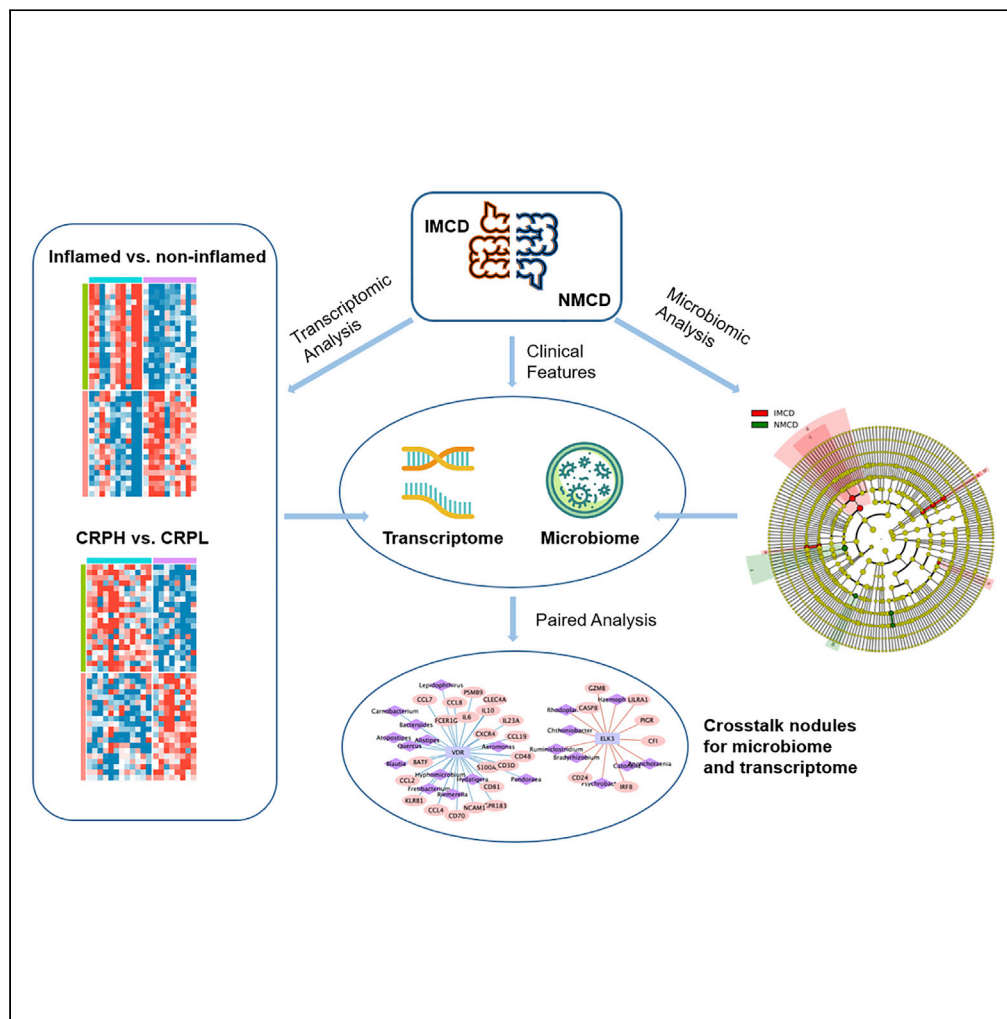## Article

Relating the transcriptome and microbiome by paired terminal ileal Crohn disease

Chenwen Cai,
Sibo Zhu, Jinlu
Tong, Tianrong
Wang, Qi Feng,
Yuqi Qiao, Jun
Shen

qiaoyuqi@renji.com (Y.Q.)
shenjun79@sina.cn (J.S.)

Highlights

Mucosa-attached
microbiota differs
between IMCD and
NMCD

Gene transcriptomes
differ between inflamed
and non-inflamed ileal
tissues

Mucosa-attached
microbiome is
significantly altered due to
segmental inflammation

Microbiome interacts with
immune-related genes
through transcription
factors

Article

Relating the transcriptome and microbiome by paired terminal ileal Crohn disease

Chenwen Cai,[1,2,5] Sibo Zhu,[3,5] Jinlu Tong,[1] Tianrong Wang,[1] Qi Feng,[4] Yuqi Qiao,[1,*] and Jun Shen[1,6,*]

SUMMARY

Management of terminal ileal Crohn disease (CD) is difficult due to fibrotic prognosis and failure to achieve mucosal healing. A limited number of synchronous analyses have been conducted on the transcriptome and microbiome in unpaired terminal ileum tissues. Therefore, our study focused on the transcriptome and mucosal microbiome in terminal ileal tissues of patients with CD with the aim of determining the role of cross-talk between the microbiome and transcriptome in the pathogenesis of terminal ileal CD. Mucosa-attached microbial communities were significantly associated with segmental inflammation status. Interaction-related transcription factors (TFs) are the panel nodes for cross-talk between the gene patterns and microbiome for terminal ileal CD. The transcriptome and microbiome in terminal ileal CD can be differently related to the local inflammatory status, and specific differentially expressed genes may be targeted for mucosal healing. TFs connect gene patterns with the microbiome by reflecting environmental stimuli and signals from microbiota.

INTRODUCTION

Crohn disease (CD) is a type of inflammatory bowel disease (IBD) (Ng et al., 2018) that may cause stenosis, perforation of lesions, sinus, fistula, and abscesses in patients, seriously affecting their quality of life. According to the Montreal classification, the terminal ileum (L1) and ileocolon (L3) are frequently involved in CD. According to the European EpiCom study, ileocecal and terminal ileal CD occurs in about 62% of Western Europeans and 75% of Eastern Europeans (Burisch et al., 2014; Golovics et al., 2014). An epidemiological study in Asia revealed that terminal ileal CD occurs in 76.8% of Asian patients (Ng et al., 2016). Based on the location of CD, distinct differences in signaling pathways and microbiota may present risk for disease progression or complication, influencing disease management decisions. Specific gene panels are predominantly associated with ileal and colonic CD (Baumgart and Sandborn, 2012; Verstockt et al., 2018). Microbial community profiles are also significantly different in patients with ileal and colonic CD, and they are independent of biopsy site, inflammatory state, or disease activities (Naftali et al., 2016).

Genetic and environmental factors play combined roles in the pathogenesis of CD (Ananthakrishnan et al., 2018; Satokari, 2015). The relationship between genetic factors and CD has been studied for a long time. A number of studies have analyzed the IBD transcriptome (Chan et al., 2018). Most studies used arrays and RNA sequencing to determine the mucosal transcriptome and the pathogenesis of IBD (Fang et al., 2015; Hong et al., 2017; Palmieri et al., 2015; Planell et al., 2013; Yang et al., 2019). Hong et al. studied paired inflammatory and non-inflammatory tissues in patients with CD (Hong et al., 2017). Interactions between gut microbes and host genes are involved in the development and progression of CD (Satokari, 2015). However, a limited number of studies have investigated the interactions between gut microbes and host genes, yielding inconsistent results due to the different methods employed (Pittayanon et al., 2019). Häsler et al. explored the relationship between the microbiome and the tissue transcriptome by analyzing unpaired inflammatory and non-inflammatory tissues involved in different diseases and different intestinal regions affected by IBD (Häsler et al., 2017).

To date, a limited number of synchronous studies have analyzed the transcriptome and microbiome in paired, specific locations of CD. Thus the precise roles of the microbiota and signaling pathways in CD are unknown. Therefore, our study focused on the transcriptome and mucosal microbiome in terminal ileal tissues of patients with CD with the aim of determining the role of cross-talk between the microbiome and transcriptome in the pathogenesis of terminal ileal CD.

