

## SUPPLEMENTARY METHODS

### Washington University study subjects and histological selection of FFPE samples

We retrospectively obtained de-identified FFPE tissue samples and demographic data from subjects who underwent ileo-colic resection surgery at Barnes-Jewish Hospital, St. Louis between 2006 and 2010. CD resections were due to disease-associated complications. Control resections were due to adenoma, carcinoma or abdominal trauma. The set of H&E-stained tissue sections from each resection were examined by two pathologists (T.C.L. and T.S.S.) to identify the FFPE tissue block that contained the uninvolving proximal margin (terminal ileum) of each resection and then this sample was further evaluated for the following histological inclusion criteria: the tissue section had 1) at least 100 well-oriented crypts, as defined by a lumen that extended from the apex to the base of the crypt, and 2) absent or minimal active or chronic inflammatory disease (cryptitis, crypt abscess, architectural distortion, pyloric gland metaplasia, neutrophil infiltrate, etc.). Subjects were recruited by the Washington University Digestive Diseases Research Core Center (DDRCC). Written informed consent was obtained from all study participants prior to inclusion in the study.

### FFPE molecular profiling of Washington University tissue specimens

Two unstained, non-baked 5-μm FFPE tissue sections on glass microscope slides were obtained and used for RNA extraction and DNase treatment with the RNeasy FFPE kit (Qiagen) according to the manufacturer's instructions. RNA concentration and quality were determined by NanoDrop and by Bioanalyzer 2100 (Agilent). The Genome Technology Access Center (GTAC) at Washington University performed microarrays as previously described<sup>1</sup> and performed RNA-seq library preparation, sequencing and read alignment. Library preparation used 150 ng of total RNA. Double-stranded cDNA was generated using the Ovation RNA-seq FFPE system (Nugen) per manufacturer's instructions. The cDNA was blunt ended, had an A base added to the 3' ends, and then had Illumina sequencing adapters ligated to the ends. Ligated fragments were then

amplified for 12 cycles using primers incorporating unique index tags. Sequencing was performed on an Illumina HiSeq2000 SR42 using single reads extending 42 bases. RNA-seq reads were aligned to the Ensembl release 76 assembly with STAR version 2.0.4b. Gene counts were derived from the number of uniquely aligned unambiguous reads by Subread:featureCount version 1.4.5. Transcript counts were produced by Sailfish version 0.6.3. Sequencing performance was assessed for total number of aligned reads, total number of uniquely aligned reads, genes and transcripts detected, ribosomal fraction known junction saturation and read distribution over known gene models with RSeQC version 2.3. A quality control workflow (outlined in Figure S1B-D) was used to select FFPE samples with similar levels of RNA degradation based on criteria used in previous reports<sup>2-4</sup>.

### Partitional clustering

The mapped gene-level transcript matrix file was filtered to exclude transcripts with a mean of 0 cpm and those that were “non-protein coding and non-annotated”. An unpaired two-tailed Mann-Whitney t test was used to identify transcripts with an unadjusted  $P < .05$  between CD and Non-IBD to use as the input for partitional clustering. Partitional clustering was completed with Partek® Genomics Suite® software, version 6.6 Copyright © 2017 (Partek Inc., St. Louis, MO, USA) using the fuzzy c-means clustering method and a fuzzifier of 2. Termination criteria were set to 1000 max iterations and an epsilon of 0.01. The distance function was Spearman Dissimilarity and the number of clusters was set to 9, which was determined to be the largest number of clusters possible with a minimal increase in the Davies-Bouldin index.

### Functional enrichment analyses

Functional enrichment analysis of the gene clusters used Enrichr (<http://amp.pharm.mssm.edu/Enrichr/>)<sup>5, 6</sup>. A subset of Cluster 5 genes was selected for more focused analysis if the gene was in one of the following lists: 1) top 100 most altered in CD, 2)

genes representing the 6 network nodes enriched for Cluster, or 3) identified in the brush border proteomics studies<sup>7, 8</sup>. Protein expression patterns in small intestine were determined using the Human Protein Atlas (HPA) v16.1<sup>9, 10</sup> and supported by further evidence in the literature when available. Brush border proteomics comparisons were based on Table S1 in McConnell *et al.* and Table 1 in Yoshida *et al.*<sup>7, 8</sup>.

### Histological validation

Brightfield images of H&E-stained resection margins were acquired with an Olympus BX51 microscope equipped with UPlanFL 10X/0.30, 20X/0.50, 40X/0.75, and 100X/1.30 Oil Iris objective lenses, an Olympus DP70 camera and DP Controller software or an Olympus DP22 camera and cellSens Standard v1.17 software. Confocal images were acquired with a Zeiss LSM880 laser scanning confocal microscope (Carl Zeiss Inc, Thornwood, NY) equipped with a 63X, 1.4 numerical aperture Zeiss Plan Apochromat oil objective. The 405nm diode, 488nm Argon, and 543nm HeNe1 (helium neon) lasers were utilized with an optimal pinhole of 1 airy unit to acquire images with ZEN 2.1 black edition software. Adobe Photoshop CS6 was used to adjust brightness and contrast, sharpen and crop images. Image acquisition and quantification was performed in a blinded fashion for all histological metrics.

To quantify microvilli length and cell height, 5 enterocytes measured per villus on 10 villi per sample, for a total of 50 cells measured per sample. Experiments supporting the rigor of this method are described in the legends of Figures S4, S5, and S6. Measured cells were visualized with the 100X objective lens and were located in the top one-third of a villus (where enterocytes have a fully mature brush border) in regions with well-oriented epithelial cells (i.e. a single layer of basally located nuclei and/or goblet cells with theca open to the lumen). For Observer 1, ten images (4080x3072; one image/villi) were acquired and microvilli length or cell height was measured using ImageJ software<sup>11</sup>. For Observer 2, microvilli length was directly measured from

the live video feed from the DP22 camera using cellSens software. The latter system was used to determine average microvilli length in the UNITI-2 cohort samples described below.

To quantify goblet cell number, two images (4080x3072) were acquired with the 20X objective lens from regions of high-quality, well-oriented tissue (range: 9-23 villi per sample; range: 184-1,769 goblet cells counted per sample). H&E-stained slides were no longer available from 8 of the samples used for transcriptional analysis.

For immunostaining, FFPE tissue sections underwent deparaffinization, antigen retrieval with Trilogy (Cell Marque), blocking with 3% BSA and 1% Triton X-100 in PBS, followed by immunostaining with lysozyme antibody (clone C-19; Santa Cruz Biotechnology sc-27958; 1:100 dilution) and detection with donkey anti-goat Alexa 594 antibody (Thermo Fisher Scientific; 1:500 dilution) or immunostaining with VIL1 rabbit monoclonal antibody (clone SP145; Abcam ab130751; 1:100 dilution) and detection with donkey anti-goat Alexa 546 (Thermo Fisher Scientific; 1:500 dilution). Donkey serum (10% vol/vol) was included in the blocking step for VIL1 immunostaining. Nuclei were visualized with bis-benzamide. (Hoechst 33258: Invitrogen) and coverslips were mounted with Fluoromount (Sigma) or ProLong Diamond (Thermo Fisher Scientific). To quantify Paneth cells per crypt, the lysozyme-positive cells were counted in a minimum of 50 crypts per sample. To quantify VIL1 staining intensity, an observer blinded to the identity of the samples binned the samples into the “Normal” or “Reduced” staining intensity categories.

### **Statistical validation of microvilli measurement method**

#### *Sampling: Number of microvilli measurements per villus*

We determined the number of enterocyte cell microvilli length measurements per villus required to reliably measure the average microvilli length. For this, we generated a data set in which microvilli length was measured on every enterocyte in the upper two-thirds of 2 villi from each of 4 ileal resection tissue sections (representing 4 distinct subjects, 48-120 enterocytes

measured per villus). We then used two statistical permutational approaches to analyze the required size of “virtual samples” of these populations to accurately determine population average. First, the random number generator function in Excel was used to randomly shuffle the order of measurements from a single villus. Using the permuted order, consecutive measurements were grouped into virtual samples of size  $n$  ( $n = 3, 5, 10, 20$ , or  $30$  cells) to form the maximal number of groups that could be formed from each villus data set using each sample size. The bioequivalence between the virtual samples and the full data sets was assessed by 2-way ANOVA followed by post hoc comparisons for the number of cells counted per sample. Second, we performed 1,000 permutations of an unpaired, two-tailed parametric t test comparing a random selection of  $n$  cell measurements to the total population of measurements collected from each villus. We measured the percentage of permuted tests that resulted in  $P \leq 0.05$  (performed in R version 3.4 using base function *sample*; <https://www.R-project.org/>). Under the assumption that a random sub-sample of cells should not be statistically different than the population mean, the permutation analysis indicates an expected false discovery rate (FDR) obtained from the selected number of cell measurements. The data from these analyses is presented in Figure S3 and indicated that measurement of 5 cells is required to accurately determine average microvilli length of a villus.

#### *Sampling: Number of enterocyte microvilli measurements per villus*

We next determined the number of villi required to reliably measure the average microvilli length in a tissue sample. For this, microvilli length was measured on 5 enterocytes in the upper one-third of a villus (this number of cells was determined in the experiment presented in Figure S3) for 50 villi in each of 4 ileal resection tissue sections (representing 4 distinct subjects). The average microvilli length of each villi was calculated from the 5 cell measurements per villi. We then used the two statistical permutational approaches described above to analyze the required size of “virtual samples” of these populations to accurately determine population average, with

the exception that average microvilli length values were used to generate the virtual samples of size  $n$  ( $n = 3, 5, 10, 20, 30$  or  $40$  villi) compared to the whole population of  $50$  villi in each sample. The data from these analyses is presented in Figure S4 and indicated that measurement of  $5$  cells per villi and  $10$  villi per sample accurately determines the average microvilli length of a sample. We statistically determined that measuring microvilli length of  $50$  cells per sample ( $5$  cells on each of  $10$  villi) gave a  $2\%$  Type I error (i.e., less than  $0.05$ ), thereby validating our ability to determine a significant difference between CD and Non-IBD control average microvilli lengths.

#### *Inter-observer reproducibility*

We examined the inter-observer reproducibility obtained using the above validated sampling method for determination of average microvilli length per sample. Both observers used the same set of H&E-stained ileal resection histological tissue sections and were blinded to the diagnosis group (Non-IBD or CD) of samples. Although the same histological tissue sections were used by each observer, different regions were likely selected for measurement by each. The observers also used different camera and software systems to collect their data. Observer 1 captured micrographs of  $10$  villi per sample using an Olympus DP70 camera with DP Controller software and then used Image J software to measure microvilli length. Observer 2 measured microvilli length directly from a live camera video feed during microscopy using an Olympus DP70 camera and Olympus cellSens software. The data from Observer 1 ( $n=60$ ) is presented in Figure 3 and Figure S5; the data from Observer 2 ( $n=57$ ; 3 H&E-stained histological tissue sections were no longer available) is presented in Figure S5. The data generated by Observer 1 and Observer 2 were strongly correlated and had similar distributions. Moreover, Observer 2 recapitulated the reduction in average microvilli length in CD samples relative to Non-IBD samples.

#### **Electron Microscopy**

Tissue was fixed in a modified Karnovsky's fixative of 3% glutaraldehyde, 1% paraformaldehyde in 0.1M sodium cacodylate buffer and then post-fixed in 2% osmium tetroxide in 0.1 M sodium cacodylate buffer for 1 hr, en bloc stained with 3% aqueous uranyl acetate for 30 min, dehydrated in graded ethanols and embedded in PolyBed 812 catalog #08792-1 (Polysciences, Hatfield, PA). Tissue blocks were sectioned at ninety nanometers thick, post stained with Venable's lead citrate and viewed with a JEOL model 1400EX electron microscope (JEOL, Tokyo, Japan). Digital images were acquired using the AMT Advantage HR (Advanced Microscopy Technology, Danvers MA) high definition CCD, 11 megapixel TEM camera. Image J (NIH) was used to measure the indicated microvillar ultrastructural features using 30,000X images (8-30 images per sample) from Non-IBD (n=3) and CD (n=4) ileal tissue samples. All metrics were assessed in a blinded fashion.

