;38(5):970–983.

51. O'Connor W, Zenewicz LA, Flavell RA. The dual nature of T(H)17 cells: shifting the focus to function. *Nat Immunol.* 2010;11(6):471–476.

52. Maxwell JR, et al. Differential roles for interleukin-23 and interleukin-17 in intestinal immunoregulation. *Immunity.* 2015;43(4):739–750.

53. Watanabe T, et al. NOD2 downregulates colonic inflammation by IRF4-mediated inhibition of K63-linked polyubiquitination of RICK and TRAF6. *Mucosal Immunol.* 2014;7(6):1312–1325.

54. Piro RM, et al. An atlas of tissue-specific conserved coexpression for functional annotation and disease gene prediction. *Eur J Hum Genet.* 2011;19(11):1173–1180.

55. Pierson E, et al. Sharing and specificity of co-expression networks across 35 human tissues. *PLoS Comput Biol.* 2015;11(5):e1004220.

56. Gaiteri C, Ding Y, French B, Tseng GC, Sibille E. Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes Brain Behav.* 2014;13(1):13–24.

57. Mayburd AL. Expression variation: its relevance to emergence of chronic disease and to therapy. *PLoS One.* 2009;4(6):e5921.

58. Mar JC, et al. Variance of gene expression identifies altered network constraints in neurological disease. *PLoS Genet*. 2011;7(8):e1002207.

59. Ning K, et al. Improved integrative framework combining association data with gene expression features to prioritize Crohn's disease genes. *Hum Mol Genet*. 2015;24(14):4147–4157.

60. Huang J, et al. eQTL mapping identifies insertion- and deletion-specific eQTLs in multiple tissues. *Nat Commun*. 2015;6:6821.

61. Li G, et al. Human genetics in rheumatoid arthritis guides a high-throughput drug screen of the CD40 signaling pathway. *PLoS Genet*. 2013;9(5):e1003487.

62. Oparina NY, et al. PXK locus in systemic lupus erythematosus: fine mapping and functional analysis reveals novel susceptibility gene ABHD6. *Ann Rheum Dis*. 2015;74(3):e14.

63. Raj T, et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science*. 2014;344(6183):519–523.

64. Westra HJ, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013;45(10):1238–1243.

65. Peters JE, et al. Insight into Genotype-Phenotype Associations through eQTL Mapping in Multiple Cell Types in Health and Immune-Mediated Disease. *PLoS Genet*. 2016;12(3):e1005908.

66. Andiappan AK, et al. Genome-wide analysis of the genetic regulation of gene expression in human neutrophils. *Nat Commun*. 2015;6:7971.

67. Westra HJ, et al. Cell Specific eQTL Analysis without Sorting Cells. *PLoS Genet*. 2015;11(5):e1005223.

68. Lee MN, et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science*. 2014;343(6175):1246980.

69. Welter D, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42(Database issue):D1001–D1006.