[1]Division of Gastroenterology and Hepatology, Key Laboratory of Gastroenterology and Hepatology, Ministry of Health, Inflammatory Bowel Disease Research Center, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai Institute of Digestive Disease, 160# Pu Jian Avenue, Shanghai 200127, China

[2]Department of Gastroenterology, Huashan Hospital North, Fudan University, No.108 LuXiang Road, Shanghai 201907, China

[3]MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200438, China

[4]Department of Radiology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, 160 Pu Jian Road, Shanghai 200127, China

[5]These authors contributed equally

[6]Lead Contact

*Correspondence:
qiaoyuqi@renji.com (Y.Q.),
shenjun79@sina.cn (J.S.)
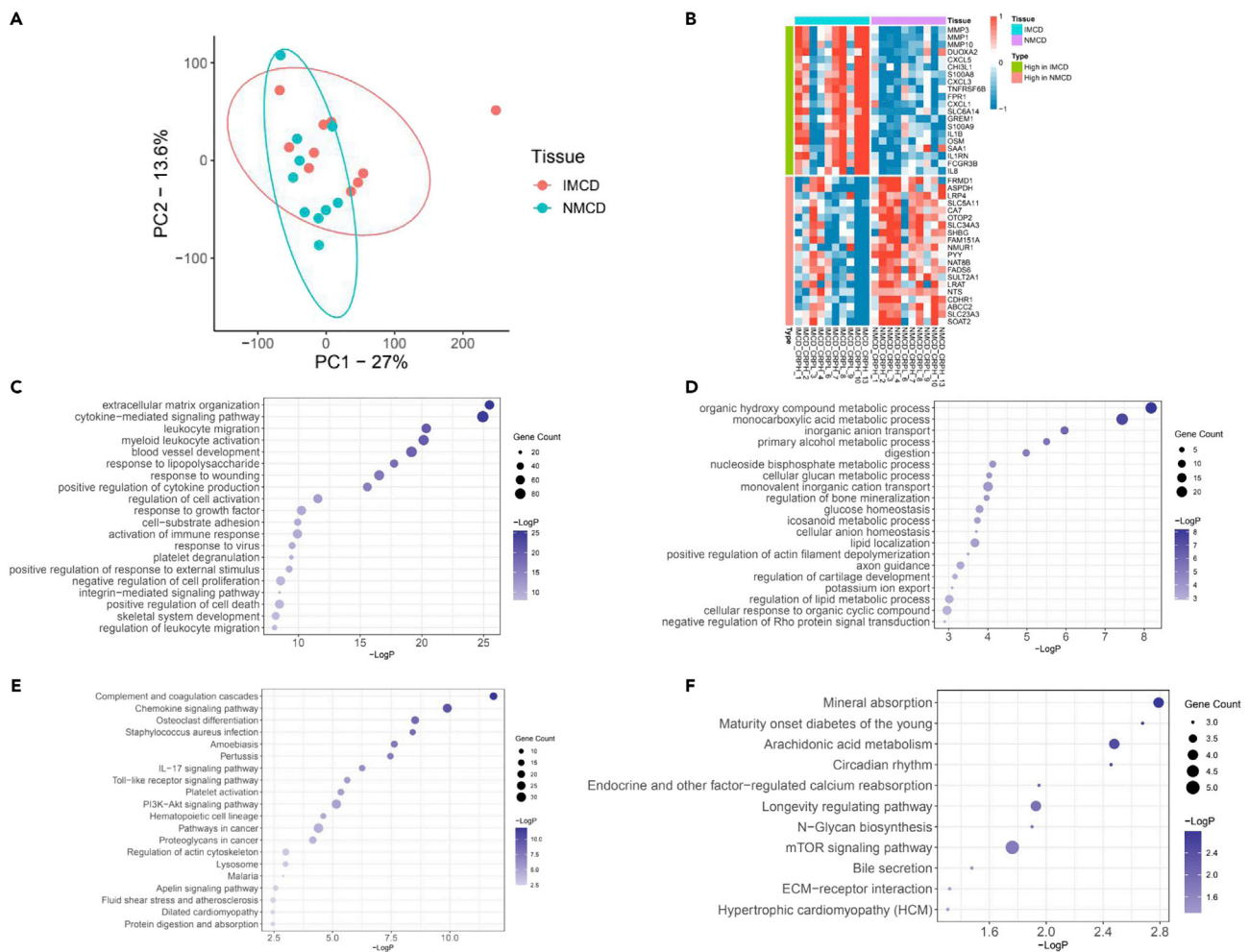https://doi.org/10.1016/j.isci.2021.102516

Figure 1. Transcriptome analysis for inflamed and non-inflamed tissue in patients with CD with terminal ileum involved

(A−F) Inflamed mucosa of Crohn disease (IMCD) and non-inflamed mucosa of Crohn disease (NMCD) samples shared a common area in principal-component analysis (PCA) (A); top 20 differentially expressed genes (DEGs) in IMCD and NMCD are detected (B); GO biological process analysis is applied for IMCD (C) and NMCD (D); KEGG pathway analysis identified different pathways in IMCD (E) and NMCD (F).

RESULTS

Identification of functional genes and inflammation-associated genes using transcriptome analysis

A total of 10 paired tissue samples from inflamed and proximal non-inflamed ileum of patients with new-onset CD were analyzed. Characteristics of the included samples are listed in Table S1. Gene patterns differ due to differences in biopsy site, inflammatory state, or disease activities. Therefore, we performed biopsy with a narrow margin, and we paired the inflamed and non-inflamed terminal ileal tissues. RNA sequencing identified a total of 22,335 genes, and principal-component analysis (PCA) was performed. However, the 2D PCA plots could not classify samples efficiently, that is, inflamed mucosa of Crohn disease (IMCD) and non-inflamed mucosa of Crohn disease (NMCD) samples were not clearly grouped by PCA (Figure 1A). However, 1,080 genes were differentially expressed between IMCD and NMCD in the terminal ileum ($p < 0.05$). Of the differentially expressed genes (DEG), 835 were upregulated and 245 were downregulated (Table S2). The top 20 genes (ranked by fold change [FC] with a cutoff p value of 0.05) were differentially regulated in either of the two subtypes (Figure 1B). *MMP3*, *MMP1*, and *DUOXA2* were highly expressed in IMCD samples, whereas *SOAT2*, *SLC23A3*, and *ABCC2* were highly expressed in NMCD samples. The Gene Ontology Biological Process (GOBP) analysis identified several functions of DEGs, and the predominant functions included overexpression of extracellular matrix organization, cytokine-mediated signaling pathway activation, leukocyte migration, myeloid leukocyte activation, and blood vessel development in IMCD (Figure 1C).
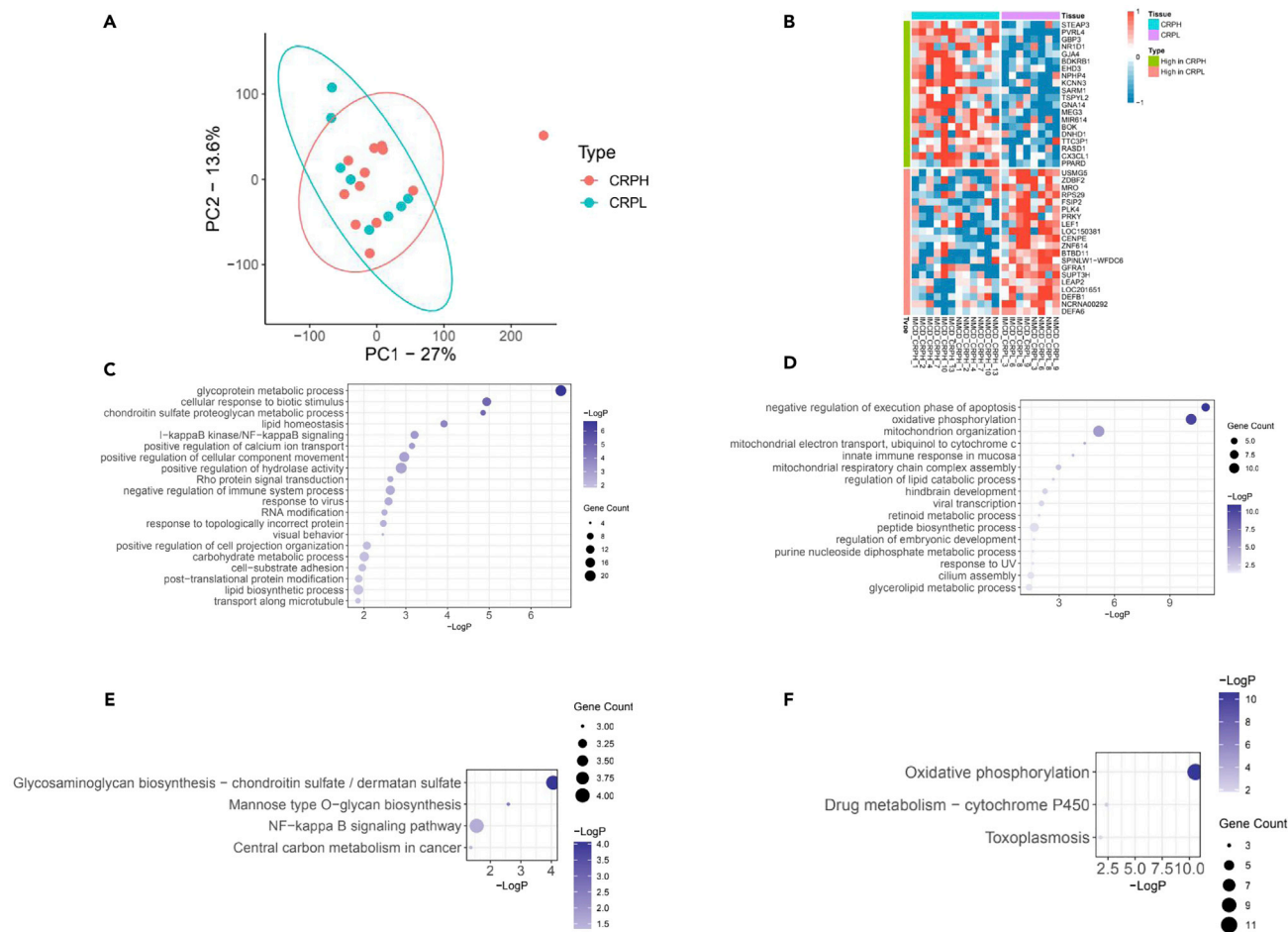
Figure 2. Transcriptome analysis for CRP-associated genes and pathways