### **Generation of the co-expression module network**

#### *Overview of the correlation graphlet analysis approach*

A network of overlapping co-expression modules was constructed using an approach that we have termed correlation graphlet analysis (CGA). CGA comprises three steps, described in further detail below: data preprocessing, construction of a master correlation network, and construction of the final network of overlapping co-expression modules. The computer code for CGA and sample data sets are included in Supplementary Materials.

In general, the overarching goal of gene network approaches is to produce biological hypotheses that lead to understanding of human disease, including identification of pathological mechanisms, therapeutic targets, and biomarkers. However, gene networks are frequently complex to the degree that it is difficult to visualize or intuit the critical biological relationships. In addition, genes may often act in several related, but distinct functional groups. This observation motivated the development of the CGA approach, which collapses closely related networked

genes into modules. These modules can be functionally informative, representing gene regulatory events occurring in a specific cell subset or in response to a particular stimulus.

The construction of gene modules is not novel as a concept or approach <sup>12</sup>, but gene modules are often treated as distinct and separate entities, which can be restrictive in the sense that the relationships between the modules are not clear. Therefore, CGA allows the modules to overlap and hence produces a module network where modules are linked through shared genes. Module overlap facilitates investigation of the functional links between modules and suggests potential multi-functional genes <sup>13</sup>. The complexity of module networks is reduced compared to gene networks, resulting in simpler network organizations where relationships are easier to intuit. For example, the module network formed in this study contained 84 module nodes and 89 edges (Figure 4B) compared to the gene correlation network, which contained over 1,300 gene nodes and almost 40,000 edges (Figure 4A). Similar approaches based on construction of networks of overlapping groups of entities have been used to gain insights from large and complex datasets in many different contexts <sup>14</sup>.

CGA employs an empirical criterion based on global network topology to select the correlation threshold used to generate the correlation network. An alternative approach, which would use a cutoff based on statistical criteria, such as Fisher's transformation *p*-value <sup>15</sup> or more sophisticated estimations <sup>16</sup>, can produce highly complex networks with large numbers of nodes and edges that are difficult to interpret and contain many false positives due to assumptions of independence between genes. CGA determines the correlation threshold by characterizing the topology (structure) of the network over a range of values until the strictest threshold at which the topology is stable has been identified. Overall, the CGA approach provides a simpler network organization for visualization of the relationships between correlated sets of genes.

#### *Data preprocessing*

The full transcriptomic dataset was assessed for outliers, and the same two CD samples were removed as was done for the partitional clustering. The remaining 68 samples (38 CD and 32 Non-IBD controls) were aligned to human genome (hg38 with ENSEMBL annotations) using ArrayStudio® software (OmicSoft®, Cary, NC) software to produce read counts and Reads Per Kilobase of transcript per Million mapped reads (RPKM) values for each gene. To focus on genes with medium-to-high expression, we excluded genes that had RPKM >1 in less than 90% of the samples within each subpopulation (CD or control). This resulted in a count data matrix of 13,858 genes across 68 samples, which was used as input for the co-expression network algorithm.

#### *Construction of a master correlation network*

The master correlation network is a binary graph with nodes corresponding to genes and edges linking the genes with an absolute Spearman correlation greater than an empirically determined threshold, similar to previously proposed correlation-based networks<sup>17</sup>. A lower bound for the threshold is identified by the 99.95<sup>th</sup> percentile of 200,000 bootstrap samples taken from all absolute Spearman correlations between genes. However, using the lower bound can lead to networks with an excessive number of edges, which may be difficult to analyze. Therefore, in contrast to other network methods, CGA does not use a pre-defined threshold, but instead selects an optimal threshold corresponding to the largest absolute correlation value that produces a stable network topology.

The change in overall network topology induced by different thresholds is measured using graphlet correlation matrices (GCMs) and graphlet correlation distance (GCD), as proposed by Yaveroğlu et al.<sup>18</sup>. Graphlet computations were performed using Orca<sup>19</sup>. The optimal threshold is chosen from a sequence of candidate cutoffs decreasing from 1.0 to the lower bound, with step size 0.005. Generally, it is expected that the GCD between consecutive networks will decrease towards zero with some oscillations as the absolute correlation decreases, indicating a stabilizing topology. In such cases where the GCD decreases monotonically with very small oscillations, the

threshold can be selected based on the GCD, such that the selected threshold is the largest absolute correlation at which all subsequent GCDs between consecutive GCMs are smaller than some topological threshold (a typical value would be 20). Once the correlation threshold is chosen and applied, only genes with at least one edge are retained in the master correlation network. The resulting network will often contain several disconnected subnetworks corresponding to subsets of genes that were unrelated at the selected threshold.

GCMs from the preprocessed transcriptomic data for the CD and Non-IBD samples were generated using various threshold values (Figure S11A-F), and the stability of the network topology was quantified using the GCD between consecutive GCMs (Figure S11G). This data set showed a non-typical pattern, where the GCD decreased and then increased, potentially indicating a shift to a different topology across the tested thresholds. When this occurs, the optimal threshold is selected as the absolute correlation value corresponding to the minimum of the GCD graph, which was determined to be 0.75 for this data set.

#### *Construction of overlapping modules*

Following the construction of the master binary network, an iterative procedure is applied to collapse groups of co-expressed genes into gene modules, which are then constructed into a simplified network. This process employs two basic steps, which are described in more detail in the subsequent paragraphs. The overview of the process is as follows. First, a core set of genes is identified and used to form module centers. Second, modules are “grown” from the module centers to construct the network. To illustrate this process, a toy data set was used to form a simple binary network composed of 16 nodes and 21 edges (Figure S12).

In simple terms, gene nodes with the most non-overlapping first-degree neighbors are selected to become module centers. In more detail, if the “K-neighborhood” of a gene is the set of genes within “K” steps from the gene along the network, then genes are prioritized for module centers based on the size of their 1-neighborhoods. Only genes with a 1-neighborhood size

greater than a user-specified cutoff (5 by default) are considered for module centers. Starting from the gene node with greatest 1-neighborhood size (Figure S12B), genes are iteratively added to the module defined by a particular module center if their 1-neighborhoods do not significantly overlap with the 1-neighborhoods defined by another module center (Figure S12C, D). Overlap is determined by a one-sided Hypergeometric test, with  $P < .05$  considered to indicate significant overlap.

Once the set of module centers has been identified, each gene in the network is grouped with the nearest center to construct the modules. If a gene is equidistant from several centers, the gene is included in each of these centers (Figure S12E). Subnetworks containing genes with a 1-neighborhood size less than the selected cutoff are collapsed into single modules that are disconnected from the primary network (for example, Module 033 containing the Paneth cell genes shown in Figure S10F). The final module network is constructed with nodes representing gene modules and edges representing overlapping (shared) genes between the modules (Figure S12F). For effective visualization of module networks, node size is determined by the number of genes contained in the module and edge thickness is determined by the number of shared genes between two modules. Module networks can be visualized within the Cytoscape<sup>20</sup> framework using a force-directed layout.

## **UNITI-2 study subjects**

The phase 3 UNITI-2 study evaluated the safety and efficacy of intravenous Ustekinumab (UST) induction in patients with moderately-to-severely active CD who had failed conventional therapies (e.g. AZA, MTX, or 6-MP)<sup>21</sup>. Microarray and histology data from the UNITI-2 study were used to assess microvilli length and gene set enrichment of the core Cluster 5 gene signature in an independent CD cohort. In addition, the impact of UST on these phenotypes was assessed. UST induction responders (defined as a Crohn's disease activity index (CDAI) decrease  $\geq 100$ ) entered IMUNITI to evaluate subcutaneous (SC) UST maintenance therapy (two dosing

schedules were assessed: UST 90 mg q12w or UST 90 mg q8w). In this phase, UST induction non-responders received UST 90 mg SC q8w if in clinical response 8 weeks after the first SC dose; placebo (PBO) induction non-responders received UST 130 mg SC q12w if in clinical response 8 weeks after the first SC dose; and PBO induction responders received PBO. Two sets of intestinal biopsies were collected during endoscopy at the induction baseline (I-Wk0), Week 8 (I-Wk8), and maintenance Week 44 (M-Wk44) from patients who consented to participate in the endoscopy sub-study. One set of biopsies was used for microarray gene expression analysis and the other set was used to prepare paraffin blocks and H&E slides for histological assessment. In this study, only ileal biopsies from UNITI-2 patients with ileal or ileo-colonic CD involvement (n=111) were used in this analysis. Ileal biopsies from healthy subjects (n=26) were analyzed as a control group for the gene expression profiling. The average microvilli length was measured in ~30% of the samples (n=63), those that contained 40 or more crypt-villus units and had adequate staining.

### **Enrichment analysis of the core Cluster 5 gene set in UST phase 3 studies**

Gene Set Variation Analysis (GSVA)<sup>22</sup> was conducted to compute enrichment scores of the core Cluster 5 gene set for each biopsy sample in the UST phase 3 microarray dataset. The Spearman correlation was assessed between the enrichment scores and Simple Endoscopic Score for Crohn's Disease (SES-CD). Differences in enrichment scores were analyzed between phenotypes (i.e. CD vs. healthy, post-treatment vs. baseline) using parametric tests in ArrayStudio version 10 (OmicSoft Corp., St. Morrisville, NC). Enrichment differences with  $P < .05$  were considered statistically significant.

### **Statistics**

Graphpad Prism v7 was used to perform statistical analysis with  $P < .05$  considered to be significant, unless otherwise indicated. Demographic variable comparisons used Fisher's exact

test or two-tailed unpaired t test. A Mann Whitney unpaired t test was used to compare CD vs. Non-IBD samples when the groups had unequal variances and a two tailed unpaired parametric t test was used when the groups had equal variances, as indicated in the figure legends. Linear regression was used to assess correlations in X-Y graphs. To determine if expression of any genes correlated with age, a 1-way regression model was used in Partek Genomics Suite v6.6 with a false discovery rate-adjusted  $P < .05$  and a 1.5-fold change between CD and Non-IBD considered to be significant.

## REFERENCES CITED IN SUPPLEMENTARY METHODS

1. VanDussen KL, Liu TC, Li D, et al. Genetic variants synthesize to produce paneth cell phenotypes that define subtypes of Crohn's disease. *Gastroenterology* 2014;146:200-9.
2. Opitz L, Salinas-Riester G, Grade M, et al. Impact of RNA degradation on gene expression profiling. *BMC Med Genomics* 2010;3:36.
3. Rentoft M, Coates PJ, Laurell G, et al. Transcriptional profiling of formalin fixed paraffin embedded tissue: pitfalls and recommendations for identifying biologically relevant changes. *PLoS One* 2012;7:e35276.
4. Romero IG, Pai AA, Tung J, et al. RNA-seq: impact of RNA degradation on transcript quantification. *Bmc Biology* 2014;12.
5. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 2013;14:128.
6. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;44:W90-7.
7. McConnell RE, Benesh AE, Mao S, et al. Proteomic analysis of the enterocyte brush border. *Am J Physiol Gastrointest Liver Physiol* 2011;300:G914-26.
8. Yoshida S, Fukutomi T, Kimura T, et al. Comprehensive proteome analysis of brush border membrane fraction of ileum of ezrin knockdown mice. *Biomed Res* 2016;37:127-39.
9. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. *Science* 2015;347:1260419.
10. Uhlen M, Oksvold P, Fagerberg L, et al. Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* 2010;28:1248-50.
11. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012;9:671-5.
12. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005;4:Article17.
13. Gillis J, Pavlidis P. The impact of multifunctional genes on "guilt by association" analysis. *PLoS One* 2011;6:e17258.
14. Lum PY, Singh G, Lehman A, et al. Extracting insights from the shape of complex data using topology. *Sci Rep* 2013;3:1236.
15. Hero AO, Rajaratnam B. Foundational Principles for Large-Scale Inference: Illustrations Through Correlation Mining. *Proc IEEE Inst Electr Electron Eng* 2016;104:93-110.

16. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 1914;10:507-521.
17. Lee HK, Hsu AK, Sajdak J, et al. Coexpression analysis of human genes across many microarray data sets. *Genome Res* 2004;14:1085-94.
18. Yaveroglu ON, Malod-Dognin N, Davis D, et al. Revealing the hidden language of complex networks. *Sci Rep* 2014;4:4547.
19. Hocevar T, Demsar J. A combinatorial approach to graphlet counting. *Bioinformatics* 2014;30:559-65.
20. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498-504.
21. Feagan BG, Sandborn WJ, Gasink C, et al. Ustekinumab as Induction and Maintenance Therapy for Crohn's Disease. *N Engl J Med* 2016;375:1946-1960.
22. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;14:7.

**LIST OF SUPPLEMENTARY TABLES**

**Table S1.** Demographic data for the 32 Non-IBD and 36 CD subjects included in the transcriptional analyses.

**Table S2.** Genes in partitional clusters.

**Table S3.** Number of overlapping genes in partitional clusters.

**Table S4.** Full functional enrichment results of partitional cluster genes.

**Table S5.** Validation of enterocyte gene enrichment of Cluster 5.

**Table S6.** Genes in co-expression network modules.

**Table S7.** Characteristics of the full UNITI-2 cohort and the subsets analyzed for gene expression and histology in this study.

**Table S8.** Core Cluster 5 genes used in GSVA.

**SUPPLEMENTARY FIGURE LEGENDS**

**Figure S1.** Quality control workflow to determine the FFPE samples included in RNA-seq transcriptional analysis. (A) Schematic illustrating use of FFPE tissue samples for sample selection, RNA extraction, and gene expression validation experiments. (B) Schematic of quality control (QC) workflow used to exclude low quality samples from the transcriptional analysis. The RNA yield of these samples ranged from 0.65 to 5.26 µg (median: 2.14 µg) with an A260/A280 ratio of 1.65 to 2.45 (median: 2.02). Using an Agilent Bioanalyzer 2100, we observed RIN values ranging from 1.2 to 7.8 (median: 2.4). To ensure similar levels of degradation, samples were selected for study only if greater than 20% of the RNA fragments were 200 base pairs or greater in length, as determined with the Bioanalyzer software. (C) Representative image of an Agilent Bioanalyzer plot from a FFPE sample included in the study. (D) Principal components analysis (PCA) of the gene expression data prior to exclusion of outlier samples based on gene expression data distribution criteria. The matrix file containing counts per million (cpm) values for mapped gene-level transcripts was imported into Partek Genomics Suite v6.6 and filtered to exclude

transcripts with mean of 0 cpm prior to PCA. Ellipses are drawn at 4 standard deviations away from the centroids of Non-IBD (blue; n=32) and CD (red; n=38) samples. Samples outside the ellipses were excluded from further transcriptional analysis (2 CD samples marked with asterisks), leaving 36 CD samples for the final analysis cohort. (E) Flowchart showing the numbers and percentages of CD and Non-IBD samples that passed histological screening and quality control (QC) metrics.

**Figure S2. Validation of microvillar gene expression in Cluster 5.** (A) Venn diagram of Cluster 5 genes (1051 unique gene symbols) and the genes encoding the proteins identified by two proteomics studies of the enterocyte brush border, McConnell *et al.* (524 unique gene symbols identified) and Yoshida *et al.* (107 unique gene symbols identified). (B) Graphs showing microarray gene expression levels for the indicated microvillar component genes. Data are expressed as relative intensity units. The same data is displayed side-by-side as scatterplots and box-and-whiskers plot for Non-IBD (n=32; blue) and CD (n=36; red). \*P<.05, \*\*P<.01, \*\*\*P<.001 by two-tailed Mann-Whitney t-test. The exact P-value is indicated for non-significant comparisons. Notably, Cluster 5 contained all of these microvilli component genes, with the exception of MYO6, which had not been included in the clustering analysis.

**Figure S3. Measurement of microvilli length on 5 enterocytes per villus accurately determines average microvilli length per villus.** Virtual samples representing 3, 5, 10, 20, or 30 cells were formed from the total population of microvilli length measurements collected from each villus (range: 48-120 enterocytes measured per sample) as described in the Supplementary Methods. (A) Graph of average microvilli lengths (displayed as mean  $\pm$  s.e.m.) calculated from the indicated virtual samples of 3, 5, 10, 20, or 30 cells or the total number of cells measured per villi. The bioequivalence between the virtual samples and the full data sets was assessed by 2-way ANOVA followed by post hoc comparisons for the number of cells counted per sample. No

statistical differences were observed between any of the virtual sample averages and the actual observed microvilli length average for any villus in any of the samples (n.s., not significant;  $P>0.05$ ). (B) Tabulation of the 2-way ANOVA results from (A). (C) Tabulation of the expected false discovery rate (FDR) percentages obtained with the permutation testing approach. The average of the FDR percentages from all villus samples for each cell selection size is listed in the bottom row. An average FDR <5.0% occurred when at least 5 cells per villus were sampled.

**Figure S4. Measurement of microvilli length on 5 enterocytes from 10 villi accurately determines average microvilli length per sample.** Virtual samples representing average microvilli length values from 3, 5, 10, 20, 30, or 40 villi were formed from the total population of 50 villi assessed per sample as described in the Supplementary Methods and with 5 cells measured per villi as determined by the experiments shown in Figure S3. (A) Graph of average microvilli lengths (displayed as mean  $\pm$  s.e.m.) calculated from the indicated virtual samples or the total population of 50 villi measured per sample. The bioequivalence between the virtual samples and the total populations was assessed by 2-way ANOVA followed by post hoc comparisons for the number of villi counted per sample. No statistical differences were observed between any of the virtual sample averages and the actual observed microvilli length average for any sample (n.s., not significant;  $P>0.05$ ). (B) Tabulation of the 2-way ANOVA results from (A). (C) Tabulation of the expected false discovery rate (FDR) percentages determined with the permutation testing approach. The average of the FDR percentages for each villi selection size is listed in the bottom row. An average FDR <5.0% occurred when at least 5 villi were sampled, but this dropped to <2.0% when at least 10 villi (with 5 cells measured per villus) were sampled.

**Figure S5. High inter-observer reproducibility for the determination of average microvilli length per sample.** Two, independent observers blinded to the diagnosis group of the samples (Non-IBD or CD) determined the average microvilli length for the same set of H&E-stained ileal

resection histological tissue sections, as described in the Supplementary Methods. (A) X-Y graph of average microvilli lengths calculated by Observer 1 vs. Observer 2 ( $n=57$ ). Statistical significance was determined by linear regression. (B) Box-and-whiskers plots showing the data distributions obtained by Observer 1 and Observer 2. No statistical difference between these data was observed by an unpaired, two-tailed parametric t test (n.s., non-significant). (C) Tabulation of the mean, standard deviation (s.d.) and sample size ( $n$ ) for the data generated by each Observer. (D) Graph of average microvilli length ( $\mu\text{m}$ ) determined by Observer 2 with the same data displayed side-by-side as scatterplots and as box-and-whisker plots for Non-IBD ( $n=26$ ; blue) and CD ( $n=31$ ; red) samples. \*\* $P = .0030$  between CD and Non-IBD by unpaired two-tailed parametric t test.

**Figure S6. Ultrastructural analysis of the enterocyte apical surface in uninvolved ileum tissue samples from CD and Non-IBD subjects.** (A-C) TEM images of enterocytes were captured from the upper portion of villi from Non-IBD ( $n=3$ ) and CD ( $n=4$ ) samples. (A) Additional representative images of the apical surface; each image represents a sample from a distinct subject. Yellow arrowheads designate rootlet ends and red arrows designate desmosomes. Bars, 1  $\mu\text{m}$ . (B) Graphs of average microvilli length, width, density and rootlet length quantified from TEM images and displayed as mean  $\pm$  s.e.m. An unpaired two-tailed parametric t test was used to determine the indicated  $P$ -values. These features were all similar between CD and Non-IBD samples. The lack of microvilli length difference between CD and Non-IBD is likely due to the much smaller sample size of the TEM study groups ( $n=3-4$  per group). An unpaired two-tailed parametric t test was used to determine the indicated  $P$ -values. (C) Graph of the source data for the average microvilli rootlet densities displayed in Figure 3D. Points represent the average measurement obtained from each TEM image for each of the samples (mean  $\pm$  s.e.m is indicated). Microvilli rootlet density was consistently reduced among the CD samples compared to the Non-IBD samples.

**Figure S7. Validation of the specificity of Cluster 5 association with microvilli length.** (A) X-Y graph of the average microvilli length and Cluster 2 or 9 average normalized gene expression. For Cluster 2, CD:  $*P = .0184$  and  $r^2 = .1617$  and Non-IBD: not significant,  $P = .1121$  and  $r^2 = .1018$  and for Cluster 9, CD:  $**P = .0053$  and  $r^2 = .2512$  and Non-IBD: not significant,  $P = .0521$  and  $r^2 = .1482$  by linear regression (lines shown on graphs). (B) Graph of average cell heights measured in Non-IBD ( $n=26$ ; blue) and CD ( $n=34$ ; red) samples. For each diagnosis, the same data are displayed side-by-side as scatterplots and as box-and-whisker plots. Not significant,  $P = .1027$  between CD and Non-IBD by unpaired two-tailed Mann-Whitney t-test. (C) X-Y graph of average cell height and Cluster 5 average normalized gene expression. CD: not significant,  $P = .1048$  and  $r^2 = .0801$  and Non-IBD: not significant,  $P = .4571$  and  $r^2 = .0233$  by linear regression (lines shown on graph).

**Figure S8. Enterocyte genes with expression levels that do not differ between CD and Non-IBD.** Graphs showing RNA-seq gene expression levels in counts per million (cpm) for the indicated genes, which have expression enriched in enterocytes. Data are displayed as scatterplots with the mean  $\pm$  s.e.m. indicated for Non-IBD ( $n=32$ ; blue) and CD ( $n=36$ ; red). Comparisons between Non-IBD and CD were not significant by two-tailed Mann-Whitney t-test. The  $P$ -value is indicated for non-significant comparisons.

**Figure S9. Goblet cell mRNA marker expression and numbers do not differ between CD and Non-IBD.** A proportional change in goblet cells relative to enterocytes in the CD vs. Non-IBD intestinal tissue samples could have been one explanation for the observed de-enrichment of the enterocyte-associated Cluster 5 genes in the CD samples. To investigate this possibility, we measured goblet cell number and mRNA marker expression in our samples.

First, we assessed the mRNA expression levels of the goblet cell markers mucin 2 (MUC2), trefoil factor 3 (TFF3), and resistin like beta (RETNLB) in our RNA-seq data set. We found no difference between the CD and Non-IBD samples. (A) Graphs showing RNA-seq gene expression levels in counts per million (cpm) for the indicated goblet cell marker genes. Data are displayed as scatterplots with the mean  $\pm$  s.e.m. indicated for Non-IBD (n=32; blue) and CD (n=36; red). Comparisons between Non-IBD and CD were not significant by two-tailed Mann-Whitney t test. The *P*-value is indicated for non-significant comparisons.

Second, we investigated goblet cell number in the same set of H&E-stained tissue sections used for histological selection at the beginning of the study. Similar to our findings with mRNA marker gene expression, we did not observe a difference in goblet cell number per villi between the CD and Non-IBD samples. (B) Representative image of H&E-stained ileal tissue section used for goblet cell quantification. A single well-oriented crypt-villus unit is shown with a dashed line separating the crypt from the villus (only villus goblet cells were counted) and arrowheads denote a few examples of goblet cells. Bar, 100  $\mu$ m. (C) Graph of average goblet cells per villus with the same data displayed side-by-side as scatterplots and as box-and-whisker plots for Non-IBD (n=26; blue) and CD (n=34; red) samples. Not significant, *P* = .7728 between CD and Non-IBD by unpaired two-tailed Mann-Whitney t test.

