Development of Prognostic Marker using Spatial Total RNA Sequencing based Multimodal Validation in Pediatric Crohn Disease

Sooyoung Jang

The Graduate School
Yonsei University
Department of Medicine

Development of Prognostic Marker using Spatial Total RNA Sequencing based Multimodal Validation in Pediatric Crohn Disease

A Dissertation Submitted
to the Department of Medicine
and the Graduate School of Yonsei University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Medical Science

Sooyoung Jang

December 2024

This certifies that the Dissertation of Sooyoung Jang is approved

Thesis Supervisor	Yu Rang Park
Thesis Committee Member	Hong Koh
Thesis Committee Member	Younghee Lee
Thesis Committee Member	Sangwoo Kim
Thesis Committee Member	Ju Yeong Kim

The Graduate School
Yonsei University
December 2024

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Yu Rang Park, for her consistent support and guidance during the course of this project. I am especially grateful for the unwavering support and advice from Professor Hong Koh. I am also grateful to Professor Sowon Park, Eun Joo Lee, and Hyeji Lim in the Department of Pediatrics, as well as MinDong Sung in the Department of Pulmonary Medicine, for their help with specimen collection and advice. And I would also like to express my appreciation to Bo Kyu Choi, Chan Young Ko, Jong Hyun Kim, JooHyun Lee, Min Kyoon Yoo and Jae Seong Hong for their invaluable assistance with the analysis. I am also grateful to the members of DHLab and MCU for their research-related feedback and emotional support, as well as to my college friends. I want to extend my heartfelt thanks to my beloved wife Minji Seo. Lastly, I would like to thank my father and my mother for raising me with love and always supporting me, and my brother. Thank you all.

This research was made possible through a grant from the MD-PhD/Medical Scientist Training Program, funded by the Korea Health Industry Development Institute (KHIDI) and the Ministry of Health & Welfare of the Republic of Korea.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Development of Prognostic Marker using Spatial Total RNA Sequencing based Multimodal Validation in Pediatric Crohn Disease

Crohn's disease is a chronic inflammatory bowel disease that is considered to be caused by the interaction between the gut microbiome and the host immune system. Numerous studies have characterized gut microbiome composition in CD patients. Advancements in high-throughput transcriptomics approaches have enhanced understanding of cellular heterogeneity in various disease contexts. The complex relationship between tissue-resident bacteria and the host immune network in CD pathogenesis remains unclear. Previous spatial transcriptomics methods were limited in identifying bacterial RNA. Here we show a novel spatial host-microbiome profiling approach that utilizes in situ poly adenylation with spatial transcriptomics for simultaneous detection of host and bacterial RNA. We developed a pipeline to detect bacteria in spatial transcriptomics data at high taxonomic resolution, including strain-level identification. Using this approach, we demonstrate that bacterial infiltration is significantly elevated in CD tissues and correlates with disease prognosis. This approach identified beneficial and pathogenic microbiome associated with CD. These findings reveal interactions between host cells and bacteria in CD, providing insights into CD pathogenesis at cellular resolution. This study offers a new approach for investigating host-microbiome interactions in various microbiome-associated diseases, potentially leading to new strategies for microbiome-based therapeutics and prognostic markers.

Key words: Gastrointestinal Microbiome; Multiomics; Bacterial translocation; Host-microbiome interactions

1. INTRODUCTION

The etiology of Crohn's disease (CD) is not clearly understood, but it is considered to be caused by an interaction between the gut microbiome and the aberrant host immune response in patients with a genetic predisposition (Neurath 2019). Several findings have emphasized the importance of bacteria in the pathogenesis of CD. Genome-wide association studies (GWAS) have reported that defects in microbe sensing, epithelial barrier function, microbicidal mechanisms, cytokine regulation, and adaptive immunity are associated with CD (Graham and Xavier 2020). Furthermore, mouse models of CD do not develop the disease when raised in germ-free conditions, but they develop CD in the presence of bacteria (Balish and Warner 2002; Kim et al. 2005). Recently, Ha et al. have reported that there is a subset of mucosa-associated gut bacteria that translocate to the intestinal tissue in CD, and these bacteria may contribute to the progression of the disease (Ha et al. 2020).

Consequently, numerous large-cohort studies have been conducted to characterize the gut microbiome composition in patients with CD (Wright et al. 2015; Hansen et al. 2012; Gevers et al. 2014; Kansal et al. 2019; Haberman et al. 2014). A notable example is the study by Gevers et al., who conducted microbiome profiling using ileal tissue samples from 447 pediatric patients with CD and 221 controls (Gevers et al. 2014). These studies suggested that the gut microbiome in CD is characterized by an increase in pathogenic bacteria such as Cutibacterium acnes and Haemophilus parainfluenzae, and a decrease in beneficial bacteria like Faecalibacterium prausnitzii (Kansal et al. 2019; Hansen et al. 2012; Gevers et al. 2014).

The identification of these alterations, particularly the reduction of beneficial bacteria, holds significant potential for the development of microbiome-based therapeutics in CD (Sorbara and Pamer 2022; Mohebali et al. 2023). The underlying mechanisms of microbiome-mediated effects are not fully understood, but evidence suggests that certain gut bacteria, particularly Clostridium clusters XIVa and IV, promote the accumulation of regulatory T (T_{reg}) cells (Sefik et al. 2015; Ohnmacht et al. 2015; Atarashi et al. 2011; 2013). These accumulation of T_{reg} cells in the intestinal mucosa subsequently regulate excessive immune responses in inflammatory bowel disease.

Recent advancements in high-throughput transcriptomic approaches, such as spatial transcriptomics, have revolutionized our understanding of cellular heterogeneity and spatial organization in complex biological systems (Tian, Chen, and Macosko 2023). However, previously developed spatial transcriptomics methods relied on poly A tail-targeting approaches, limiting their

application to mRNA from eukaryotic organisms, such as humans and mice (Ståhl et al. 2016). These methods faced challenges in identifying non-host RNA, particularly bacterial RNA, which lacks poly A tails (McKellar et al. 2022). This limitation has hindered a comprehensive understanding of host-microbiome interactions and their potential roles in the pathogenesis of microbiome-associated diseases, such as CD. Nevertheless, Galeano et al. recently demonstrated that bacteria within tumors may play a crucial role in cancer metastasis using a spatial transcriptomics method targeting bacterial 16S rRNA (Galeano Niño et al. 2022). However, these 16S rRNA-targeting approaches have limited taxonomic resolution when profiling the microbiome, restricting the identification of bacteria to the genus level, thereby hindering species-level discrimination (Galeano Niño et al. 2022; Saarenpää et al. 2023; Lötstedt et al. 2023).

In this study, we aim to investigate the interactions between bacteria and host cells at cellular resolution in pediatric CD tissue samples using recently developed spatial total RNA sequencing, which utilizes yeast poly A polymerase to simultaneously detect both host and non-host RNAs. We also assess the prognostic value of tissue microbiome analysis in CD and explore the impact of tissue-resident bacteria on the immune response within CD tissues. Furthermore, we aim to identify pathogenic and beneficial microbiomes associated with CD at the cellular level and investigate the alteration of the host transcriptome in response to these microbiomes. Our study incorporates deep learning approaches of histology image to detect tissue morphology changes associated with bacterial infiltration in CD. Additionally, we have developed a refined decontamination process using semi-supervised learning methods to identify bacterial presence without the need for metagenomic shotgun sequencing validation.

2. MATERIALS AND METHODS

2.1. Participant and ethics

This prospective study included patients with CD from the division of Gastroenterology, Hepatology, and Nutrition, Department of Pediatrics, Sinchon Severance Hospital, in Seoul. This tertiary teaching hospital diagnoses approximately 25 new CD cases annually. Ethical approval was obtained from the Institutional Review Board of Severance Hospital, Yonsei University College of Medicine (IRB number: 4-2022-1127).

We recruited pediatric patients aged between 7 and 18 years between Nov 01, 2022, and Apr 30,

2024. For the CD group, we included patients newly diagnosed with CD. The control group consisted of patients with irritable bowel syndrome who underwent endoscopy but showed no evidence of inflammation on endoscopic examination, histopathological analysis of biopsy specimens, and blood tests. For all participants, both the patients and their parents or legal guardians provided informed consent for participation in the study. Clinical relapse of pediatric CD was defined as a Pediatric Crohn's Disease Activity Index (PCDAI) score of ≥30, indicating a worsening of the disease (Hyams et al. 1991).