(A−F) PCA cannot discriminate CRP high (CRPH) and low (CRPL) groups (A); the top 20 genes of CRPH and CRPL (B); GO biological process analysis is used for CRPH (C) and CRPL (D); KEGG pathway analysis identified different pathways in CRPH (E) and CRPL (F).

Additionally, genes related to organic hydroxyl and monocarboxylic acid metabolism were predominantly overexpressed in NMCD (Figure 1D). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed to identify the collection of manually drawn pathway maps, representing molecular interaction, reaction, and relation networks. Interestingly, DEGs in the KEGG pathway analysis were less than those in the GOBP analysis. The KEGG pathway analysis identified DEGs mainly related to complement and coagulation cascades, cytokine signaling pathway, and osteoclast differentiation in IMCD (Figure 1E), and it identified overexpressed genes related to mineral absorption, arachidonic acid metabolism, longevity-regulating pathway, and mammalian target of rapamycin signaling pathway in NMCD (Figure 1F).

Identification of C-reactive protein-associated genes and pathways using transcriptome analysis

A total of 415 genes were differentially expressed between C-reactive protein (CRP) high CRPH and CRP low (CRPL) groups ($p < 0.05$). Of the DEGs, 303 were upregulated and 112 were downregulated (Table S3). However, PCA demonstrated that CRP was not discriminated effectively as an inflammatory biomarker between the gene expression profiles (Figure 2A). The top 20 genes (ranked by FC with a cutoff p value of 0.05) were differentially regulated in either of the two subtypes (CRPH and CRPL). *STEAP3*, *PCRL4*, and *GBP3* were highly expressed in CRPH, whereas *DEFA6*, *NCRNA00292*, and *DEFB1* were highly expressed in CRPL (Figure 2B). The GOBP analysis identified predominant functional categories, such as overexpression of glycoprotein metabolic process, cellular response to biotic stimulus, positive regulation of cellular component movement, and positive regulation of hydrolase activity in CRPH (Figure 2C). Additionally, genes associated
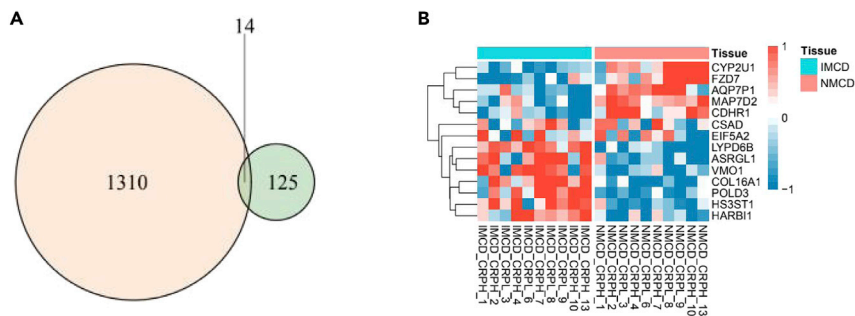
Figure 3. Venn diagram for differently expressed genes according to inflammatory state or systemic inflammatory responses

Fourteen DEGs of IMCD versus NMCD between CRPH (1,310 DEGs) and CRPL (125 DEGs) were identified by Venn diagram (A). Up- and downregulated gene clusters are determined (B).

with negative regulation of the execution phase of apoptosis, oxidative phosphorylation, and mitochondrion organization were predominantly overexpressed in CRPL (Figure 2D). However, the KEGG pathway analysis identified only a few differentially overexpressed genes related to glycosaminoglycan biosynthesis, such as chondroitin sulfate/dermatan sulfate synthesis and NF-κB signaling pathway, in CRPH (Figure 2E), and genes related to the oxidative phosphorylation pathway ware overexpressed in CRPL (Figure 2F).

An intersection of the DEGs of IMCD and NMCD between CRPH (1,310 DEGs) and CRPL (125 DEGs) was observed (Figure 3A). Fourteen DEGs, including the upregulated gene clusters *CSAD*, *EIF5A2*, *LYPD6B*, *ASRGL1*, *VMO1*, *COL16A1*, *POLD3*, *HS3ST1*, and *HARBI1* and the downregulated clusters *CDHR1*, *MAP7D2*, *FZD7*, *CYP2U1*, and *AQP7P1*, were screened out (Figure 3A) to form a diagram (Figure 3B). The 14 DEGs may reflect a local, segmental inflammatory state independent of or before systemic inflammatory responses.

Microbiome analysis detected subtle alteration in the ileal mucosa-attached microbiome

Analysis of 20 ileal mucosa-attached bacterial samples resulted in a total of 44.7 million 16S rRNA amplicons with an average of 2.24 million reads sequenced per sample. A total of 19 phyla and 3,120 operational taxonomical units (OTUs) were detected, and the most dominant phyla were Proteobacteria (68.39%), Actinobacteria (11.86%), Firmicutes (7.99%), and Bacteroidetes (7.17%). The percentages of these bacteria were 69.91%, 12.09%, 6.87%, and 5.69%, in IMCD and 66.90%, 11.66%, 9.09%, and 8.62%, in NMCD, respectively (Figure 4A). OTU and Shannon-Wiener index in IMCD and NMCD samples did not show any significant difference in bacterial richness and diversity (Figure 4B). Similarly, the 3D principal coordinate analysis plot using the Bray-Curtis index did not reveal any significant difference between IMCD and NMCD microbial communities (Figure 4C). Interestingly, linear discriminant analysis (LDA) and linear discriminant analysis effect size (LEfSe) revealed distinct clusters based on the inflamed or non-inflamed mucosa of CD. LDA indicated that Pseudomonadales, *Bacillus*, and Micrococcaceae were abundant in NMCD samples, whereas Ruminococcaceae, *Haemophilus*, Leptotrichiaceae, *Leptotrichia*, Burkholderiales, and Comamonadaceae were abundant in IMCD samples (Figure 4D). The cladogram generated from LEfSe analysis indicated that IMCD samples exhibited an increased proportion of *Leptotrichia*, Leptotrichiaceae, Comamonadaceae, Burkholderiales, *Haemophilus*, and Ruminococcaceae when compared with the NMCD samples. IMCD samples demonstrated a decreased proportion of Pseudomonaldales and Micrococcaceae compared with the NMCD samples (Figure 4E). The combination of these observations indicates that the mucosa-attached bacterial communities were significantly altered by segmental inflammation.