Third, we correlated the expression of the goblet cell mRNA marker genes to goblet cell number and to the average gene expression of each partitional cluster in our study. Goblet cell number per villi correlated with the gene expression of MUC2, but not TFF3, RETNLB or any of the partitional clusters (data for Cluster 5 is shown). (D) X-Y graphs of average goblet cells per villus and gene expression (cpm) of the indicated goblet cell marker genes for the samples in (C). Linear regression (lines shown on graphs) was used to determine the statistical significance of the associations (MUC2: \*\**P* = .0080 and  $r^2$  = .1153; TFF3: not significant, *P* = .4060 and  $r^2$  = .0119; RETNLB: not significant, *P* = .5662 and  $r^2$  = .0057). (E) X-Y graph of average goblet cells per villus and Cluster 5 average normalized gene expression for the samples in (C). CD: not significant, *P*

=.3609 and  $r^2$ =.0261 and Non-IBD: not significant,  $P$  =.5625 and  $r^2$ =.0142 by linear regression (lines shown on graph). Taken together, the evidence did not support a proportional change in goblet cells being linked or causal of the Cluster 5 gene expression changes in the CD samples.

**Figure S10. Paneth cell number and gene expression are increased and correlated in CD intestine.** A proportional change in Paneth cells relative to enterocytes in the CD vs. Non-IBD intestinal tissue samples could have been another explanation for the observed de-enrichment of the enterocyte-associated Cluster 5 genes in the CD samples. To investigate this possibility, we measured Paneth cell number and mRNA marker expression in our samples.

First, we assessed the mRNA expression levels of the Paneth cell markers Defensin alpha 5 (DEFA5), Defensin alpha 6 (DEFA6), Lysozyme (LYZ), phospholipase A2 group IIA (PLA2G2A), and regenerating family member 3 alpha (REG3A) in our RNA-seq data set. The expression of these markers was increased in CD relative to Non-IBD samples. (A) Graphs of RNA-seq gene expression displayed as counts per million (cpm) for the indicated Paneth cell marker genes. Data are displayed as scatterplots with the mean  $\pm$  s.e.m. indicated for Non-IBD ( $n=32$ ; blue) and CD ( $n=36$ ; red). \* $P$  <.05, \*\* $P$  <.01, \*\*\*\* $P$  <.0001 for Non-IBD vs. CD by two-tailed unpaired Mann Whitney t test.

Second, we investigated Paneth cell number in serial tissue sections to those used for RNA extraction. Similar to our findings with mRNA marker gene expression, Paneth cell number per crypt was increased in CD relative to Non-IBD samples. (B) Representative image of an ileal crypt base immunostained with lysozyme (red) to visualize Paneth cells. Nuclei were visualized with bis-benzamide (blue). Four well-oriented Paneth cells are outlined in white. Bar, 10  $\mu$ m. (C) Graph of average Paneth cells per crypt, with the same data displayed side-by-side as scatterplots and as box-and-whisker plots for Non-IBD ( $n=32$ ; blue) and CD ( $n=36$ ; red) samples. \*\* $P$  =.0012 between CD and Non-IBD by two-tailed unpaired t test.

Third, we correlated the expression of the Paneth cell mRNA marker genes to Paneth cell number per crypt and to the average gene expression of each partitional cluster in our study. Paneth cell number per crypt correlated with the gene expression of each of the Paneth cell marker genes. (D) X-Y graphs of average Paneth cells per crypt and counts per million (cpm) gene expression for the indicated Paneth cell markers in Non-IBD ( $n=32$ ) and CD ( $n=36$ ). DEFA5: \*\*\*\* $P<.0001$  and  $r^2=.3031$ ; DEFA6: \*\*\*\* $P <.0001$  and  $r^2=.2324$ ; LYZ: \*\*\*\* $P <.0001$  and  $r^2=.2165$ ; PLA2G2A: \*\*\*\* $P <.0001$  and  $r^2=.3811$ ; REG3A: \*\*\*\* $P <.0001$  and  $r^2=.2845$  by linear regression (lines shown on graphs). The Paneth cell marker genes belonged to Cluster 1 (Table S2), but Paneth cell number did not correlate with the average gene expression of Cluster 1 nor any other gene cluster (data for Cluster 5 is shown). (E) X-Y graph of average Paneth cells per crypt and Cluster 1 or Cluster 5 average normalized gene expression for Non-IBD ( $n=32$ ; blue) and CD ( $n=36$ ; red) samples. For Cluster 1, CD: not significant,  $P = .5578$  and  $r^2=.0102$  and Non-IBD: not significant,  $P = .8835$  and  $r^2=.0007$ ; for Cluster 5, CD: not significant,  $P = .6196$  and  $r^2=.0073$  and Non-IBD: not significant,  $P = .2341$  and  $r^2=.0469$  by linear regression (lines shown on graphs). Even so, given that a cell type associated with gene expression in Cluster 1 appeared to be changing in relationship to enterocyte gene expression and because the elevated pathways enriched in Cluster 1 could potentially impact enterocyte gene expression, we tested for a potential linkage between Cluster 1 and 5 gene expression. Although data from a few subjects demonstrated an inverse correlation between Cluster 1 and Cluster 5, this relationship was not consistent across the majority of the samples (see main text Figure 4C).

Fourth, we used our network analysis to examine whether the Paneth cell marker genes were linked to the Cluster 5-enriched region. (F) The Paneth cell gene markers did not map to the large, primary connected component of the gene-gene correlation network, but rather formed an isolated module present in the gene correlation network. This module (M033) is not displayed in the co-expression module network in Figure 4 due to its small number of associated genes, but this subnetwork is listed in Table S6 and the partitional cluster enrichment analysis for this module

in the network is presented in Figure S14. The isolation of the Paneth cell gene module supports the lack of correlation between Paneth cell mRNA marker expression or cell numbers with Cluster 1 or Cluster 5 average gene expression. Taken together, the evidence did not support the gene expression differences in Cluster 1 being linked or causal of those observed in Cluster 5.

**Figure S11. Empirical determination of a correlation threshold using graphlet correlation matrices.** Network topology was characterized using graphlet correlation matrices (GCMs) for gene correlation networks constructed with absolute correlation threshold cutoffs of (A) 0.92, (B) 0.88, (C) 0.84, (D) 0.80, (E) 0.76, and (F) 0.75. (G) The Frobenius norm between consecutive GCMs for cutoffs from 0.92 to 0.74 decreases towards zero, with a minimum at 0.75.

**Figure S12. Module network construction.** (A-F) A toy data set was used to demonstrate how module networks are constructed. (A) A master gene correlation network containing 16 nodes representing genes and 21 edges representing correlated expression between genes. (B) The node with the greatest 1-neighborhood size (green with bold outline) is added first to the set of module centers. The green nodes without bold outline are the 1-neighbors of this module center. (C) The node with second greatest 1-neighborhood size (blue with bold outline) is next added to the set of module centers because its 1-neighborhood (blue nodes without bold outline) does not overlap with the neighborhood of the first module center. (D) The node with third greatest 1-neighborhood size (bold red outline) is not included in the set of module centers because its 1-neighborhood (nodes with non-bold red outline) significantly overlaps with the 1-neighborhood of the previously defined module centers. (E) Remaining nodes are then assigned to the nearest module centers to complete module formation. The green node with the blue outline is equidistant from two module centers and represents an overlapping gene. (F) To construct the module network, the gene nodes are collapsed such that nodes now represent modules containing the

indicated number of genes and edges represent shared or overlapping genes between two modules.

**Figure S13. Source data for determining the enrichment of partitional cluster genes in the co-expression network modules.** (A) The same co-expression module network as presented in Figure 6B, but with the module number listed beside each module. (B) Table of results for the partition cluster enrichment analysis for each of the modules in the network in (A) as well as the Paneth cell subnetwork (Module 033) shown in Figure S11F. The module numbers are colored to match the enrichment legend in (A). The number of overlapping genes between each module and cluster are shown in the table and these are shaded according to the statistical significance of the overlap, as determined by one-sided hypergeometric test.

**Figure S14. Impact of UST induction on core Cluster 5 gene expression in CDAI-defined responders and non-responders.** Graphs of the enrichment scores for the core Cluster 5 gene set of induction phase, CDAI-defined non-responders (NR) and responders (R) who received either placebo (PBO) or ustekinumab (UST) and of healthy Non-IBD controls at the I-Wk8 time point vs. baseline I-Wk0. (B) Graphs of the enrichment scores for the core Cluster 5 gene set of induction phase NR and R groups and of healthy Non-IBD controls at the I-Wk0 time point. (C, D) Graphs of the CDAI scores (C) and SES-CD scores (D) of induction phase NR vs. R at the I-Wk0 time point. (A-D) An unpaired two-tailed parametric t test was used to determine statistical significance. Sample size was insufficient to perform a similar analysis of the NR and R groups at the M-Wk44 time point.

**Table S1. Demographic data for the 32 Non-IBD and 36 CD patients included in the transcriptional analyses.**

Characteristic	Non-IBD n (%)	CD n (%)	P
Male sex	11 (34.4)	16 (44.4)	.4616 <sup>a</sup>
Caucasian	29 (90.6)	32 (88.9)	>.9999 <sup>a</sup>
Never smoker	18 (56.3)	17 (47.2)	.4777 <sup>a</sup>
Have family history of IBD		8 (22.2)	
Age at diagnosis			
A1 ( $\leq$ 16 yrs)		3 (8.3)	
A2 (17-40 yrs)		25 (69.4)	
A3 ( $>$ 40 yrs)		8 (22.2)	
Disease location			
L1 (Ileal)		22 (61.1)	
L2 (Colonic)		6 (16.7)	
L3 (Ileocolonic)		7 (19.4)	
L4 (upper small bowel)		1 (2.8)	
Disease behavior			
B1 (Non-stricturing, non-penetrating)		8 (22.2)	
B2 (Stricturing)		17 (47.2)	
B3 (Penetrating)		11 (30.6)	
Presence of upper small bowel disease		5 (13.9)	
Presence of perianal disease		8 (22.2)	
5-aminosalicylic acid (ASA) treatment		22/32 (68.8)	
Steroids treatment		22/30 (73.3)	
Antibiotics treatment		18/28 (64.3)	
Anti-TNF treatment		18/31 (58.1)	
Average age at surgery (years)	60.8	37.4	<b>&lt;.0001<sup>b</sup></b>

Values in bold are significant; <sup>a</sup> Fisher's exact test; <sup>b</sup> Two-tailed t-test.

**Table S3. Number of overlapping genes in fuzzy c-means clusters.**

**Table S7. Characteristics of the full UNITI-2 cohort and the subsets analyzed for gene expression and histology in this study <sup>a</sup>.**

Characteristic	UNITI-2 full cohort (n=511)	Gene expression subset (n=110)	Histology subset (n=54)
Male, number (%)	243 (47.6)	53 (48.2)	25 (46.3)
Age, mean (SD)	38.6 (13.4)	38.3 (13.4)	37.1 (13.9)
Disease duration, mean (SD)	10.0 (9.6)	9.0 (9.3)	8.6 (9.6)
CDAI, mean (SD)	301.9 (58.9)	308.6 (58.3)	306.9 (54.7)
CRP median	7.8	8.0	7.0
FCALP median	471.9	304.4 <sup>b</sup>	245.7 <sup>b</sup>
Ileal SES-CD, mean (SD)		4.5 (3.4)	3.1 (3.2) <sup>c</sup>

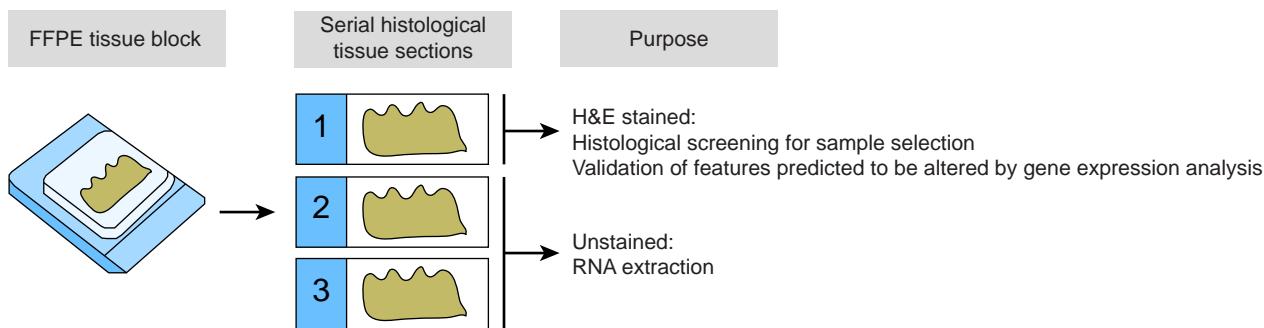
<sup>a</sup> Patients with colon-only disease involvement were excluded in all 3 populations.