2.2. Sample collection and storage

Tissue samples were collected from the terminal ileum during endoscopic examination. For pediatric CD patients, both inflamed and non-inflamed tissues were obtained, while control group samples consisted of only non-inflamed tissue. Immediately after collection, these tissue samples were embedded in optimum cutting temperature (OCT) compound (SciGen Scientific, Gardena, CA, USA), frozen and stored at -80°C. The storage periods ranged from 1 week to 3 months. The stored samples were subsequently process for spatial total RNA sequencing and bulk shotgun metagenome sequencing. The quality of RNA extracted from the samples was assessed using the RNA Integrity Number (RIN), with samples falling below the quality threshold being excluded from further analysis.

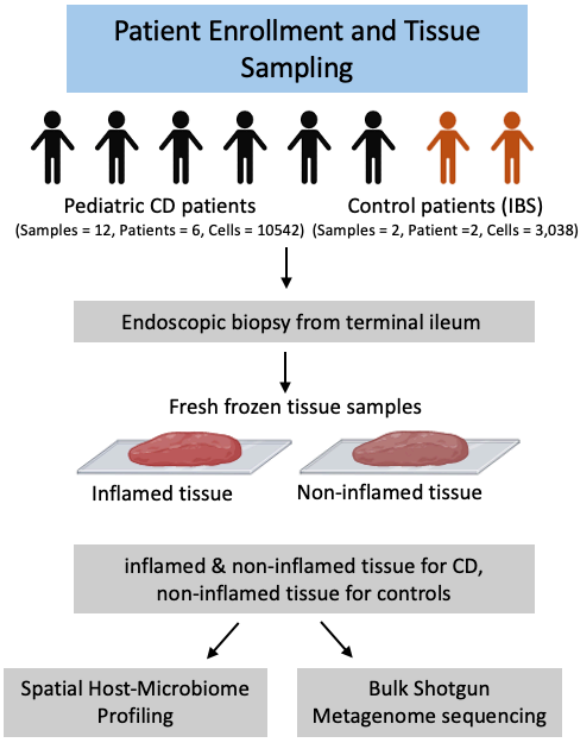

Figure 1. Schematic diagram of the overall study design

2.3. Spatial total RNA sequencing

The quality of tissue blocks was evaluated by isolating RNA from consecutive tissue sections and measuring the RNA integrity number (RIN) using an Agilent 4200 TapeStation system (Agilent Technologies, Santa Clara, CA, USA). Samples with a RIN $\geq$ 4 were considered suitable for further experiment. Due to the small size of the tissues, multiple sections were placed on each sample block for analysis. For CD group, both inflamed and non-inflamed tissue sections from the same patient were mounted side by side, while only non-inflamed tissue sections were mounted for control group. Following the Visium protocol (CG000160, Rev D, 10x Genomics), the tissue samples underwent methanol fixation and H&E staining. The stained sections were imaged using a Nikon Eclipse Ti2 microscope.

Based on the approach of McKellar et al., we conducted in situ polyadenylation (McKellar et al. 2022). This procedure included equilibration with wash buffer containing Protector RNase Inhibitor (Roche, catalog no. 3335402001), followed by incubation with an enzyme mix containing yeast poly A polymerase (Thermo Scientific, catalog no. 74225Z25KU). We used the Visium Tissue Optimization Kit from 10x Genomics to determine the optimal tissue permeabilization time. After the in situ polyadenylation step, we prepared final sequencing libraries using standard Visium library preparation protocol (CG000239, Rev F, 10x Genomics). The prepared libraries were combined and sequenced on a NovaSeq 6000 system (Illumina, San Diego, CA, USA) using the NovaSeq 6000 S1 Reagent Kit v1.5 (200 cycles, 20028318, Illumina). We sequenced to a depth of approximately 120M reads per sample.

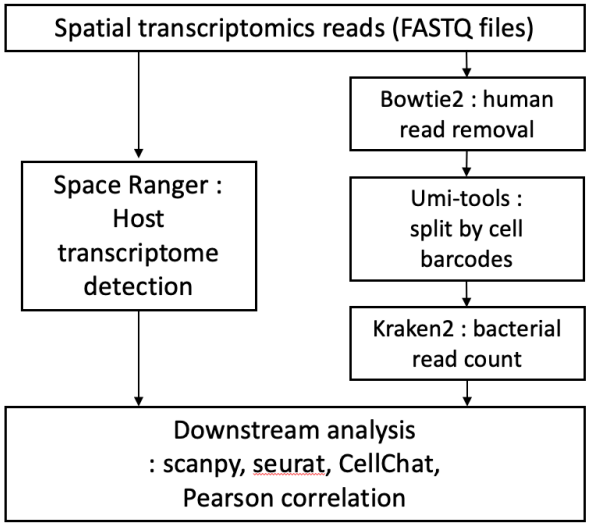

Figure 2. Schematic diagram of the spatial host-microbiomeo profiling

2.4. Bulk shotgun metagenome sequencing

Genomic DNA quality and quantity were assessed via fluorometry (Qubit, Invitrogen) and gel electrophoresis. Samples with a DNA Integrity Number ≥6, as measured by an Agilent 4200 TapeStation, were used for further analysis. The library was prepared using the Illumina TruSeq

Nano DNA Library Prep Kit. For each sample, 100 ng of DNA was fragmented to 350 bp using a Qsonica 800 R2, followed by Illumina adapter ligation and PCR amplification. Libraries sized 500–600 bp were selected and quantified using TapeStation 4200 (Agilent Technologies) and KAPA Library Quantification Kit (Kapa Biosystems). Sequencing was performed on an Illumina NovaSeq 6000 platform using 150 bp paired-end reads.sample.

2.5. Host transcriptome analysis

Host transcriptomics data were preprocessed using Space Ranger v1.3.1 (10x Genomics). Downstream analyses with the output count matrices were performed using Scanpy v1.9.3 (Wolf, Angerer, and Theis 2018). As the Visium method captures multiple cells within each spot, we employed Cell2location, a computational tool that maps single-cell RNA sequencing data onto spatial transcriptomics data (Kleshchevnikov et al. 2022). For this process, we utilized previous single-cell RNA sequencing data from pediatric patients with CD (Elmentaite et al. 2020). Cell2location integrates this single-cell data with our spatial transcriptomics data, allowing us to identify areas enriched with specific cell types across the analyzed tissue samples. Low-quality spots with >40% mitochondrial reads were excluded from the analysis (Elmentaite et al. 2021).

2.6. Spatial microbiome profiling

We used Bowtie2 v2.5.1 to align the spatial transcriptomics read to the GRCh38 human reference genome for human read removal (Langmead and Salzberg 2012). UMI-tools v1.1.1 was used to separate the fastq files by cell barcode for independent analysis per spots (Smith, Heger, and Sudbery 2017). Bacterial reads were identified using Kraken2 v2.1.1 with the publicly available PlusPF database (version 2022/09/08, https://benlangmead.github.io/aws-indexes/k2) (Wood, Lu, and Langmead 2019). It contains RefSeq reference sequences for archaea, bacteria, viruses, plasmids, protozoa, fungi, and humans, as well as UniVec_Core sequences. We calculated counts per million (CPM) mapped reads by dividing bacterial read counts by total sequencing reads for normalization.

2.7. Microbiome profiling in bulk shotgun metagenome data

In the analysis of bulk shotgun metagenome data, we employed a process similar to our spatial microbiome detection method. Bowtie2 aligned reads to the GRCh38 human reference genome for

human read removal (Langmead and Salzberg 2012). Subsequently, we identified bacterial reads using Kraken2 with the same publicly available PlusPF database used in our spatial analysis (Wood, Lu, and Langmead 2019).

2.8. Spatial microbiome decontamination process

We developed a stepwise decontamination process to reduce false positive species in our spatial microbiome profiling data. Initially, we excluded species with low spatial microbiome sequencing read counts across all samples, specifically those with <50 total reads across all spatial microbiome sequencing samples. We then compared the profiling results with the bulk metagenome sequencing data from the same tissue samples. Species with <50 reads from bulk sequencing or with >1.5 ratio of spatial and bulk sequencing read percentages were removed from our spatial microbiome dataset.