Co-abundance analysis of microbiome and host transcriptome detected microbiota-associated transcription factors

Transcriptioon factors (TFs) are nodes, where cells respond to extracellular information as environmental stimuli and signals from microbiota. To determine the host-microbiome cross-talk, DEGs, TFs, and TF-regulated genes with abundant bacterial OTUs were correlated. After screening 1,080 DEGs, which were differentially expressed between IMCD and NMCD, TFs were identified to determine the cross-talk between gene patterns and microbiome for terminal ileal CD. After sorting according to Pearson correlation, the top 100 DEGs were used to draw co-expressing networks. A small panel of TFs (IRF4, AHR, MEIS2, ZEB1, SMARCAI, CREB5, ARNTL2, and MEIS1) connected the microbiota (*Eikenella*, *Moraxella*, *Candldatus*,
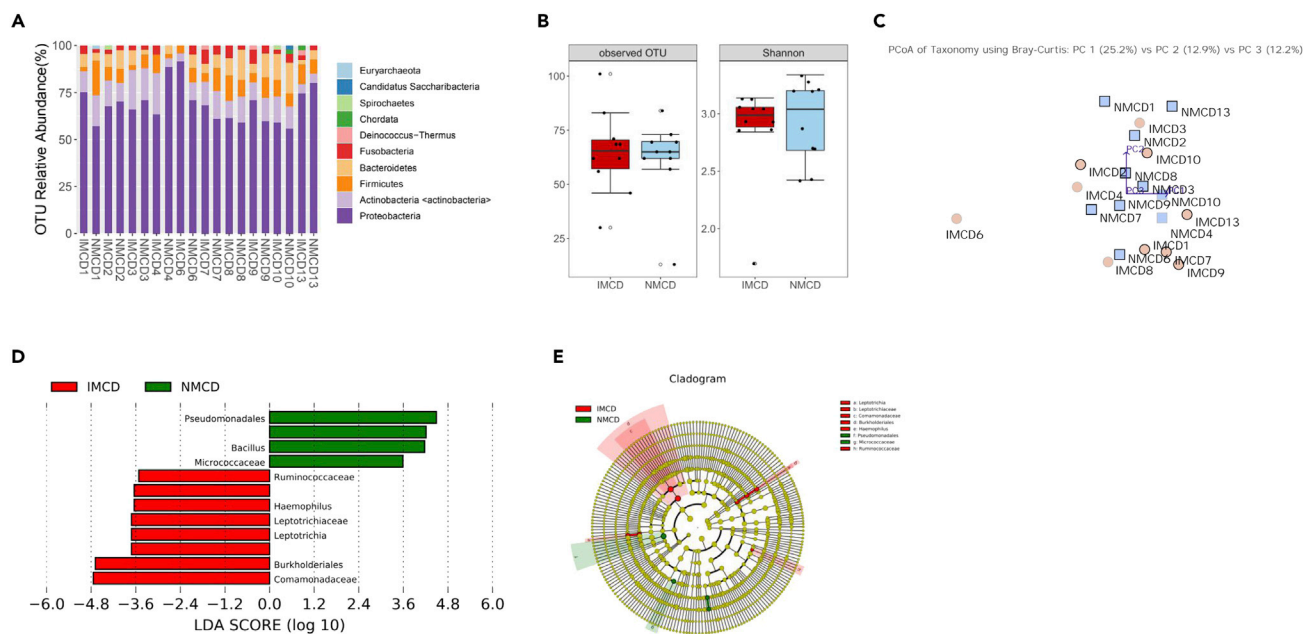
Figure 4. Microbiome analysis detected mucosa-attached microbiota

(A—E) OTU relative abundance for bacteria richness in IMCD and NMCD (A); OTU, Shannon-Wiener index, and principal coordinate analysis (PCoA) plot in IMCD and NMCD samples did not show significant difference (box plots include the minimum and maximum data values, the median, and lower and upper quartiles) (B and C); linear discriminant analysis (LDA) (D) and cladogram generated from LEfSe analysis (E) indicated distinct clusters based on inflamed or noninflamed mucosa of CD.

Halyseosphaera, Rhodococcus, Punica, Butyrivibrio, Mycoplasma, Faecalibacterium) to the majority of DEGs (Figure 5A). Two individual TFs (VDR and ELK3) were at the center of linking microbiota and DEGs (Figure 5B), which may present trigger points for analysis of host-microbiome cross-talk in terminal ileal CD.

The biological processes and signaling pathways of genes positively associated with TFs are shown in Figures 5C and 5D, respectively. Cytokine-mediated signaling pathway, lymphocyte activation, and regulation of cytokine production are the three predominant biological processes in IMCD, whereas cytokine-cytokine receptor interaction is the predominant signaling pathway in NMCD. The pathways involved in genes, negatively associated with TFs, were not enriched. The biological and signal pathways involved in TF-related genes of NMCD are shown in Figures 5E and 5F (positive correlation) and Figures 5G and 5H (negative correlation). Interestingly, cytokine-mediated signaling pathway and cytokine-cytokine receptor interaction are the common pathways related to TFs in terminal ileal CD, independent of the inflammatory status. The pathways involved in positive or negative correlation with TFs were partly same (Figures 5E–5H). We used DEGs (Figure 5I) to construct a heatmap, and we observed that *VDR* was significantly increased in the inflammation site, whereas *ELK3* was significantly decreased. Taken together, these findings suggest remarkable interactions between the host transcriptome and the mucosal microbiome, which may be mediated by key TFs and TF panels.

DISCUSSION

In patients with terminal ileal CD, significant differences in the gene transcriptomes of inflamed and non-inflamed tissues, including genes belonging to the chemokine ligand (CXCL), matrix metalloproteinase (MMP), and interleukin (IL) families, were observed. Some DEGs only reflected local inflammation in specific CD segments, and they remained unaffected by systemic inflammation. Some differences in the mucosa-attached microbiota were observed between IMCD and NMCD. Intestinal microbes in the terminal ileum interact with immune-related genes through core TFs. This interaction may be related to the pathogenesis of CD.

Several studies have been conducted to determine the relationship between the expression of intestinal genes, composition of intestinal microbes, and onset of CD. Some gene variants, such as *NOD2/CARD15*, are closely related to the pathogenesis of CD (Heliö et al., 2003). Susceptibility genes associated
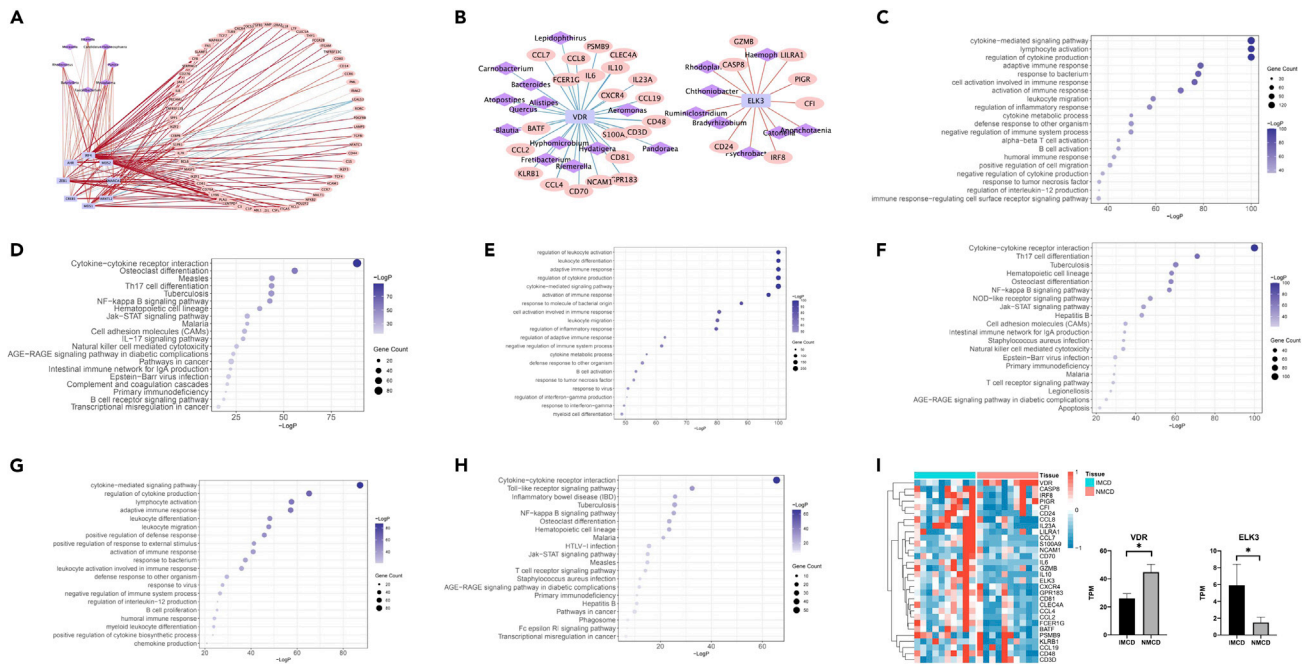
Figure 5. Co-abundance analysis of microbiome and host transcriptome