<sup>b</sup> P-values <0.05 by Wilcoxon test comparing gene expression or histology subset vs. UNITI-2 full cohort.

<sup>c</sup> P-value <0.05 by Wilcoxon test comparing gene expression subset vs. histology subset.

## Supplementary Figure S1

**A**

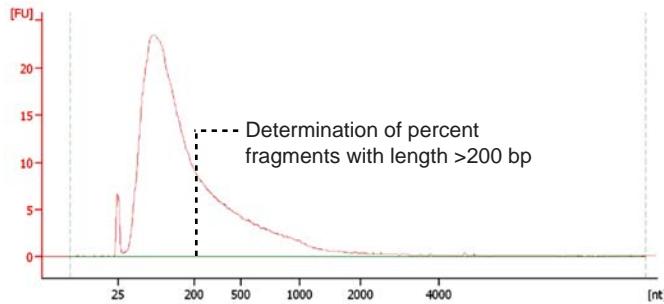


**B** Quality control workflow

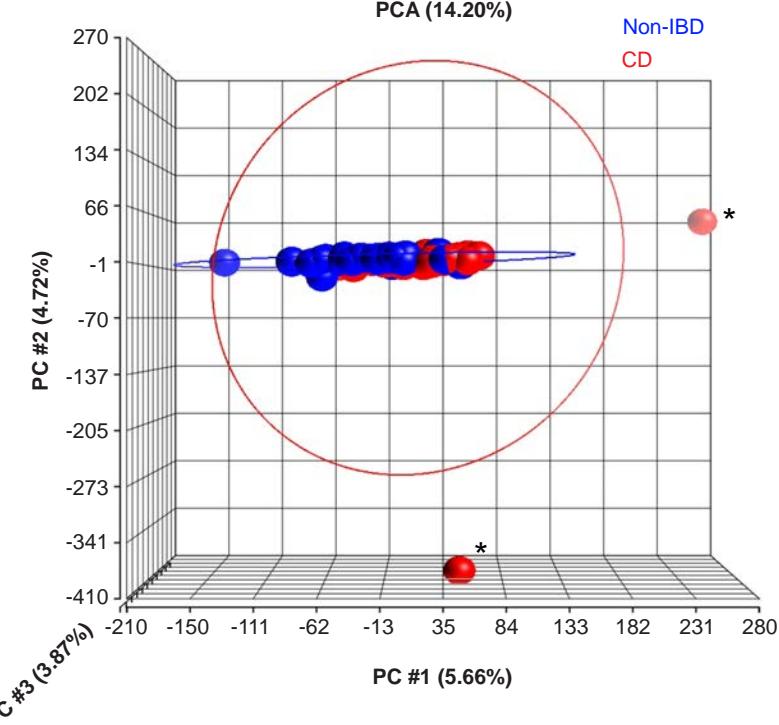
To be included in the study, a sample must...

1. Pass histological screening criteria
2. Provide sufficient RNA mass
3. Have >20% of RNA fragments with length >200 bp as determined by Bioanalyzer (Pre-seq QC)
4. Pass all technical RNA-seq QC metrics (Pre-seq QC)
5. Meet minimal criteria for # of reads, % unique reads on target, and genes detected (Post-seq QC)
6. Meet gene expression data distribution criteria (Post-seq QC)

**C** Representative Bioanalyzer plot from study sample



**D** PCA of gene expression data with outliers included



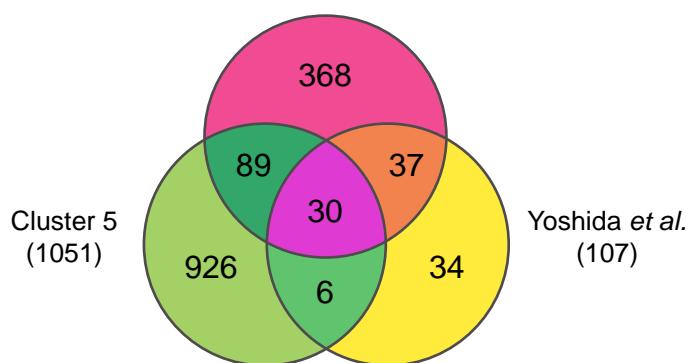
**E**

Samples selected based on histological pre-screening	Diagnosis	Sufficient RNA yield	QC Pre-seq	QC Post-seq
Resection margins without active acute inflammation n = 76	Non-IBD n = 37	n = 35 (95%)	n = 32 (86%)	n = 32 (86%)
	CD n = 39	n = 39 (100%)	n = 38 (97%)	n = 36 (92%)

Supplementary Figure S2

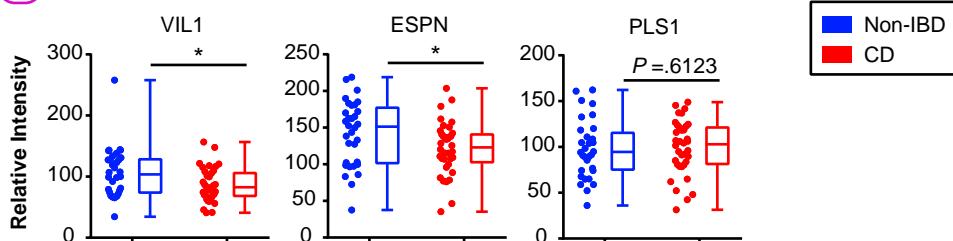
**A**

McConnell *et al.*  
(524)

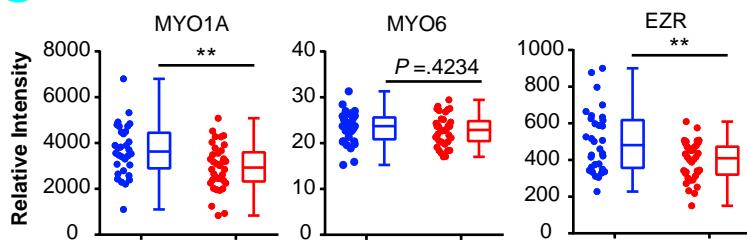


**B**

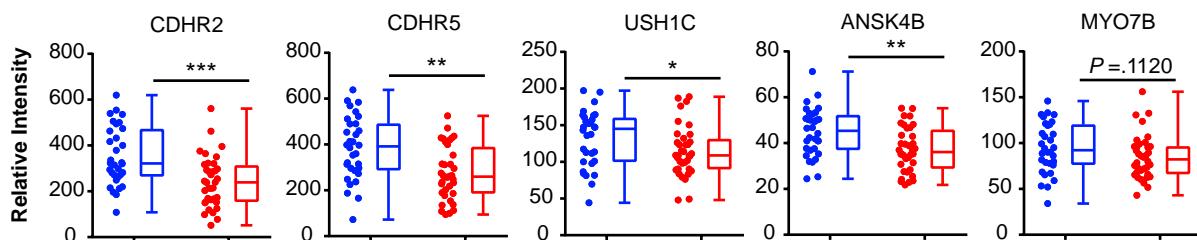
1 F-actin bundling



2 Membrane-cytoskeleton crosslinking

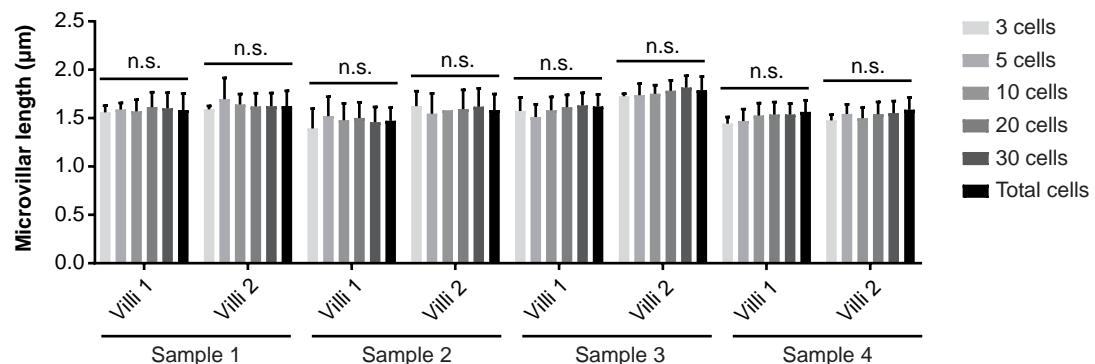


3 Intermicrovillar adhesion



### Supplementary Figure S3

#### A Determination of the number of cells per villi to measure for accurate sampling:



#### B

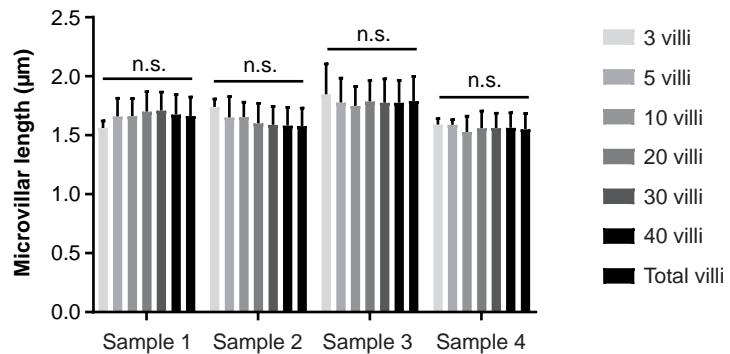
Variable	% of total variation	P-value	F-statistic
Sample	10.48	<0.0001	19.30
Cells counted	0.58	0.187	1.50
Interaction	1.13	0.999	0.41

#### C

Sample	Villus	3 cells	5 cells	10 cells	20 cells	30 cells
Sample 1	Villus 1	8.0%	4.3%	2.7%	0.3%	0.0%
Sample 1	Villus 2	7.2%	6.5%	3.1%	1.4%	0.1%
Sample 2	Villus 1	5.7%	3.6%	3.5%	1.9%	1.1%
Sample 2	Villus 2	6.7%	4.3%	3.8%	2.1%	0.7%
Sample 3	Villus 1	3.5%	4.5%	3.0%	1.1%	0.2%
Sample 3	Villus 2	7.3%	4.5%	2.2%	0.8%	0.0%
Sample 4	Villus 1	8.7%	5.8%	2.9%	0.9%	0.3%
Sample 4	Villus 2	7.0%	5.4%	2.3%	1.0%	0.4%
<b>AVERAGE</b>		<b>6.8%</b>	<b>4.9%</b>	<b>2.9%</b>	<b>1.2%</b>	<b>0.4%</b>