To mitigate potential false positives arising from host read contamination, we evaluated the results of single and double human read-removal processes (Garrido-Sanz, Senar, and Piñol 2022). The single human read-removal process used Bowtie2, while the double human read-removal process used both Bowtie2 and BWA v0.7.17; all aligned with the GRCh38 human reference genome (Langmead and Salzberg 2012; Li and Durbin 2009). This process led to two observations. First, certain species exhibited a drastic reduction in read counts following the double human read removal, indicating that they might be false positive; second, the read counts from genuine gut-residing species also decreased moderately, suggesting that certain true signals were affected (Figure 4). We classified the species as potential false positives if their read counts after double removal dropped to <2% of those observed after the single removal.

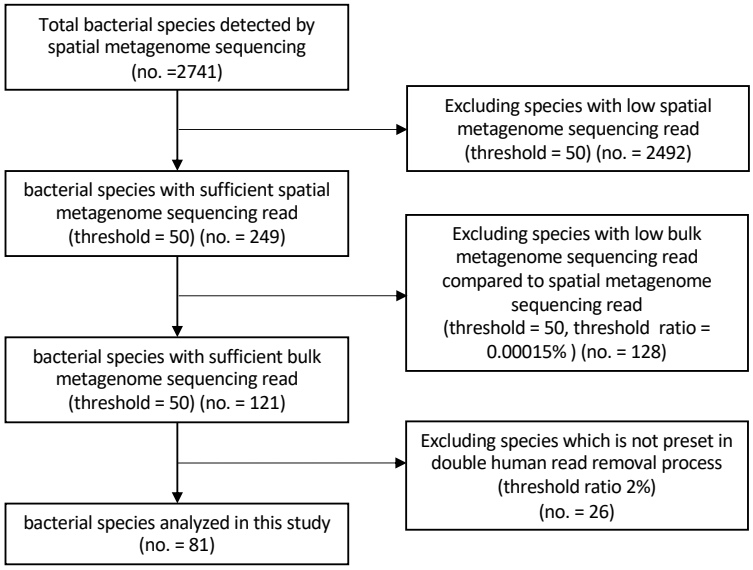

Figure 3. Determination of bacterial species for analysis in the pediatric Crohn's disease spatial microbiome study

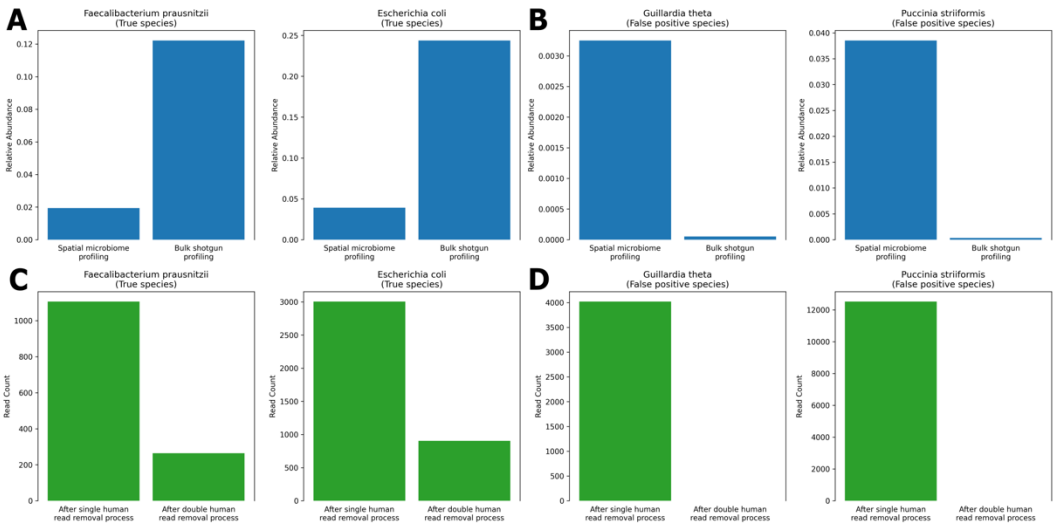

Figure 4. Comparison of Spatial Microbiome Profiling and Bulk Shotgun Sequencing with Human Read Removal Approaches for Accurate Bacterial Identification

2.9. Comparative analysis of bacterial infiltration

The average number of bacterial reads, in CPM, per spot was compared across the control, CD non-inflamed, and CD inflamed tissues. The number of cells with bacterial infiltration was assessed for each cell type. For each patient, we compared the average number of bacterial reads per spot with their time to relapse and the Rutgeerts score of ileal endoscopic findings (Rutgeerts et al. 1990).

2.10. Differential gene expression and gene set enrichment analysis

Differentially expressed genes were identified based on the presence of specific bacterial species. This analysis was conducted separately for each detected bacteria, and we assessed the consistency of gene expression changes across different bacterial species. Each cell type previously identified by Cell2location was analyzed, enabling cell type-specific characterization of host-microbiome interactions. Gene set enrichment analysis was performed using the BioPlanet 2019 database to identify biological pathways associated with genes consistently expressed differentially in the presence of bacteria (Huang et al. 2019). Additionally, we conducted an analysis of tissue-specific gene expression between cells exposed to beneficial and pathogenic microbiomes.

2.11. Quantification of microbial effects on cell viability

We investigated the impact of specific bacterial species on cell viability using the percentage of mitochondrial reads, a widely used marker of cell viability in single-cell RNA sequencing and spatial transcriptomics analyses (Ilicic et al. 2016). Cells with ≤10% mitochondrial reads were defined as viable, while those with >10% were defined as damaged. We calculated the relative risk (RR) of reduced cell viability following exposure to specific bacterial species to quantify the effect of bacterial exposure on cellular integrity. This analysis was performed exclusively on samples from participants with CD to examine the impact of bacterial exposure. Multiple testing correction was performed using the Benjamini–Hochberg method to control the false discovery rate at 0.05. The RR was calculated as follows:

$$RR = \frac{Number\ of\ damaged\ cells\ exposed\ to\ bacteria\ /\ Total\ number\ of\ cells\ exposed\ to\ bacteria}{Number\ of\ damaged\ cells\ not\ exposed\ to\ bacteria\ /\ Total\ number\ of\ cells\ not\ exposed\ to\ bacteria}$$

We also calculated Population Attributable Risk Percent (PARP) by integrating the RR values of specific bacteria with their tissue prevalence. This provided a measure of the overall contribution of

each bacterial species to cellular damage within the entire tissue environment. The PARP was calculated using the following formula, where P is the proportion of cells exposed to the specific bacteria:

$$PARP\ (\%) = \left(\frac{P\times(RR-1)}{P\times(RR-1)+1}\right)\times 100$$

2.12 Coexistence of bacteria across cell types

To analyzea bacterial distribution patterns, we assessed the spatial location of specific bacteria across spots using dimension reduction. We performed Uniform Manifold Approximation and Projection (UMAP) analysis on the spatial presence information of bacteria, compressing it into two-dimensional data for visualization. For each spot, we examined which cell types coexisted with specific bacteria using Cell2location analysis. Additionally, for each spot, we integrated the cell type proportions estimated by Cell2location with bacterial presence data to determine the frequency of coexistence between specific bacterial species and various cell types.

2.13. Correlation analysis of microbiome and host transcriptome

We performed Pearson correlation analysis to evaluate the relationship between the presence of each bacterial species and host gene expression levels. This analysis was performed exclusively on samples from participants with CD. Principal Component Analysis (PCA) was then applied to the resulting correlation matrix for dimension reduction. We analyzed the contribution of individual genes to principal component 2 (PC2) and displayed the top 20 genes with the highest absolute contribution.