(A−I) The core panel of TFs determined main correlation connecting microbiota and DEGs (A); VDR and ELK3 are in the central position (B); the biological process (C) and signaling pathways (D) of genes positively related to TFs in IMCD; the biological process and signaling pathways of genes positively related to TFs in NMCD (E−H); heatmap and bar plots confirmed that VDR was significantly increased in the inflammation site, whereas ELK3 was significantly decreased (paired t test, *p < 0.05, data are represented as mean ± SEM) (I).

with the onset of CD differ in different populations (Ng et al., 2012). IBD-related transcriptomic studies have deepened our understanding of disease pathogenesis and prognosis (Chan et al., 2018). Analyses of the colonic tissues of patients with IBD have been conducted using cDNA microarrays, and abnormal regulatory genes related to disease pathogenesis (*CYLD, CDH11, ARNTL2,* and *RORA*) have been reported (Costello et al., 2005; Palmieri et al., 2015). Transcriptome analysis of patients with IBD of the colon and terminal ileal tissues has been conducted using RNA sequencing (Häsler et al., 2017; Hong et al., 2017). Häsler et al. (2017) investigated the relationship between gut microbes and transcriptomes, and they observed that IBD presented a special weakening relationship between the microbiota, host gene regulation, and splicing of host mucosal transcriptome. The combination of susceptibility genes, gut microbes, and transcriptome reflects the complicated pathogenesis and similar phenotypes of CD.

Considering that the terminal ileum is one of the major parts involved in CD, it presents difficulties in management of CD due to fibrosis and failure to achieve mucosal healing. In the present study, we found that multiple genes belonging to the CXCL, MMP, and IL families showed significantly different expression levels between IMCD and NMCD. These findings are similar to those of previous studies (Dobre et al., 2018; Hong et al., 2017; Jakubowska et al., 2016; Schuppan and Freitag, 2004). A significantly increased expression of MMPs, especially an increased level of MMP3 in mononuclear cells and fibroblasts, has been reported in the fistulas of CD (Kirkegaard et al., 2004; Panés and Rimola, 2017). These gene families are closely associated with immune-related pathways. In GOBP analysis, these gene families were involved in the development of cytokine-mediated signaling pathway, leukocyte migration, and myeloid leukocyte activation, and the results were similar to the KEGG analysis. Genes such as *DOUX2* (Haberman et al., 2014), *S100A8/A9* (Okada et al., 2019), *CHI3L1* (Deutschmann et al., 2019), *TNFRSF6B* (Eriksson et al., 2008), and *SLC6A14* (Luther et al., 2018) are also associated with CD. Although the colonic and terminal ileal tissues share a similar inflammatory status in CD, they may have different transcriptomes due to different genetic microenvironments and microbiomes.

Therapy using a treat-to-target approach followed by regular monitoring until the goal is achieved is critical for long-term therapeutic effects in CD. Invasive endoscopic procedures and non-invasive biomarkers such as CRP

or calprotectin are used in the evaluation of CD. CRP indicates systemic inflammatory conditions. CRP also acts as a predictor for mucosal healing, deep remission, and long-term outcomes for biologics, and thus, it indicates disease activity (Lin et al., 2020). Endoscopic evaluation and non-invasive biomarkers such as CRP may not be paralleled in CD. The subgroup of patients with CD presented endoscopic lesions and normal CRP. DEGs are affected by systemic inflammatory conditions; therefore, we examined the changes of gene transcriptomes and intestinal microbiomes in IMCD and NMCD by distinguishing the CRP levels of patients with a cutoff of 10 mg/L. Fourteen DEGs were unaffected by the systemic inflammatory state. It has been previously reported that the *CDHR1* and *FDZ7* are expressed in the colonic tissues of patients who are unresponsive to anti-TNF-α therapy (Luther et al., 2018). Owing to the limited number of studies, the role of DEGs in the pathogenesis of CD is inconclusive. However, DEGs are related to local mucosal healing rather than systemic inflammation or concomitant infection. Patients with CD who achieve clinical remission can still possess intestinal ulcers. Our results presented candidate pathways and genes as targets of mucosal healing.

To understand the differences in terminal ileal mucosa-associated microbiome and its relationship with the transcriptome, the same samples were used to analyze the gut microbiome. Proteobacteria, Actinobacteria, Firmicutes, Bacteroidetes, and Fusobacteria were detected in the terminal ileum of patients with CD, consistent with a recent study (Fan et al., 2020). The microbial abundance and diversity in patients with terminal ileal CD are low when compared with those of healthy controls and non-CD diseased controls (Häsler et al., 2017). Our study also indicated similar bacterial genera (Ruminococcaceae, *Haemophilus* [Wright et al., 2017], *Leptotrichia* [Salem et al., 2019], Burkholderiales [Sim et al., 2010], and Comamonadaceae [Chen et al., 2014]) with an increased transcriptome abundance in IMCD, consistent with previous studies (Henke et al., 2019).

We explored the interactions between immune genes, TFs, and gut microbiome. The results of the co-expression of OTU-TF-immune genes indicated that some gut microbes were closely related to the expression of immune-related genes. For example, in IMCD, *IRF4* interacted with several gut microbes such as *Rhodococcus* and *Butyrivibrio*. *IRF4* has been linked to the pathogenesis of CD (Watanabe et al., 2014; Xu et al., 2012). Similar situations exist for other TFs, such as AHR (Benson and Shepherd, 2011; Wu et al., 2019), ZEB1 (de Barrios et al., 2019), CREB5 (Jostins et al., 2012), and ARNTL2 (Palmieri et al., 2015). In NMCD, we observed that a decreased expression of *VDR* was associated with intestinal microbes. The relationship between the decreased *VDR* expression levels and CD has been reported previously (Garg et al., 2019; Gisbert-Ferrándiz et al., 2020). *ELK3* may also be associated with the pathogenesis of CD (Uniken Venema et al., 2019). The underlying mechanisms of the interactions of TFs with intestinal microbes leading to CD are still unclear. However, intestinal microbes may interact with immune-related pathways through TFs. The interaction may be a connection with CD local mucosal damage healing. As the immune pathways involved in IMCD and NMCD are highly similar, the determination of high TF expression in them is not difficult. Thus, there may be some interactions between the gut microbes and the host immune system.

Synchronous analyses suggest that there is an interaction between host genes and gut microbes in CD. Although we do not know the causal relationship of this interaction, the synchronous analyses provide the basis and research ideas for future research to explore the causal relationship between gut microbes and host gene expression changes. The strengths of our study are synchronous analyses of transcriptome and microbiome in paired, specific locations of CD. Second, endoscopic evaluations and non-invasive biomarkers, such as CRP, may not be paralleled to CD. Furthermore, the DEG matrix, which was not correlated with higher CRP levels, reflected local, segmental inflammatory status independent of or before systemic inflammatory response, and it could present candidate pathways and genes as targets for mucosal healing. Finally, a network of host-microbiome cross-talk was detected by microbiota-associated TFs.