## Supplementary Figure S4

**A** Determination of the number of villi per sample to measure for accurate sampling:

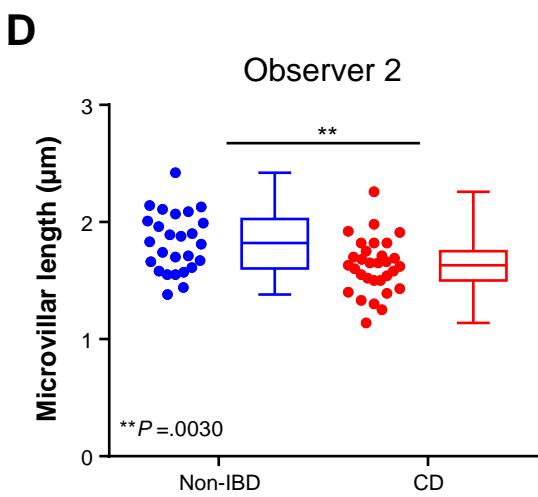
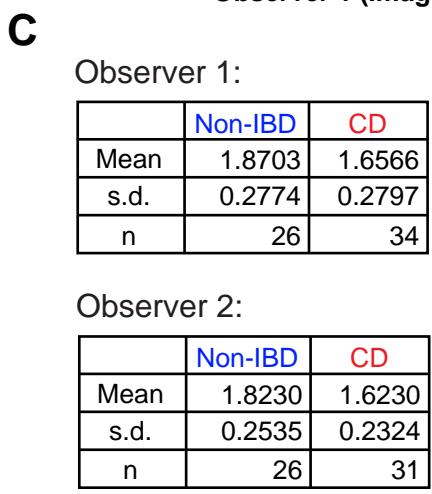
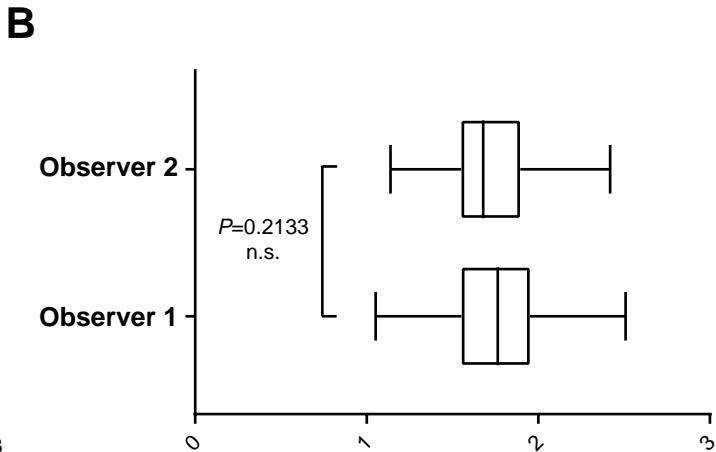
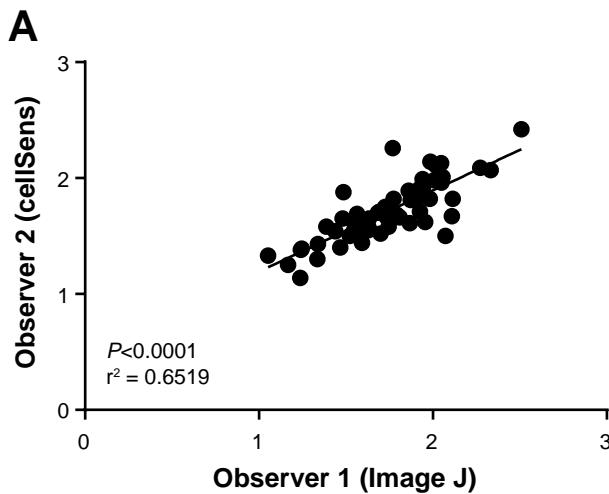


<b>B</b>	<b>Variable</b>	<b>% of total variation</b>	<b>P-value</b>	<b>F-statistic</b>
	Sample	9.46	<0.0001	21.56
	Villi counted	0.24	0.950	0.27
	Interaction	1.97	0.763	0.75

<b>C</b>	<b>Sample</b>	<b>3 villi</b>	<b>5 villi</b>	<b>10 villi</b>	<b>20 villi</b>	<b>30 villi</b>	<b>40 villi</b>
	Sample 1	6%	3%	2%	0%	0%	0%
	Sample 2	8%	7%	3%	0%	0%	0%
	Sample 3	5%	4%	3%	1%	0%	0%
	Sample 4	5%	4%	2%	0%	0%	0%
	<b>AVERAGE</b>	<b>6%</b>	<b>4%</b>	<b>2%</b>	<b>0%</b>	<b>0%</b>	<b>0%</b>

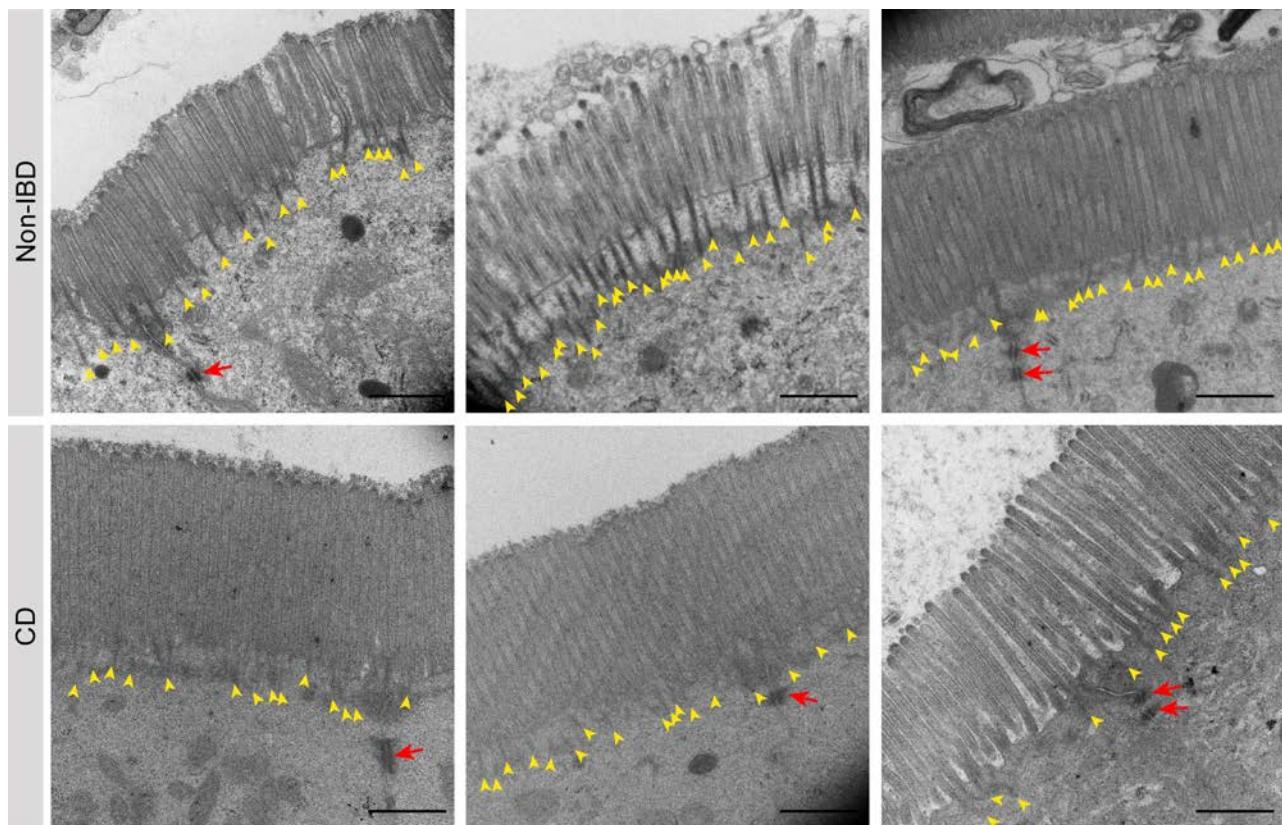
Supplementary Figure S5

Inter-observer reproducibility:

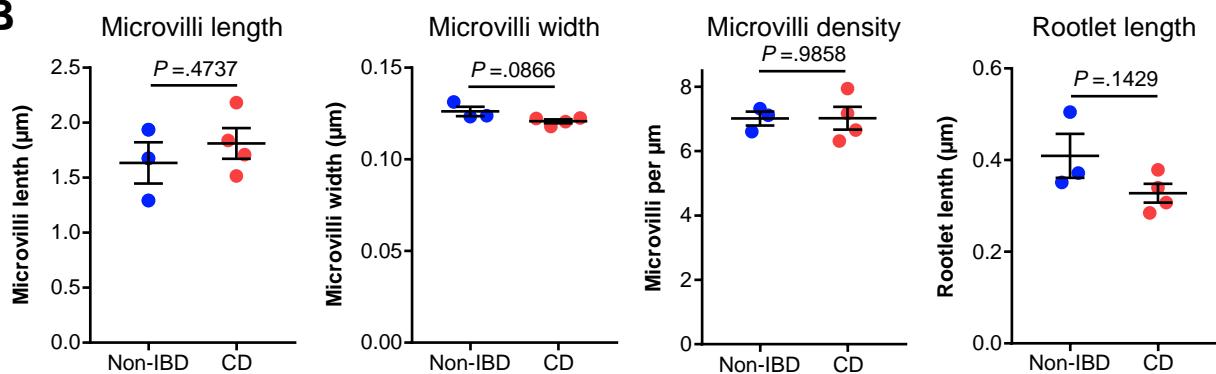


Supplementary Figure S6

**A**

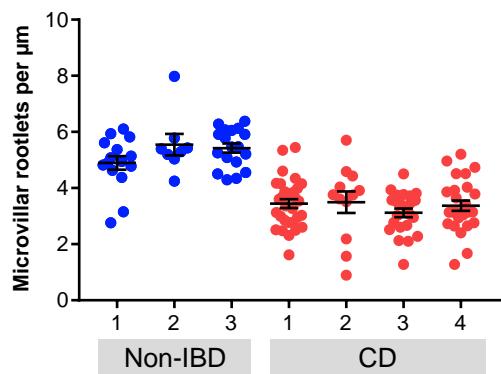


**B**



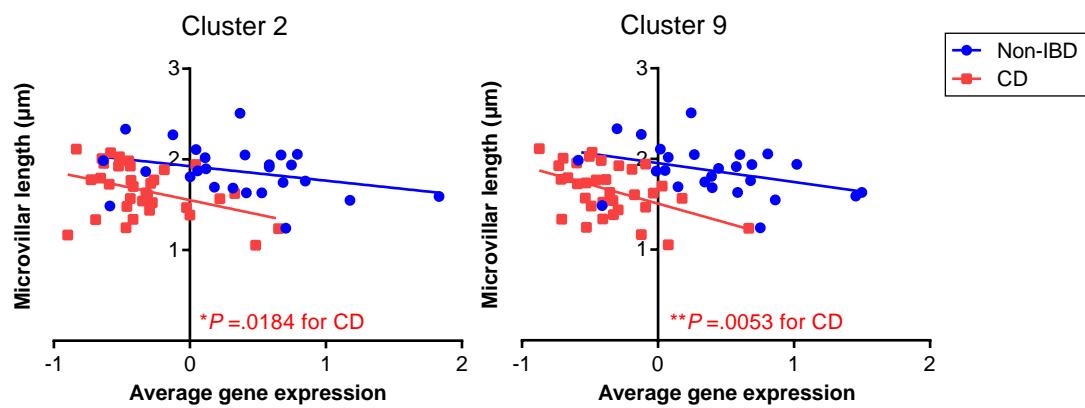
**C**

Source data for rootlet density (Fig. 3D)