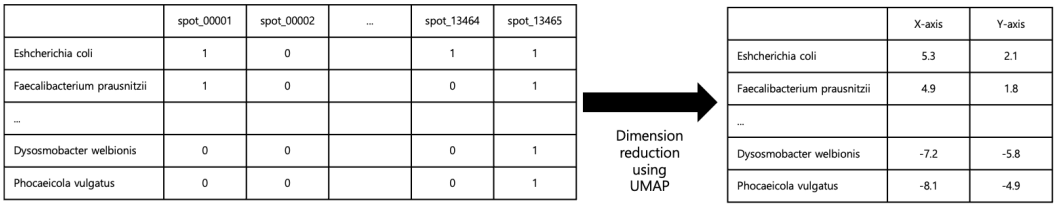

Figure 5. Schema of dimension reduction for analysis of bacteria coexistence

2.14. Deep Learning Analysis of Bacterial Translocation

We developed deep learning models to detect bacterial translocation in high-resolution H&E-stained images of the same tissue sections used for spatial transcriptomics. We trained separate models to identify the presence of all bacteria, beneficial microbiomes, pathogenic microbiomes, and individual bacterial species. To prepare the data, we divided the images into 224x224 pixel patches centered on barcode locations, selecting only those containing over 60% tissue. We fine-tuned pre-trained DenseNet121 models using an Adam optimizer (learning rate 0.001) for up to 25 epochs.

To further analyze the relationship between cell morphology and bacterial infiltration, we clustered the 224x224 pixel patches from the H&E-stained images into five groups based on their morphological characteristics. Pearson correlation analysis was then used to evaluate the association between these morphological clusters and the presence of various bacterial types.

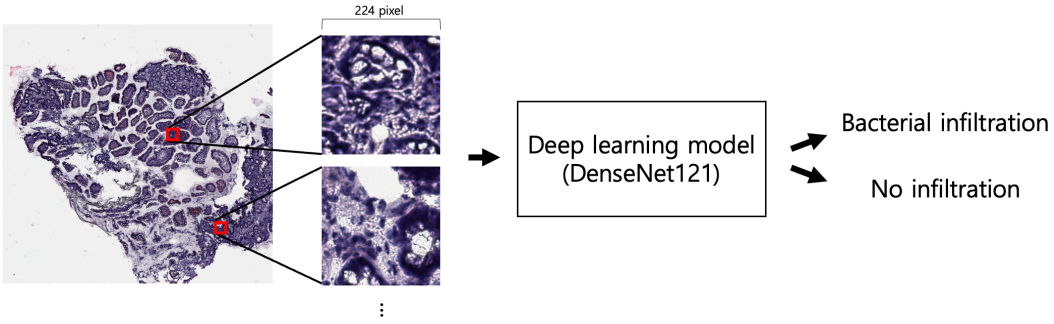

Figure 6. Schema of Deep learning for bacterial translocation induced tissue changes

2.15. Advanced decontamination method using semi-supervised learning model

When performing sequencing on specimens with mixed human and bacterial content, misidentification of human DNA as bacterial DNA produce substantial false positive species.(Gihawi et al. 2023; Gihawi, Cooper, and Brewer 2023) While our study validated the presence of bacterial species through metagenomic shotgun sequencing of identical specimens, such validation may not be feasible for all reasearch settings. We hypothesized that true bacterial species and contamination species would differentially affect host gene expression patterns. Using deep

learning models, we evaluated whether Pearson correlation values between bacterial presence and host gene expression changes, which represent host-bacterial interactions, could distinguish true bacterial species from contaminants.

Considering situations where bulk microbiome validation is not available, we classified species into definitive bacterial candidates, definitive false contaminant candidates, and unlabeled species by only comparing single and double human read removal processes. We employed the ContrastiveMixup model, a semi-supervised deep learning approach that utilizes both labeled and unlabeled data to improve classification performance.(Darabi et al. 2021) This model enhances the learning of labeled data through unlabeled data while simultaneously predicting labels for unlabeled data.

3. RESULTS

3.1. Spatial Microbiome Profiling Reveals Bacterial Translocation Patterns

A total of 14 terminal ileal tissue samples were biopsied, including 12 from inflamed and non-inflamed tissues of six children with pediatric CD (mean age 13.5 years, SD 2.1) and two from non-inflamed tissues of two children without CD as the control group (mean age 15.0 years, SD 1.0). Spatial host-microbiome sequencing and bulk shotgun metagenome sequencing were performed on all collected tissues. For the CD group, tissues were collected at the time of diagnosis.

Table 1. Patient characteristics of this study

	Crohn's disease	Control (Irritable bowel syndrome)
number of tissues	12	2
number of patients	6	2
PCDAI score	34.7(4.5)	-
Age	13.5(2.1)	15.0(1.0)
Gender		
male	5	0
female	1	2

initial CRP		21.1(27.8)	3.9(3.7)
Location, n (%)			-
	L1 terminal ileum	0(0%)	-
	L2 colon	0(0%)	
	L3 ileocolon	6(100%)	-
Behavior, n(%)			-
	B1 inflammatory	6(100%)	
	B2 stricturing	0(0%)	-
	B3 penetrating	0(0%)	-
perianal involvement, n (%)		4(66%)	-

* Abbreviations: PCDAI: Pediatric Crohn's Disease Activity Index; CRP: C-reactive protein

To capture both host RNA and bacterial RNA simultaneously, in-situ polyadenylation was performed using yeast poly A polymerase prior to host spatial transcriptome sequencing, following the method of McKellar et al (McKellar et al. 2022). Spatial microbiome profiling was conducted by removing human reads using Bowtie2, followed by the identification of bacterial reads using Kraken2 (Langmead and Salzberg 2012; Wood, Lu, and Langmead 2019). Additionally, bacterial species that were nearly absent in the bulk metagenome shotgun sequencing performed on the same tissue were excluded from the spatial microbiome profiling results. Furthermore, bacterial species that were markedly reduced after additional human read removal using BWA were also excluded, as they were considered potential human read contaminants (Figure 1) (Li and Durbin 2009). As a result, 81 species of bacteria were identified, which are commonly found in the gut.

Cell2location was used to map the spatial distribution of cell types in our spatial transcriptomics data by integrating previously published single-cell RNA sequencing data from pediatric patients with CD (Kleshchevnikov et al. 2022; Elmentaite et al. 2020). This approach identified various areas enriched with certain cell-types, such as enterocyte-rich, immune cell-rich, and Tuft cell-rich areas (Figure 7). Considering the samples from the same participant, the immune cell-rich areas were found to be more abundant in the inflamed CD tissues compared to the non-inflamed tissues (Figure 8).

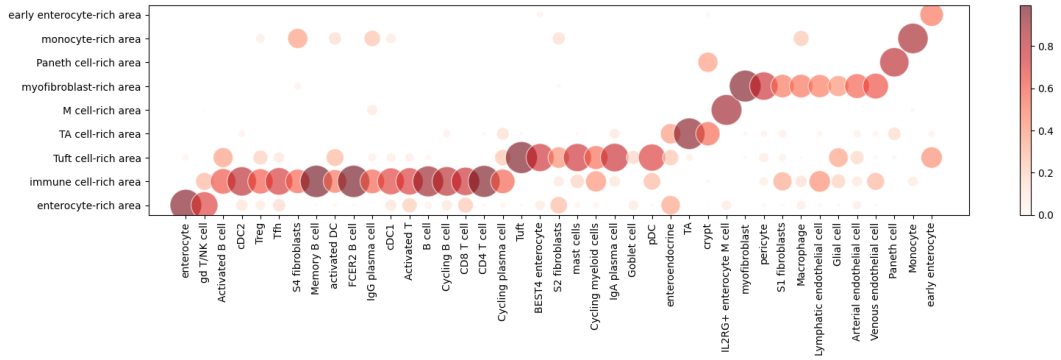

Figure 7. Cell type annotation using Cell2location. Cell type deconvolution was performed using Cell2location. Based on the results, areas were annotated according to co-occurring cell type combinations, such as immune cell-rich area, enterocyte-rich area, Tuft cell-rich area, and Paneth cell-rich area. The composition of cell types for each area is displayed.