Limitations of the study

Our study is limited by its small sample size, which limited the power to assess solid relationships between the transcriptome and microbiome. As this study did not analyze the differences between pre- and post-treatment (with administration of medicines) CD, we could not confirm whether the microbiome and host transcriptome were influenced by disease activity.

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jun Shen (shenjun79@sina.cn).

Materials availability

Materials availability has been described in the section of transparent methods.

Data and code availability

All data and code supporting the findings of this study are available within the paper and its supplemental information files (Data S1). Our RNA-seq data have been uploaded to the GEO database (GSE171244).

METHODS

All methods can be found in the accompanying Transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102516.

AUTHOR CONTRIBUTIONS

C.C., J.T., and T.W. collected the samples and analyzed data; S.Z. analyzed the data and performed the bioinformatics analysis; Q.F. reviewed the data and conclusions; Y.Q. and J.S. presented the idea of this paper, supported the funding, analyzed the conclusions, and drafted and revised the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Ananthakrishnan, A.N., Bernstein, C.N., Iliopoulos, D., Macpherson, A., Neurath, M.F., Ali, R.A.R., Vavricka, S.R., and Fiocchi, C. (2018). Environmental triggers in IBD: a review of progress and evidence. Nat. Rev. Gastroenterol. Hepatol. *15*, 39–49.

Baumgart, D.C., and Sandborn, W.J. (2012). Crohn's disease. Lancet (London, England) *380*, 1590–1605.

Benson, J.M., and Shepherd, D.M. (2011). Aryl hydrocarbon receptor activation by TCDD reduces inflammation associated with Crohn's disease. Toxicol. Sci. *120*, 68–78.

Burisch, J., Pedersen, N., Cukovic-Cavka, S., Brinar, M., Kaimakliotis, I., Duricova, D., Shonova, O., Vind, I., Avnstrom, S., Thorsgaard, N., et al. (2014). East-West gradient in the incidence of inflammatory bowel disease in Europe: the ECCO-EpiCom inception cohort. Gut *63*, 588–597.

Chan, S.N., Low, E.N.D., Raja Ali, R.A., and Mokhtar, N.M. (2018). Delineating inflammatory bowel disease through transcriptomic studies: current review of progress and evidence. Intestinal Res. *16*, 374–383.

Chen, L., Wang, W., Zhou, R., Ng, S.C., Li, J., Huang, M., Zhou, F., Wang, X., Shen, B., et al. (2014). Characteristics of fecal and mucosa-associated microbiota in Chinese patients with inflammatory bowel disease. Medicine *93*, e51.

Costello, C.M., Mah, N., Häsler, R., Rosenstiel, P., Waetzig, G.H., Hahn, A., Lu, T., Gurbuz, Y., Nikolaus, S., Albrecht, M., et al. (2005). Dissection of the inflammatory bowel disease transcriptome using genome-wide cDNA microarrays. PLoS Med. *2*, e199.

de Barrios, O., Sanchez-Moral, L., Cortés, M., Ninfali, C., Profitós-Pelejà, N., Martínez-Campanario, M.C., Siles, L., Del Campo, R., Fernández-Aceñero, M.J., Darling, D.S., et al. (2019). ZEB1 promotes inflammation and progression towards inflammation-driven carcinoma through repression of the DNA repair glycosylase MPG in epithelial cells. Gut *68*, 2129–2141.

Deutschmann, C., Sowa, M., Murugaiyan, J., Roesler, U., Röber, N., Conrad, K., Laass, M.W., Bogdanos, D., Sipeki, N., Papp, M., et al. (2019). Identification of chitinase-3-like protein 1 as a novel neutrophil antigenic target in Crohn's disease. J. Crohn's Colitis *13*, 894–904.

Dobre, M., Milanesi, E., Mănuc, T.E., Arsene, D.E., Țieranu, C.G., Maj, C., Becheanu, G., and Mănuc, M. (2018). Differential intestinal mucosa transcriptomic biomarkers for Crohn's disease and ulcerative colitis. J. Immunol. Res. *2018*, 9208274.

Eriksson, A., Jennische, E., Flach, C.F., Jorge, A., and Lange, S. (2008). Real-time PCR quantification analysis of five mucosal transcripts in patients with Crohn's disease. Eur. J. Gastroenterol. Hepatol. *20*, 290–296.

Fan, H.N., Zhu, P., Lu, Y.M., Guo, J.H., Zhang, J., Qu, G.Q., and Zhu, J.S. (2020). Mild changes in the mucosal microbiome during terminal ileum inflammation. Microb. Pathogenesis *142*, 104104.

Fang, K., Grisham, M.B., and Kevil, C.G. (2015). Application of comparative transcriptional genomics to identify molecular targets for pediatric IBD. Front. Immunol. *6*, 165.

Garg, M., Royce, S.G., Tikellis, C., Shallue, C., Sluka, P., Wardan, H., Hosking, P., Monagle, S., Thomas, M., Lubel, J.S., et al. (2019). The intestinal vitamin D receptor in inflammatory bowel disease: inverse correlation with inflammation but no relationship with circulating

vitamin D status. Ther. Adv. Gastroenterol. *12*, 1756284818822566.

Gisbert-Ferrándiz, L., Cosín-Roger, J., Hernández, C., Macias-Ceja, D.C., Ortiz-Masiá, D., Salvador, P., Esplugues, J.V., Hinojosa, J., Navarro, F., Calatayud, S., et al. (2020). Diminished vitamin D receptor protein levels in Crohn's disease fibroblasts: effects of vitamin D. Nutrients *12*, 973.

Golovics, P.A., Mandel, M.D., Lovasz, B.D., and Lakatos, P.L. (2014). Inflammatory bowel disease course in Crohn's disease: is the natural history changing? World J. Gastroenterol. *20*, 3198–3207.

Häsler, R., Sheibani-Tezerji, R., Sinha, A., Barann, M., Rehman, A., Esser, D., Aden, K., Knecht, C., Brandt, B., Nikolaus, S., et al. (2017). Uncoupling of mucosal gene regulation, mRNA splicing and adherent microbiota signatures in inflammatory bowel disease. Gut *66*, 2087–2097.

Haberman, Y., Tickle, T.L., Dexheimer, P.J., Kim, M.O., Tang, D., Karns, R., Baldassano, R.N., Noe, J.D., Rosh, J., Markowitz, J., et al. (2014). Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. J. Clin. Invest. *124*, 3617–3633.

Heliö, T., Halme, L., Lappalainen, M., Fodstad, H., Paavola-Sakki, P., Turunen, U., Färkkilä, M., Krusius, T., and Kontula, K. (2003). CARD15/NOD2 gene variants are associated with familially occurring and complicated forms of Crohn's disease. Gut *52*, 558–562.

Henke, M.T., Kenny, D.J., Cassilly, C.D., Vlamakis, H., Xavier, R.J., and Clardy, J. (2019). Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. Proc. Natl. Acad. Sci. U S A *116*, 12672–12677.

Hong, S.N., Joung, J.G., Bae, J.S., Lee, C.S., Koo, J.S., Park, S.J., Im, J.P., Kim, Y.S., Kim, J.W., Park, W.Y., et al. (2017). RNA-seq reveals transcriptomic differences in inflamed and noninflamed intestinal mucosa of Crohn's disease patients compared with normal mucosa of healthy controls. Inflamm. Bowel Dis. *23*, 1098–1108.

Jakubowska, K., Pryczynicz, A., Iwanowicz, P., Niewiński, A., Maciorkowska, E., Hapanowicz, J., Jagodzińska, D., Kemona, A., and Guzińska-Ustymowicz, K. (2016). Expressions of matrix metalloproteinases (MMP-2, MMP-7, and MMP-9) and their inhibitors (TIMP-1, TIMP-2) in inflammatory bowel diseases. Gastroenterol. Res. Pract. *2016*, 2456179.

Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature *491*, 119–124.

Kirkegaard, T., Hansen, A., Bruun, E., and Brynskov, J. (2004). Expression and localisation of matrix metalloproteinases and their natural inhibitors in fistulae of patients with Crohn's disease. Gut *53*, 701–709.