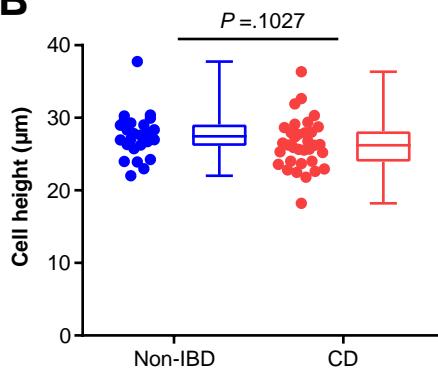


Supplementary Figure S7

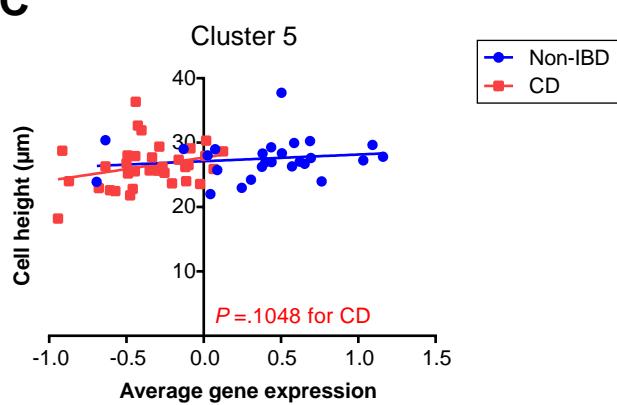
**A**



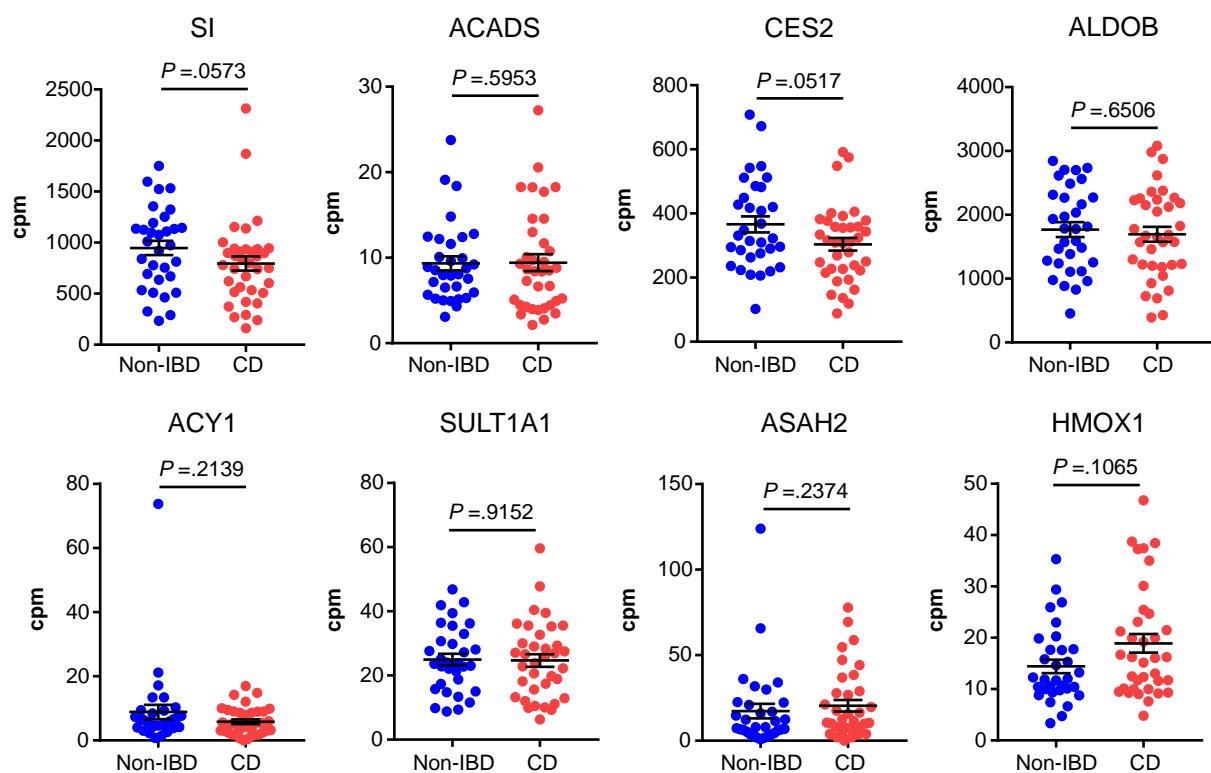
**B**



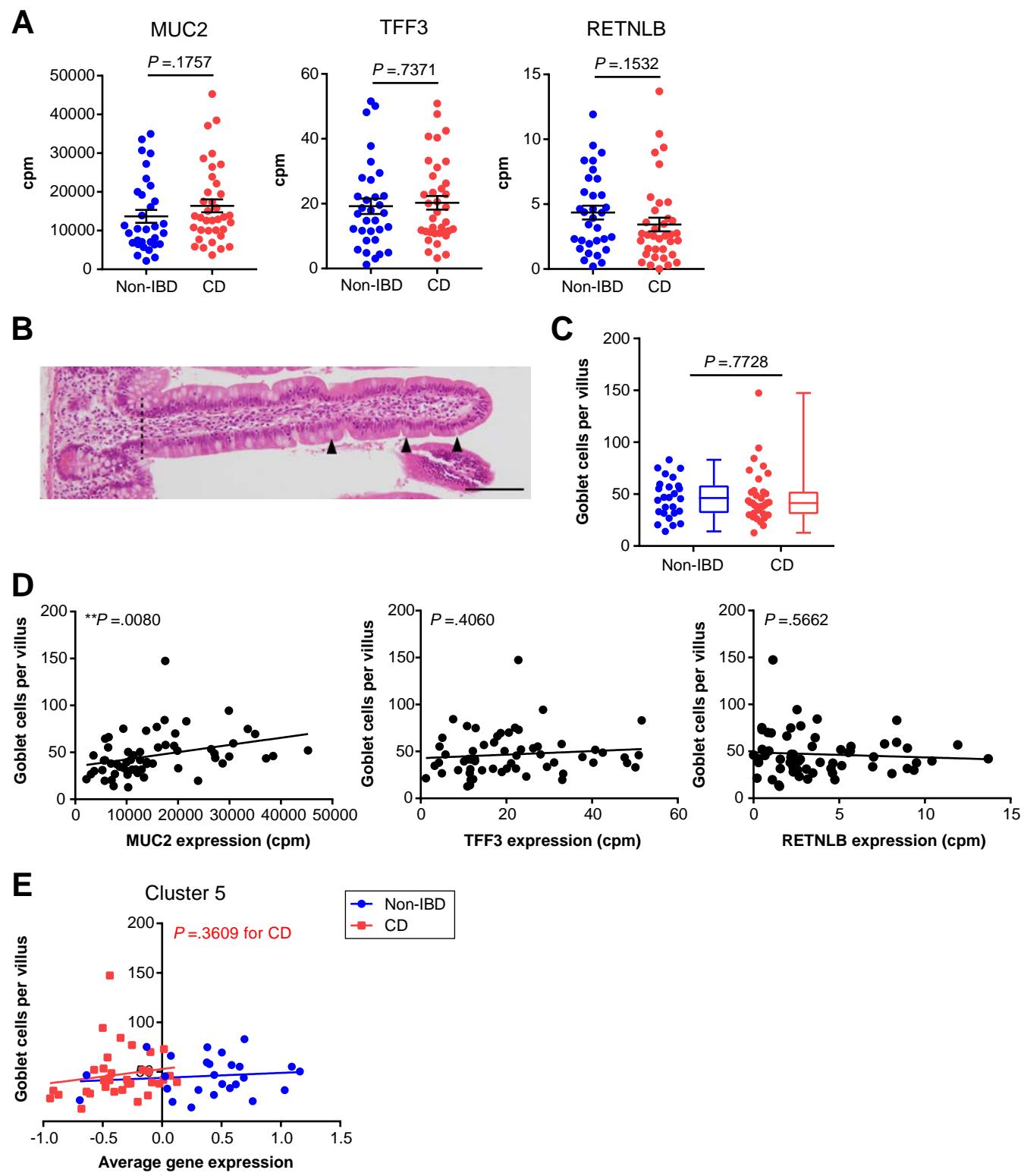
**C**



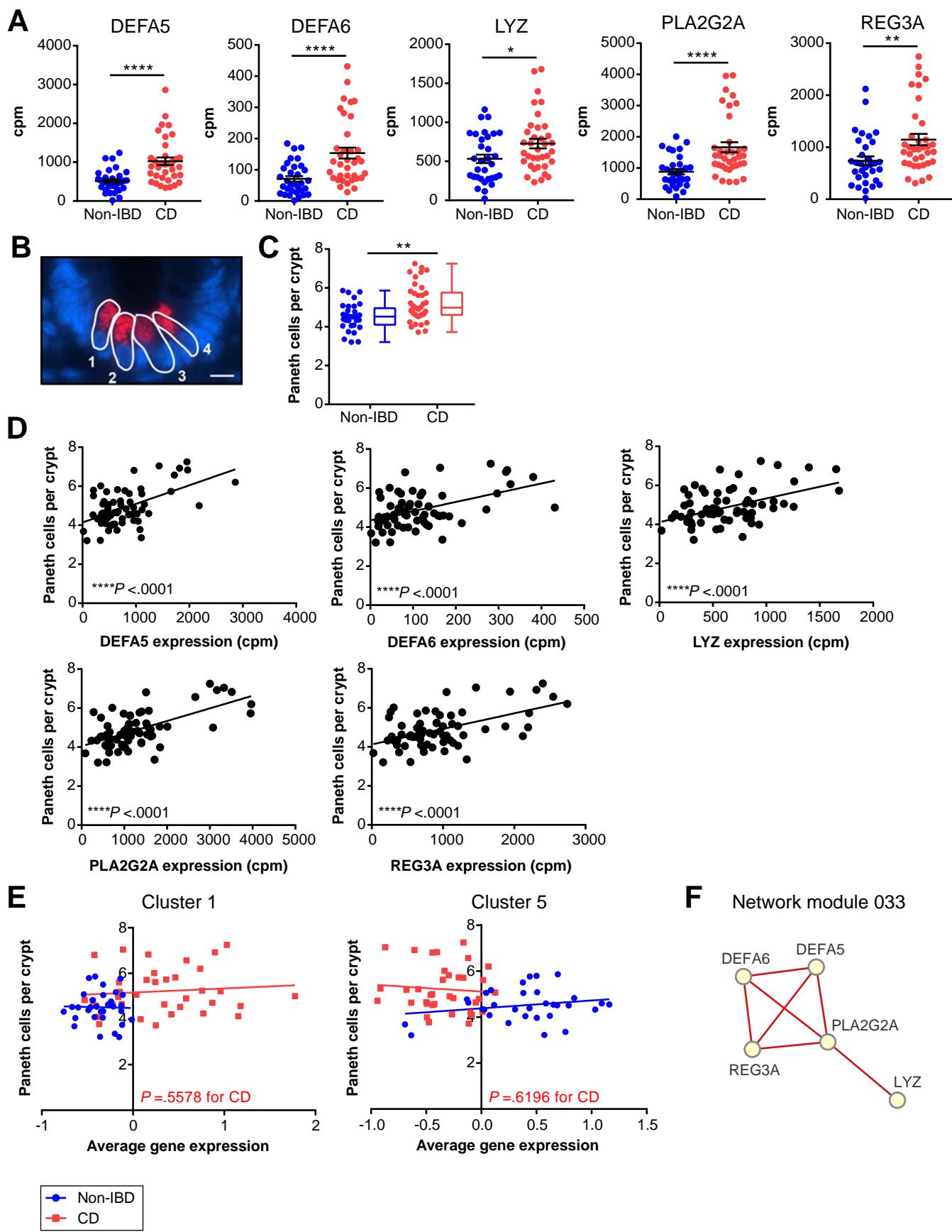
Supplementary Figure S8



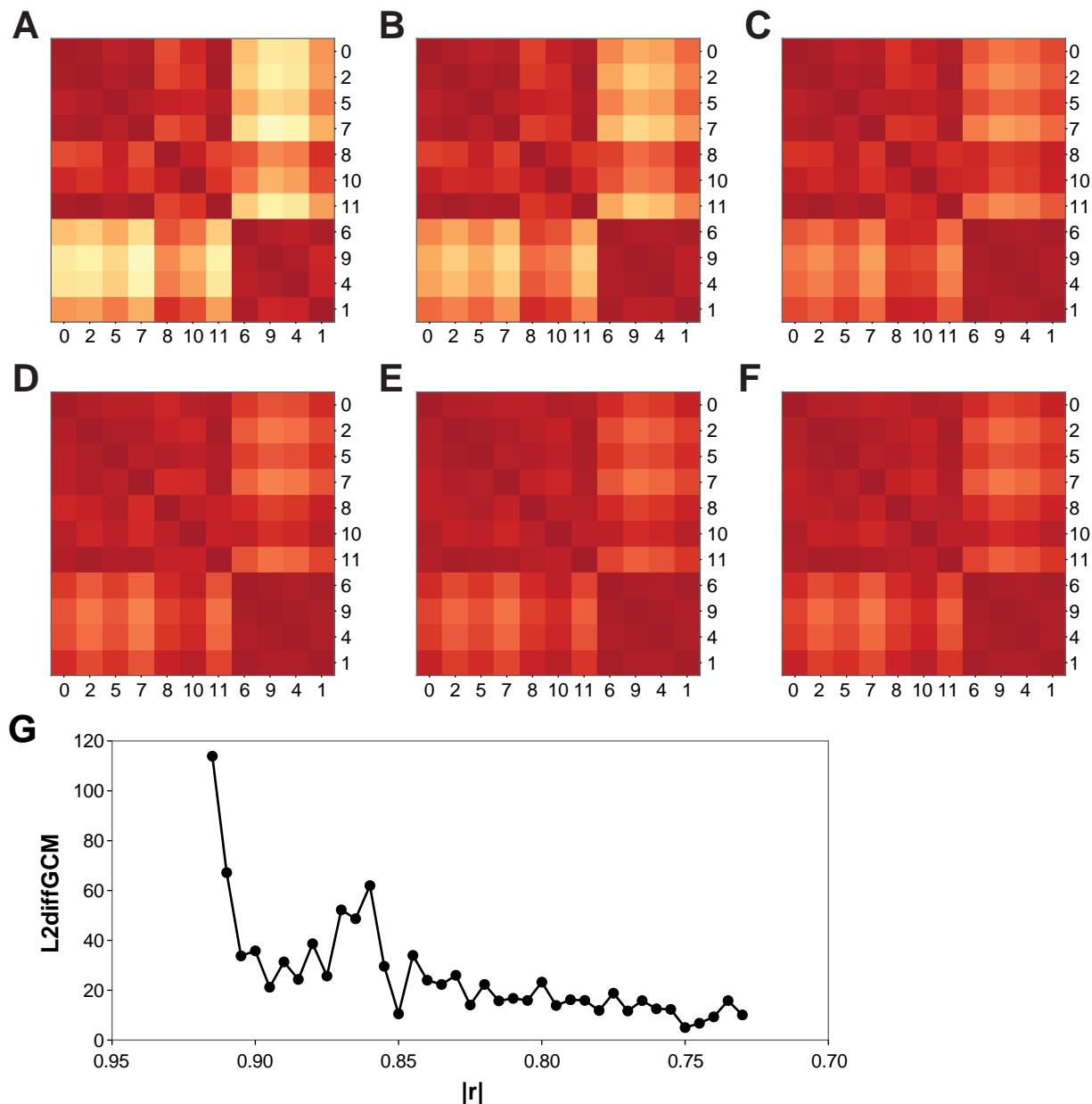
Supplementary Figure S9



Supplementary Figure S10

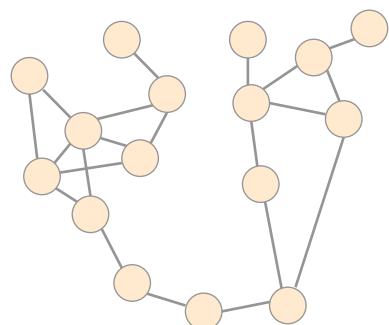


Supplementary Figure S11

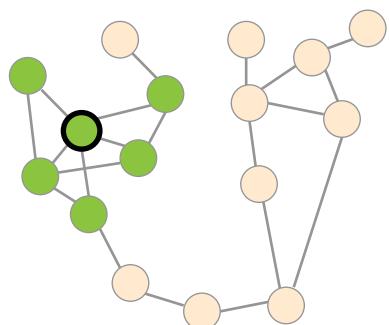


Supplementary Figure S12