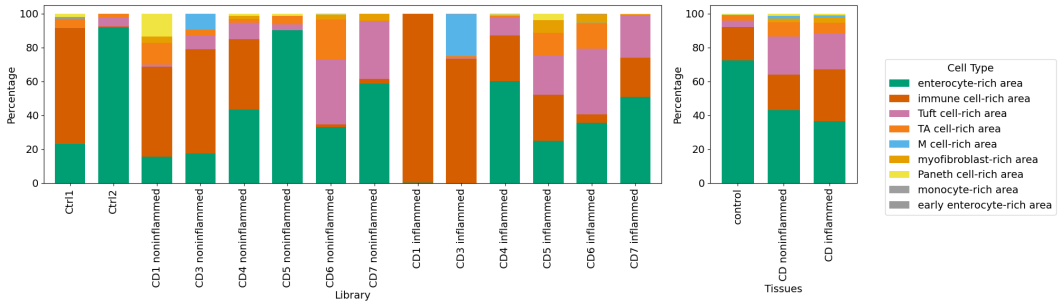

Figure 8. Stacked Percentage Bar Chart of Cell Types by sample and tissue type. a. Stacked Percentage Bar Chart of Cell Types by sample. b. Stacked Percentage Bar Chart of Cell Types by sample. CD: Crohn's disease; Ctrl: Control

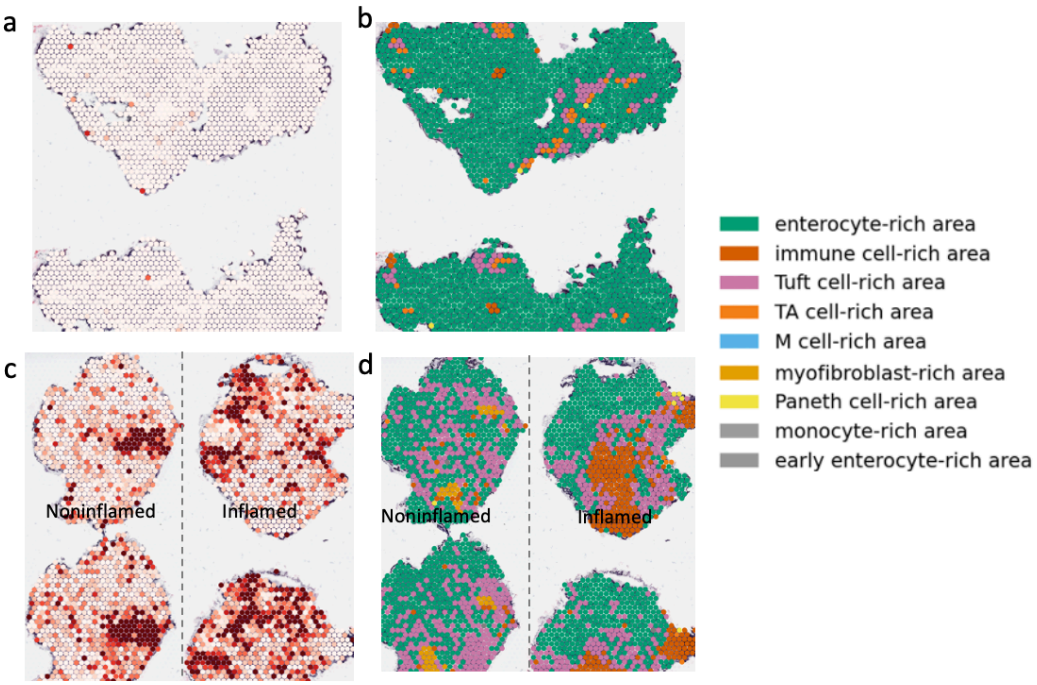

Figure 9. Distribution of total bacterial read counts between tissue types. Distribution of total bacterial read counts in contol group tissue (Ctrl 2). b. Distribution of cell types in contol group tissue (Ctrl 2). c. Distribution of total bacterial read counts in CD patient tissue (CD4). noninflammed tissue(left), inflammed tissue(right). d. Distribution of cell types in CD patient tissue (CD4). CD: Crohn's disease; Ctrl: Control

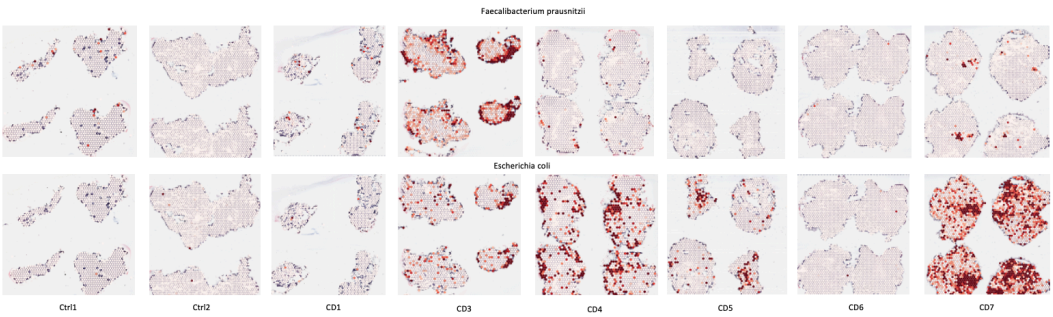

Figure 10. Distribution of various bacteria in various tissue types. Faecalibacterium prausnitzii and Escherichia coli infiltration patterns in control (Ctrl 1-2) and Crohn's disease (CD 1-7) tissues.

CD: Crohn's disease; Ctrl: Control

Furthermore, the bacterial read count per cell was higher in both non-inflamed (0.409 ± 0.028, mean ± standard error) and inflamed CD tissues (0.892 ± 0.115) compared with that in the controls (0.011 ± 0.002), with the highest count observed in inflamed CD tissues (Figure 11a). Notably, we observed increased counts in M cell-rich areas (8.660 ± 2.516) (Figure 11b), consistent with the known function of M cells in sampling luminal bacteria (Dillon and Lo 2019).

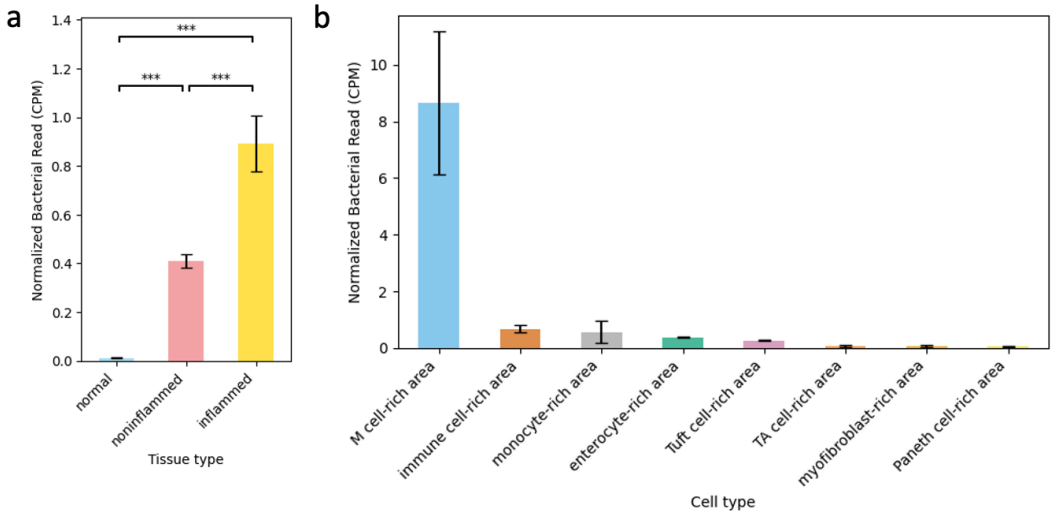

Figure 11. Comparison of total bacterial read counts between tissue types a. The average number of total bacterial reads per cells in control, CD non-inflamed, and CD inflamed tissue were compared. b. The average number of total bacterial reads per cell were compared between each cell types. P-values indicate statistical significance (*P < 0.05 and **P < 0.005, ***P < 0.0005)

3.2. Bacterial Translocation Predicts Crohn's Disease Prognosis

The bacterial read count per cell was highest in the CD group with relapse, followed by the CD group without relapse, and lowest in the control group. Within the relapse group, the time to relapse was shorter for participants who had a higher count at the time of diagnosis (Figure 12a). We also observed an association between bacterial read counts and the severity of endoscopic findings in the ileum, which was assessed by the Rutgeerts score (Figure 12b) (Rutgeerts et al. 1990). These

findings suggest the potential for predicting CD prognosis by assessing the extent of bacterial translocation in intestinal tissues.