Lin, X., Qiu, Y., Feng, R., Chen, B., He, Y., Zeng, Z., Zhang, S., Chen, M., and Mao, R. (2020). Normalization of C-reactive protein predicts better outcome in patients with Crohn's disease with mucosal healing and deep remission. Clin. Translational Gastroenterol. *11*, e00135.

Luther, J., Gala, M., Patel, S.J., Dave, M., Borren, N., Xavier, R.J., and Ananthakrishnan, A.N. (2018). Loss of response to anti-tumor necrosis factor Alpha therapy in Crohn's disease is not associated with emergence of novel inflammatory pathways. Dig. Dis. Sci. *63*, 738–745.

Naftali, T., Reshef, L., Kovacs, A., Porat, R., Amir, I., Konikoff, F.M., and Gophna, U. (2016). Distinct microbiotas are associated with ileum-restricted and colon-involving Crohn's disease. Inflamm. Bowel Dis. *22*, 293–302.

Ng, S.C., Shi, H.Y., Hamidi, N., Underwood, F.E., Tang, W., Benchimol, E.I., Panaccione, R., Ghosh, S., Wu, J.C.Y., Chan, F.K.L., et al. (2018). Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet (London, England) *390*, 2769–2778.

Ng, S.C., Tsoi, K.K., Kamm, M.A., Xia, B., Wu, J., Chan, F.K., and Sung, J.J. (2012). Genetics of inflammatory bowel disease in Asia: systematic review and meta-analysis. Inflamm. Bowel Dis. *18*, 1164–1176.

Ng, S.C., Zeng, Z., Niewiadomski, O., Tang, W., Bell, S., Kamm, M.A., Hu, P., de Silva, H.J., Niriella, M.A., Udara, W.S., et al. (2016). Early course of inflammatory bowel disease in a population-based inception cohort study from 8 countries in Asia and Australia. Gastroenterology *150*, 86–95.e83, quiz e13-84.

Okada, K., Okabe, M., Kimura, Y., Itoh, H., and Ikemoto, M. (2019). Serum S100a8/A9 as a potentially sensitive biomarker for inflammatory bowel disease. Lab. Med. *50*, 370–380.

Palmieri, O., Mazzoccoli, G., Bossa, F., Maglietta, R., Palumbo, O., Ancona, N., Corritore, G., Latiano, T., Martino, G., Rubino, R., et al. (2015). Systematic analysis of circadian genes using genome-wide cDNA microarrays in the inflammatory bowel disease transcriptome. Chronobiol. Int. *32*, 903–916.

Panés, J., and Rimola, J. (2017). Perianal fistulizing Crohn's disease: pathogenesis, diagnosis and therapy. Nat. Rev. Gastroenterol. Hepatol. *14*, 652–664.

Pittayanon, R., Lau, J.T., Leontiadis, G.I., Tse, F., Yuan, Y., Surette, M., and Moayyedi, P. (2019). Differences in gut microbiota in patients with vs without inflammatory bowel diseases: a systematic review. Gastroenterology.

Planell, N., Lozano, J.J., Mora-Buch, R., Masamunt, M.C., Jimeno, M., Ordas, I., Esteller, M., Ricart, E., Pique, J.M., Panes, J., et al. (2013). Transcriptional analysis of the intestinal mucosa of patients with ulcerative colitis in remission reveals lasting epithelial cell alterations. Gut *62*, 967–976.

Salem, F., Kindt, N., Marchesi, J.R., Netter, P., Lopez, A., Kokten, T., Danese, S., Jouzeau, J.Y., Peyrin-Biroulet, L., and Moulin, D. (2019). Gut microbiome in chronic rheumatic and inflammatory bowel diseases: similarities and differences. United Eur. Gastroenterol. J. *7*, 1008–1032.

Satokari, R. (2015). Contentious host-microbiota relationship in inflammatory bowel disease—can foes become friends again? Scand. J. Gastroenterol. *50*, 34–42.

Schuppan, D., and Freitag, T. (2004). Fistulising Crohn's disease: MMPs gone awry. Gut *53*, 622–624.

Sim, W.H., Wagner, J., Cameron, D.J., Catto-Smith, A.G., Bishop, R.F., and Kirkwood, C.D. (2010). Novel Burkholderiales 23S rRNA genes identified in ileal biopsy samples from children: preliminary evidence that a subtype is associated with perianal Crohn's disease. J. Clin. Microbiol. *48*, 1939–1942.

Uniken Venema, W.T., Voskuil, M.D., Vila, A.V., van der Vries, G., Jansen, B.H., Jabri, B., Faber, K.N., Dijkstra, G., Xavier, R.J., Wijmenga, C., et al. (2019). Single-Cell RNA sequencing of blood and ileal T cells from patients with Crohn's disease reveals tissue-specific characteristics and drug targets. Gastroenterology *156*, 812–815 e822.

Verstockt, B., Smith, K.G., and Lee, J.C. (2018). Genome-wide association studies in Crohn's disease: past, present and future. Clin. Translational Immunol. *7*, e1001.

Watanabe, T., Asano, N., Meng, G., Yamashita, K., Arai, Y., Sakurai, T., Kudo, M., Fuss, I.J., Kitani, A., Shimosegawa, T., et al. (2014). NOD2 downregulates colonic inflammation by IRF4-mediated inhibition of K63-linked polyubiquitination of RICK and TRAF6. Mucosal Immunol. *7*, 1312–1325.

Wright, E.K., Kamm, M.A., Wagner, J., Teo, S.M., Cruz, P., Hamilton, A.L., Ritchie, K.J., Inouye, M., and Kirkwood, C.D. (2017). Microbial factors associated with postoperative Crohn's disease recurrence. J. Crohn's Colitis *11*, 191–203.

Wu, C.Q., Lin, Q.R., Ying, S.J., Luo, J.K., Hong, W.J., Lin, Z.J., and Jiang, Y. (2019). Association of Crohn's disease with aryl hydrocarbon receptor gene polymorphisms in patients from Southeast China. Immunol. Invest. *48*, 809–821.

Xu, W.D., Pan, H.F., Ye, D.Q., and Xu, Y. (2012). Targeting IRF4 in autoimmune diseases. Autoimmun. Rev. *11*, 918–924.

Yang, L., Tang, S., Baker, S.S., Arijs, I., Liu, W., Alkhouri, R., Lan, P., Baker, R.D., Tang, Z., Ji, G., et al. (2019). Difference in pathomechanism between Crohn's disease and ulcerative colitis revealed by colon transcriptome. Inflamm. Bowel Dis. *25*, 722–731.

Supplemental information

Relating the transcriptome and microbiome

by paired terminal ileal Crohn disease

Chenwen Cai, Sibo Zhu, Jinlu Tong, Tianrong Wang, Qi Feng, Yuqi Qiao, and Jun Shen

Supplementary Table 1. Characteristics of included patients, related to all figures.

	CD (n = 10)
Sex, n (%)	
Male	6 (60.0)
Female	4 (40.0)
Age, yrs, mean ± SD	30.9 ± 10.4
Montreal classification, n (%)	
Age at diagnosis	
A1	1 (10.0)
A2	7 (70.0)
A3	2 (20.0)
Location	
L1	8 (80.0)
L2	0 (0)
L3	2 (20.0)
L4	0 (0)
Behavior	
B1	7 (70.0)
B2	2 (20.0)
B3	1 (10.0)
With perianal lesion	4 (40.0)
Disease activity, n (%)	
Inactive (CDAI < 150)	6 (60.0)
Active (CDAI≥150)	4 (40.0)
CRP level	
CRP < 10mg/L	4 (40.0)
CRP ≥ 10mg/L	6 (60.0)

yrs, years; SD, standard deviation; CDAI, Crohn's disease activity index; CRP, C-reactive protein

TRANSPARENT METHODS

- KEY RESOURCES TABLE

- CONTACT FOR REAGENT AND RESOURCE SHARING

- SUBJECT DATAILS

- METHOD DETAILS

  O RNA extraction

  O RNA sequencing

  O 16s RNA sequencing and microbiome analysis

  O Transcriptome analysis

  O Analysis of host-microbiome cross-talk

  O Analysis of C-reactive protein (CRP)-associated genes and pathways

- QUANTIFICATION AND STATISTICAL ANALYSIS

- DATA AVAILABILITY

CONTACT FOR REAGENT AND RESOURCE SHARING

Reagents and resources included in the current study are available from the

corresponding author on reasonable request. All requests will be fulfilled by Jun Shen

(shenjun79@sina.cn) and Yuqi Qiao (qiaoyuqi@renji.com).