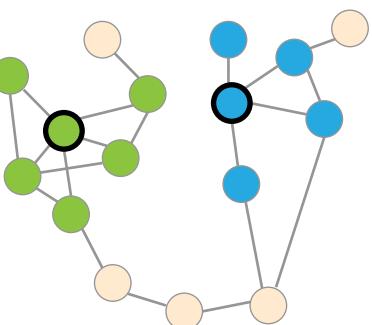
A



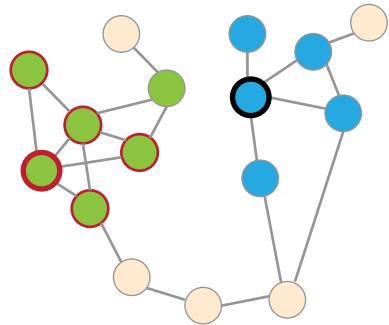
B



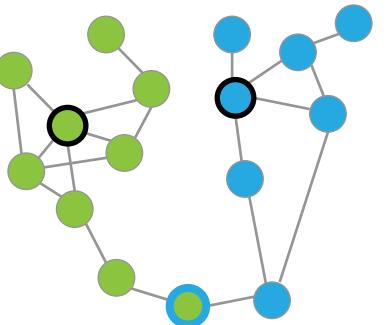
C



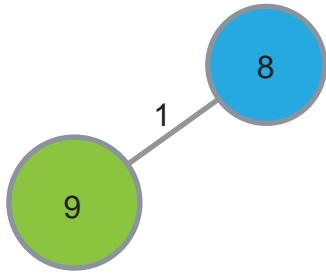
D



E

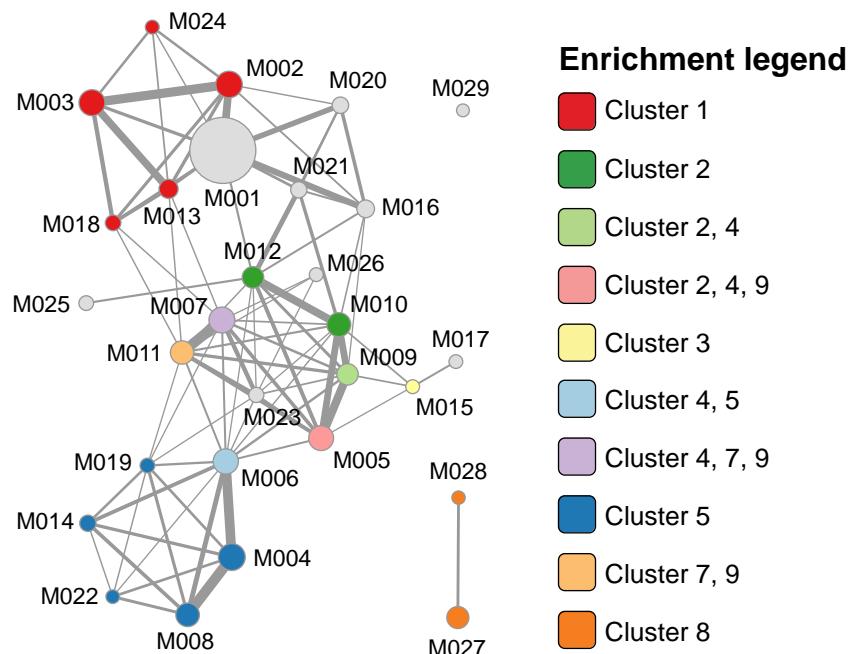


F

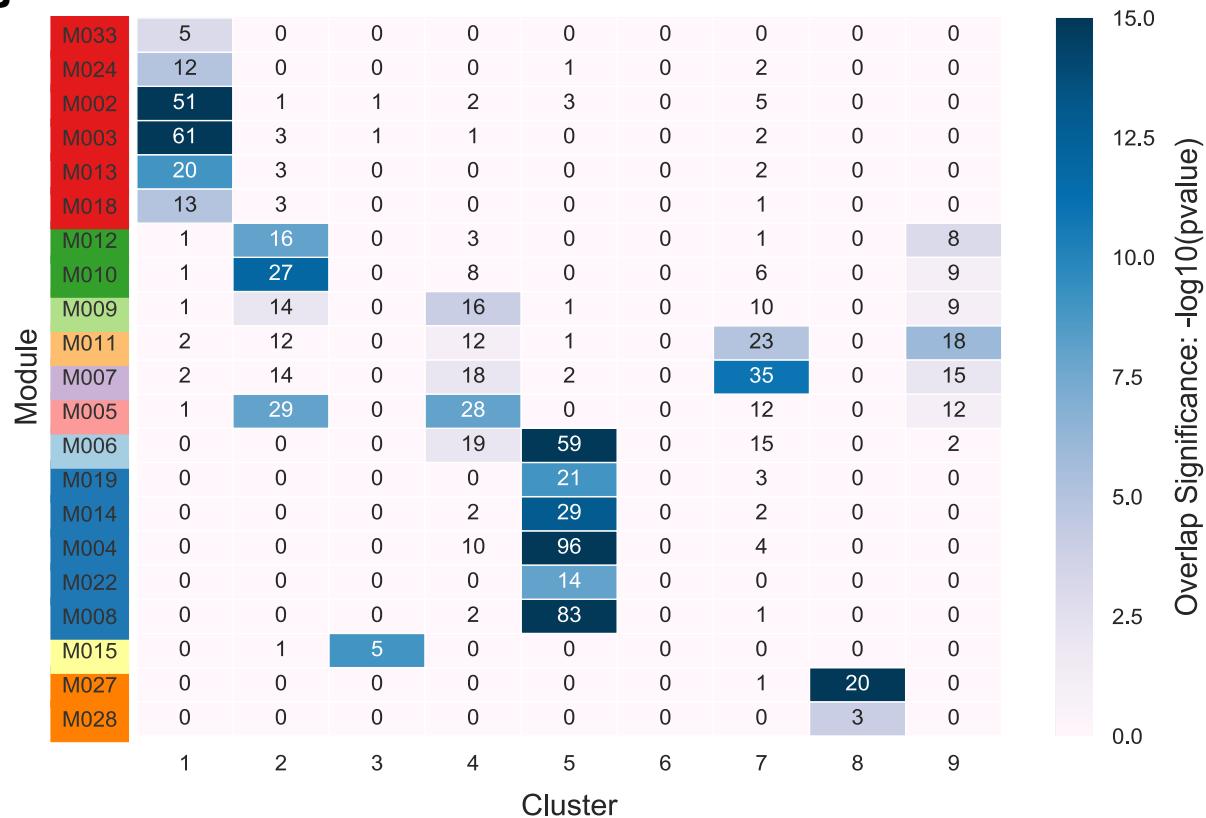


Supplementary Figure S13

**A**

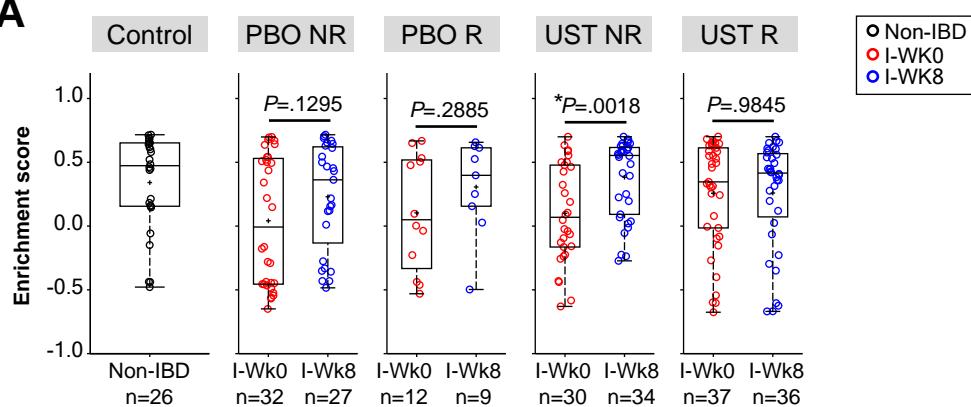


**B**

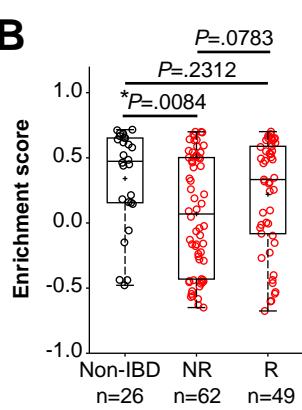


Supplementary Figure S14

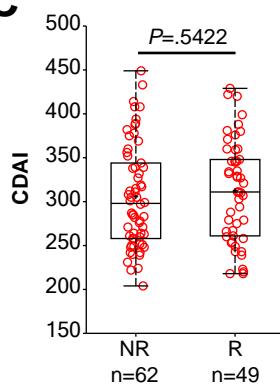
**A**



**B**



**C**



**D**