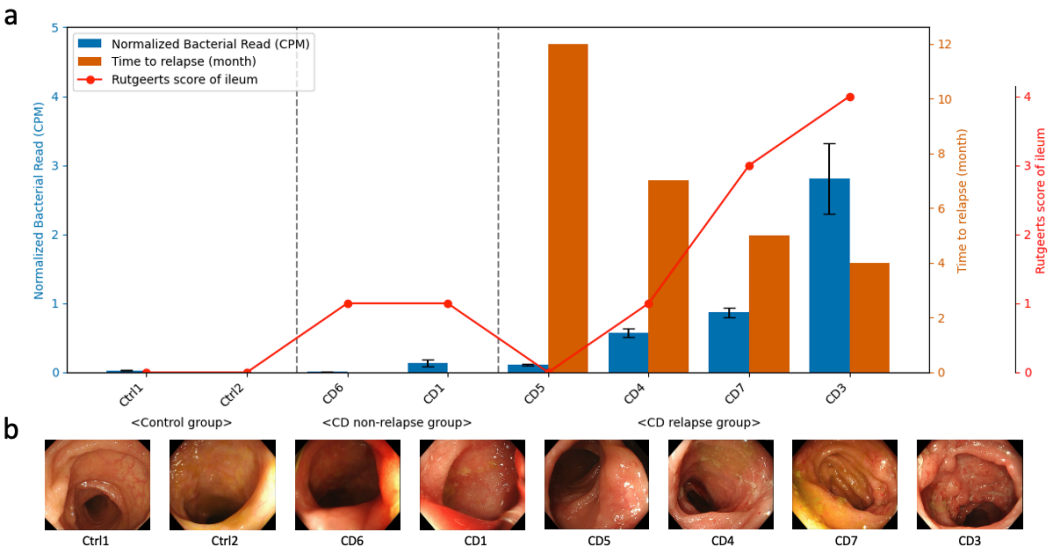

Figure 12. Comparison of bacterial infiltration, time to relapse, and endoscopic severity in CD. a. The graph displays normalized bacterial read counts (blue bars), time to relapse (orange bars), and Rutgeerts scores (red line) across control, non-relapse CD, and relapse CD groups. b. Endoscopic images of ileum for each participants, showing varying degree of mucosal inflammation corresponding to their group classification.

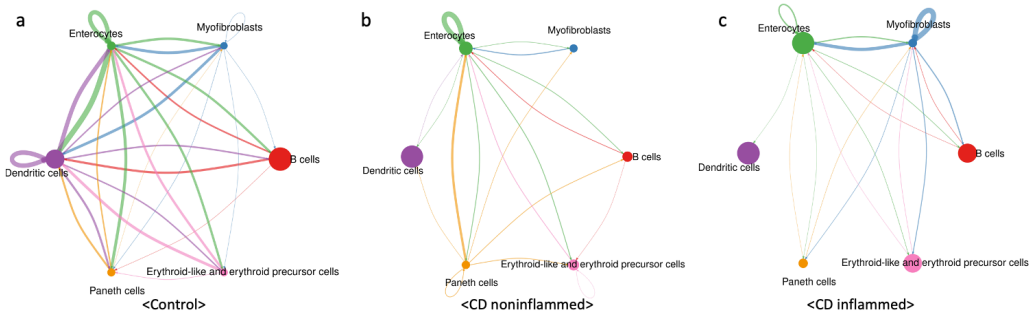

Figure 13. Cell-cell interaction analysis between tissue. Cell-cell interaction were analyzed using

CellChat (Jin et al. 2021). a. Cell-cell interaction analysis in control tissue. b. Cell-cell interaction analysis in CD noninflammed tissue. c. Cell-cell interaction analysis in CD inflammed tissue.

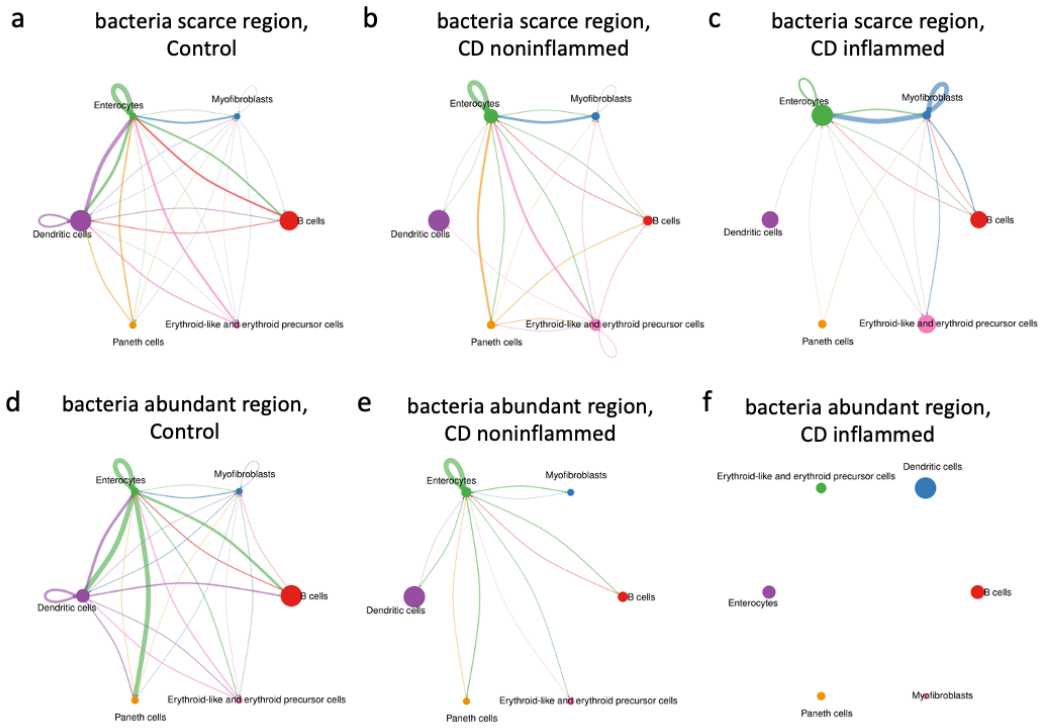

Figure 14. Cell-cell interaction analysis between tissue in bacteria-scarce region and bacteria-abundant region. Cell-cell interactions were analyzed using CellChat (Jin et al. 2021). a. Cell-cell interaction analysis of control tissue in bacteria scarce region. b. Cell-cell interaction analysis of CD noninflammed tissue in bacteria scarce region. c. Cell-cell interaction analysis of CD inflammed tissue in bacteria scarce region. d. Cell-cell interaction analysis of control tissue in bacteria abundant region. e. Cell-cell interaction analysis of CD noninflammed tissue in bacteria abundant region. f. Cell-cell interaction analysis of CD inflammed tissue in bacteria abundant region. Bacteria scarce regions and bacteria abundant regions were divided by a CPM normalized bacteria read count value of 5.

Given that Crohn's disease (CD) is characterized by impaired microbial sensing, defects in microbicidal mechanisms, and dysregulated cytokine responses, we conducted a cell-cell interaction

analysis to explore potential differences in immune response networks. We analyzed these interactions separately in bacteria-scarce and bacteria-abundant regions. In control tissue, signaling from enterocytes to other cells increased significantly in bacteria-abundant regions. However, in CD tissue, this signaling showed minimal variation between bacteria-scarce and bacteria-abundant regions. These findings suggest that the mechanism by which enterocytes sense bacteria and transmit signals to initiate immune responses is weakened in CD.

3.3. Bacterial Presence Activates Host Immune Responses

Genes differentially expressed based on the presence of 81 bacterial species were identified. Subsequently, the consistency of this expression pattern was examined across different bacterial species to identify genes whose expression was consistently induced or suppressed by the presence of bacteria. Furthermore, using the BioPlanet database, we investigated the functions of gene that showed consistent overexpression in the presence of the 81 bacterial species (Huang et al. 2019). Gene expression analysis in all cells showed that bacterial presence induced upregulation of immune system components, including B cell receptor signaling (Figure 15a, 15b). Meanwhile, lipid absorption pathway of the intestine, such as chylomicron-mediated lipid transport, were downregulated in the presence of bacteria (Figure 15c).