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE or Version
Materials	
RNA-later	Qiagen, Hilden, Germany
TruSeq stranded Total RNA kit	Illumina, San Diego, USA
Random hexamers	Qiagen, Hilden, Germany
KAPA HiFi HotStart ReadyMix	KAPA Biosystems, Waltham, USA
Nextera XT Index Kit v2	Illumina, San Diego, USA
Instrument	
Light microcopy	Olympus, Tokyo, Japan
Agarose gel electrophoresis	Thermo Fisher Scientific, Shanghai Branch, China
Illumina NovaSeq system	150 nucleotide paired end reads, Illumina
Software	
Megan	Version 6.20.17
R	Version 4.0.1
Galaxy online tool	https://huttenhower.sph.harvard.edu/galaxy/
Metascape	https://metascape.org/gp/ index.html
GraphPad Prism	Version 8.0.2
package	
EdgeR	Release (3.12)
Procomp	Version 4.0.1
PandaSeq	Version 2.11

Silva database	Release: 104
Fastqc	Version 0.11.9
Trimmomatic	Version 0.36
Hisat2	Version 2.0.5
FeatureCounts	Version 1.5.2
Pheatmap	Version 1.0.12
VennDiagram	Version 1.6.20

SUBJECT DATAILS

Between January 2019 and December 2019, new-onset CD patients from the

Inflammatory Bowel Disease Research Center, Renji Hospital, School of Medicine,

Shanghai Jiao Tong University were recruited in the study. Approximately 5 mg (four

mucosal biopsies) of tissue samples were extracted from the inflamed and proximal

non-inflamed ileum of CD patients (n=10). Biopsies were obtained during

endoscopies, and they were stored in RNA-later (Qiagen, Hilden, Germany) at 4 °C

overnight and at -80 °C for long-term storage. Characteristics of the included samples

are listed in Table S1. IMCD and NMCD were extracted from the biopsy samples,

and they were confirmed by light microcopy (Olympus, Tokyo, Japan). Three months

prior to recruitment, all patients were instructed to avoid antibiotics, steroids,

immunosuppressants, and biologics. The study was approved by the Institutional

Review Board of Renji Hospital, School of Medicine, Shanghai Jiao Tong University

(IRB no. 2018[221]). All patients provided written informed consent prior to sampling and data collection.

METHOD DETAILS

RNA extraction

Total RNA was extracted from cell lysates by grinding the tissues with an abrasive rod to augment cell disruption. RNA was prepared as instructed, and the integrity of RNA was assessed using agarose gel electrophoresis (Thermo Fisher Scientific, Shanghai Branch, China): Add 1/5 volume of Total RNA Extraction Reagent in chloroform, shake vigorously, and incubate on ice for 15min. Centrifuge at 12000xg for 15min at 4°C. Take the supernatant to a new tube, add an equal volume of ice bath isopropanol, mix upside down, -20°C for 20 min or more. Centrifuge 12000xg, 15min, 4°C. Remove supernatant, add an equal volume of 75% ice ethanol with Total RNA Extraction Reagent, and suspend the precipitate. Centrifuge 12000xg, 5min, 4°C. Remove supernatant, dry the precipitate moderately. Dissolve RNA precipitate with RNase-free water (50-100ul).

RNA sequencing

Strand-specific RNA sequencing libraries were prepared using 10 ng of total RNA from each sample using the TruSeq stranded Total RNA kit (Illumina, San Diego, USA), and they were sequenced on an Illumina NovaSeq system (150 nucleotide paired end reads, Illumina). Quality control and read alignment and processing were

conducted using the package EdgeR and 'procomp' was performed with R software version 4.0.1. (https://cran.r-project.org/). The data of RNA sequencing was listed in Supplementary File 1.

16s RNA sequencing and microbiome analysis

Transcriptionally active ileal mucosa-attached bacterial profiles were generated using the RNA obtained from mucosal biopsy (converted to cDNA using random hexamers, Qiagen). The V3-V4 variable region of 16S rRNA was amplified using bacterial 16S rRNA gene-specific composite primers (337F and 805R) using the KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Waltham, USA). Libraries were prepared using KAPA HiFi HotStart ReadyMix and Nextera XT Index Kit v2 (Illumina). The following 16S amplicon PCR primers were used: F5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWG CAG;

R5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTA TCTAATCC. Pooled amplicon libraries were sequenced using the Illumina NovaSeq system (2×150 bp), and they were processed using the PandaSeq and Malt algorithms. The 16S double-end data were spliced using PandaSeq (V2.11) and then compared with the silva database (release: 104) using malt (V0.4.1). The compared daa files were Taxonomy annotated using Megan (V6.20.17) and OUT abundance quantification was performed. Species abundance was analyzed using megan, and the top 10 OUTs were plotted as histograms using R. Alpha diversity, PCoA were also

analyzed using Megan. lefse analysis was done using Galaxy online tool (https://huttenhower.sph.harvard.edu/galaxy/). The data of 16s RNA sequencing was listed in Supplementary File 2 and Supplementary File 3.

Transcriptome analysis

Nextra data were quality controlled (QC) by fastqc (v0.11.9). Low quality reads and Nextra junctions were removed after QC using Trimmomatic (v0.36). And the Hg19 reference genome was compared using Hisat2 (v2.0.5). The raw expression matrix was obtained by quantification using FeatureCounts (v1.5.2). The raw expression matrix was corrected for depth by count per million (CPM) and logarithmized. PCA analysis was performed using the R package 'prcomp' function. Pairwise or unpaired screening for differential genes (FC>1.5, < 0.67 and P< 0.05) was performed using student t. The top 10 genes with the largest fold change up- and down-regulation were heat mapped using 'Pheatmap' (V1.0.12). Functional enrichment analysis was conducted on Metascape (https://metascape.org/gp/ index.html) using KEGG and GOBP, and bubble maps were plotted by R. VennDiagram (V1.6.20) was used to plot venn diagrams.

Analysis of host-microbiome cross-talk

To quantitatively measure host-microbiome cross-talk, Spearman's rank correlation coefficient was calculated for all DEGs and OTUs present in the samples between the respective gene expression levels and the respective OTU abundance. To determine

the statistical significance of individual correlation coefficients, false discovery rate was estimated using the Benjamini-Hochberg test.

In the host-microbiome cross-talk map, the microbes and top 20 genes related to TF were selected, and they were compared between IMCD and NMCD. A paired t-test was applied, and $p < 0.05$ was considered statistically significant. Six samples obtained from three patients, who consumed antibiotics in the past three months, were excluded.

The list of human transcription factors (TFs) was downloaded from the HumanTFdb database and intersected with the DEGs (http://bioinfo.life.hust.edu.cn/HumanTFDB#!). The spearman correlations of transcription factors in DEGs with OTU and immune gene (download from https://www.nanostring.com/) were calculated separately using the 'corr.test' function in Psych package, and the P-values were corrected using BH. The correlation results were filtered according to BH<0.05, and the OUT-TF-immune gene co-expression network was constructed by using TF as a mediator.

Analysis of C-reactive protein (CRP)-associated genes and pathways

We used a dichotomous approach to determine the impact of disease activity on DEGs. Patients were divided into two groups: CRP low (CRPL) ($< 10mg/L$) and CRP high (CRPH) ($\geq 10mg/L$). A Venn diagram was used to identify different comparisons such as IMCD/NMCD and CRPH/CRPL and to analyze overlapping and

unique features.

QUANTIFICATION AND STATISTICAL ANALYSIS

Pairwise or unpaired screening for differential genes (FC>1.5, < 0.67 and P< 0.05) was performed using student t. The statistical analysis was carried out using GraphPad Prism software Version 8.0.2 (San Diego, CA, USA).

DATA AVAILABILITY

All data and code supporting the findings of this study are available within the paper and its supplemental information files (Data S1). Our RNA-Seq data has been uploaded to the GEO database (GSE171244).