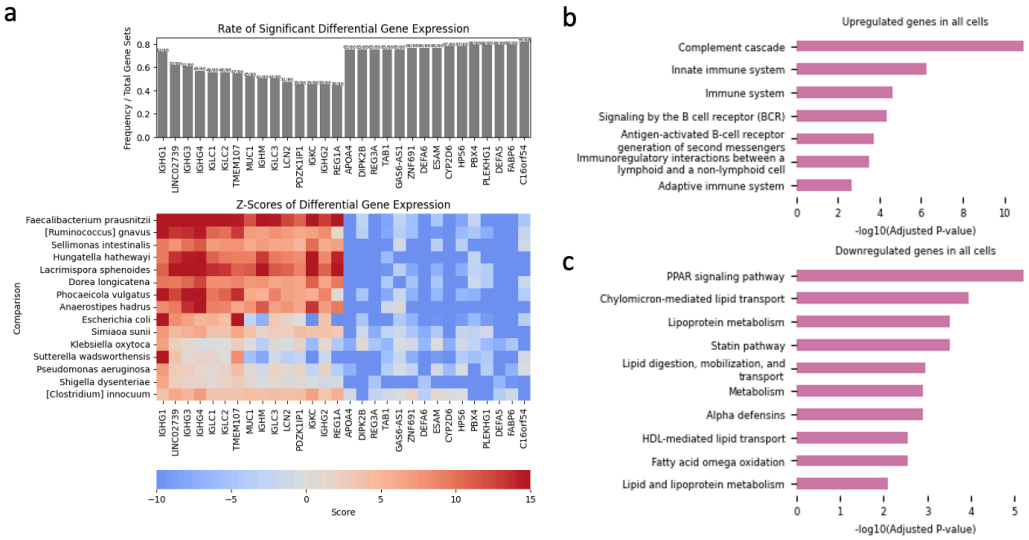

Figure 15. Differential gene expression and pathway analysis in response to bacterial presence in all cells. a. The top panel shows the rate of significant differential gene expression in response to

various bacteria in all cells. The heatmap below displays Z-scores of differential gene expression for specific bacterial species. b. Gene set enrichment analysis of upregulated genes in all cells in response to bacterial presence. c. Gene set enrichment analysis of downregulated genes in all cells in response to bacterial presence.

Gene expression analysis of immune cell-rich and enterocyte-rich areas demonstrated consistent upregulation of genes encoding immunoglobulin components in response to bacterial presence, including IGHG4, IGKC, IGHA1, JCHAIN, and IGHG3 in immune cell-rich areas and IGHG1, IGHG3, IGHG4, IGLC2, IGKC, IGHA1, IGLC3, and IGHM in enterocyte-rich areas (Figure 16a, 16b) (Mikocziova, Greiff, and Sollid 2021). Functional enrichment analysis revealed that upregulated genes in both areas were involved in bacterial defense mechanisms, including interleukin signaling and innate immune system pathway (Figure 16c, 16d) (Huang et al. 2019). Moreover, the genes in the immune cell-rich areas showed a higher degree of consistency in their differential expression compared with those in the enterocyte-rich areas, suggesting that the transcriptional response of immune cells to bacterial presence is more uniform and robust, underlining their central role in the bacterial defense response.

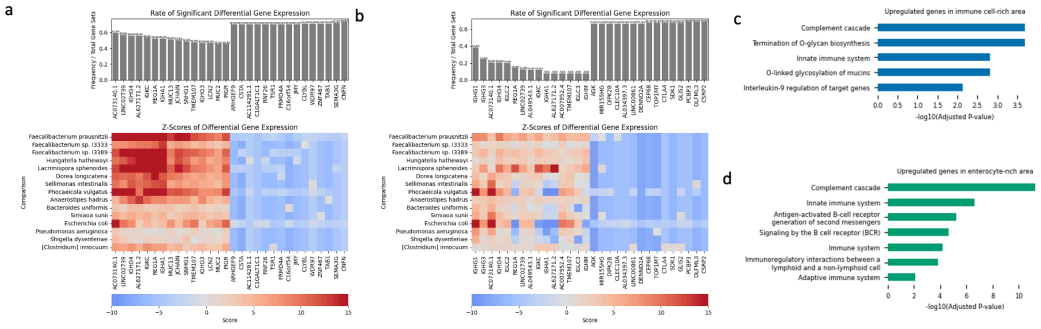

Figure 16. Differential gene expression and pathway analysis in response to bacterial presence. a. The top panel shows the rate of significant differential gene expression in response to various bacteria in immune cell-rich area. The heatmap below displays Z-scores of differential gene expression for specific bacterial species in immune cell-rich area. b. The top panel shows the rate of significant differential gene expression in response to various bacteria in enterocyte-rich area. The heatmap below displays Z-scores of differential gene expression for specific bacterial species in enterocyte-rich area. c. Gene set enrichment analysis of upregulated genes in immune cell-rich areas in response to bacterial presence. d. Gene set enrichment analysis of upregulated genes in

response to bacterial presence in enterocyte-rich area.

3.4. Bacterial Species Distinctly Impact Cell Viability in Crohn's Disease

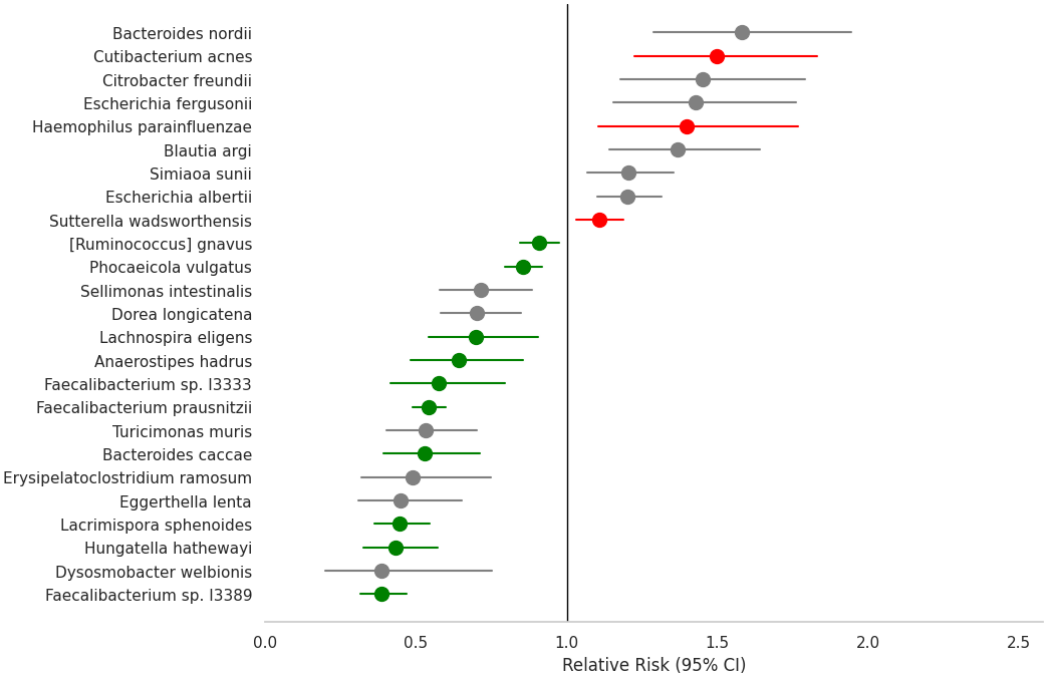

Figure 17. Relative risk of reduced cell viability following various bacterial exposure. Cells with ≤ 10% mitochondrial reads were defined as viable cells, while those with > 10% were defined as damaged cells. We performed multiple testing correction using the Benjamini-Hochberg method with a significance threshold of 0.05, and only statitically significant results after this correction were visualized. Species previously reported to increase the risk of CD are shown in red, while those known to decrease the CD risk are shown in green.

We investigated intestinal barrier disruption, a hallmark of CD, at the cellular level by using the percentage of mitochondrial reads, a widely used marker of cell viability in single-cell RNA sequencing and spatial transcriptomics analyses (Neurath 2019; Boyapati et al. 2018; Ilicic et al. 2016; D'incà et al. 2006). The RR values of reduced cell viability were calculated following exposure to specific bacterial species to quantify the effect of bacterial exposure on intestinal barrier

disruption. We performed multiple testing corrections using the Benjamini–Hochberg method with a significance threshold of 0.05 and visualized only the significant results after this correction. This analysis was performed solely on samples from participants with CD, without including those from control participants, to specifically focus on the impact of bacterial exposure on intestinal barrier integrity in the context of CD. The RR values and their 95% confidence intervals were displayed for various bacterial species (Figure 17). Species previously reported to increase CD risk are shown in red, while those known to decrease it are shown in green (Gevers et al. 2014; Kansal et al. 2019) (Supplementary Table 1). This demonstrates that previously known beneficial and pathogenic microbiomes were consistently observed in our study.

Bacteroides caccae (RR = 0.53; 95% confidence interval [CI], 0.39–0.71), F. prausnitzii (RR = 0.54; 95% CI, 0.49–0.60), Phocaeicola vulgatus (RR = 0.85; 95% CI, 0.80–0.92), and Ruminococcus gnavus (RR = 0.91; 95% CI, 0.84–0.98) were confirmed as beneficial microbes, aligning with previous findings. Anaerostipes hadrus (RR = 0.64; 95% CI, 0.48–0.85), Hungatella hathewayi (RR = 0.43; 95% CI, 0.33–0.57), Faecalibacterium sp. I3333 (RR = 0.57; 95% CI, 0.41–0.79), and sp. I3389 (RR = 0.38; 95% CI, 0.31–0.47) were previously identified as beneficial at the genus level, even though their species-level association with CD were not reported. Cutibacterium acnes (RR = 1.50; 95% CI, 1.22–1.83), H. parainfluenzae (RR = 1.40; 95% CI, 1.10–1.17), and Sutterella wadsorthensis (RR = 1.11; 95% CI, 1.03–1.19) were confirmed as pathogenic microbes at the species level, consistent with prior studies.

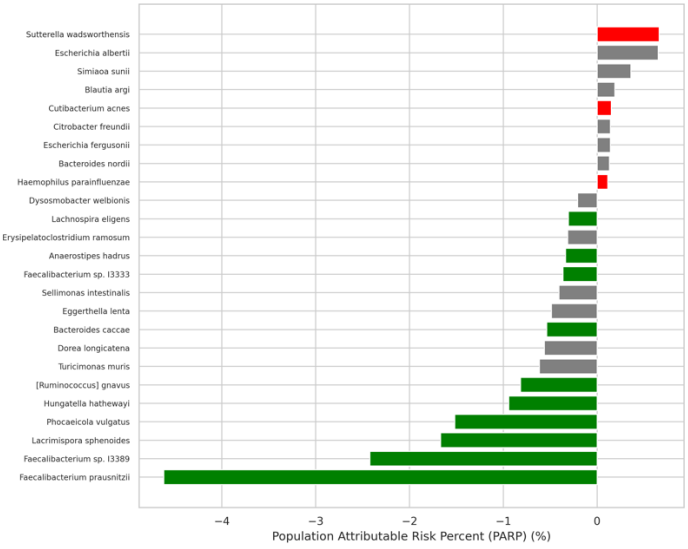

Figure 18. Population Attributable Risk Percent (PARP) for intestinal barrier disruption by bacterial species. Cells with ≤ 10% mitochondrial reads were defined as viable cells, while those with > 10% were defined as damaged cells. We performed multiple testing correction using the Benjamini-Hochberg method with a significance threshold of 0.05, and only statitically significant results after this correction were visualized. Species previously reported to increase the risk of CD are shown in red, while those known to decrease the CD risk are shown in green.

We also calculated the PARP by integrating the RR values of specific bacteria with their gut prevalence. This provided a measure of the overall contribution of each bacterial species to intestinal barrier disruption within the entire gut environment. F. prausnitzii, a highly abundant gut bacterium known for its beneficial properties in CD, was the most influential in reducing intestinal barrier disruption (by -3.1%).

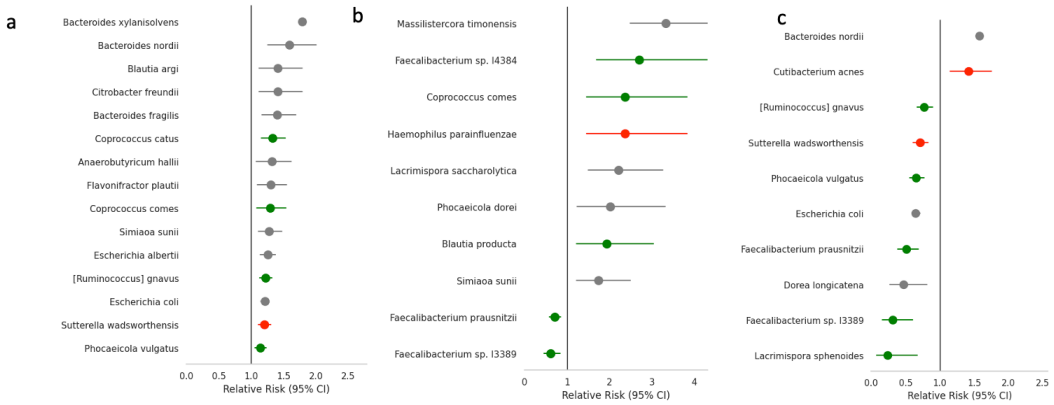

Figure 19. Relative risk of reduced cell viability following bacterial exposure in various cell types. Relative risk (RR) of reduced cell viability after bacterial exposure in a. enterocyte-rich areas, b. immune cell-rich areas, and c. tuft cell-rich areas. We performed multiple testing correction using the Benjamini-Hochberg method with a significance threshold of 0.05, and only statitically significant results after this correction were visualized. Species previously reported to increase the risk of CD are shown in red, while those known to decrease the CD risk are shown in green.

When the RR was examined specifically in enterocyte-rich areas (Figure 19a), all bacteria were

found to reduce cell viability, even species such as Phocaeicola vulgatus, which was initially found to inhibit this reduction during analysis of the entire cell population. In contrast, analyses of immune cell- and tuft cell-rich areas revealed that some bacteria exhibited a protective effect by inhibiting the reduction in cell viability (Figure 19b, 19c). Additionally, bacterial exposure increased the fraction of damaged cells in enterocyte-rich areas, but decreased it in tuft cell-rich, immune cell-rich, and myofibroblast-rich areas. This suggests that the beneficial effect of certain gut microbiomes may be facilitated through interactions with cell types other than enterocytes, potentially including immune and tuft cells. These findings highlight the complex interplay between various host gut cell types and the resident bacteria.

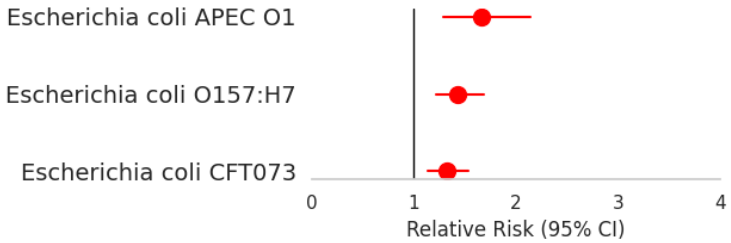

Figure 20. Relative risk of reduced cell viability following exposure to various Escherichia coli strains. Relative risk (RR) of reduced cell viability after exposure of various Escherichia coli strains.

To account for the heterogeneity of Escherichia coli, a highly abundant species with diverse strains ranging from commensal to pathogenic, we calculated the RR values for 42 different E. coli strains characterized at the strain level, although only 15.2% were identified at this taxonomic resolution. We identified E. coli APEC O1 (RR = 1.66; 95% CI, 1.29–2.14), O157:H7 (RR = 1.43; 95% CI, 1.22–1.68), and CFT073 (RR = 1.32; 95% CI, 1.14–1.53) as part of the pathogenic microbiome, which is consistent with previous reports (Johnson et al. 2007; Kao et al. 1997; Riley et al. 1983).

3.5. Beneficial and Pathogenic Microbiomes Distinctly Modulate Host Transcription

We analyzed differentially expressed genes between cells exposed to beneficial and to pathogenic microbiomes. Using 16 beneficial and nine pathogenic members identified in this study, we formed 144 beneficial-pathogenic microbiome pairs. For each pair, we compared gene expressions in

response to the beneficial versus pathogenic microbiome. We assessed the consistency of these changes across the different pairs, identifying genes consistently upregulated or downregulated in the presence of beneficial or pathogenic bacteria.

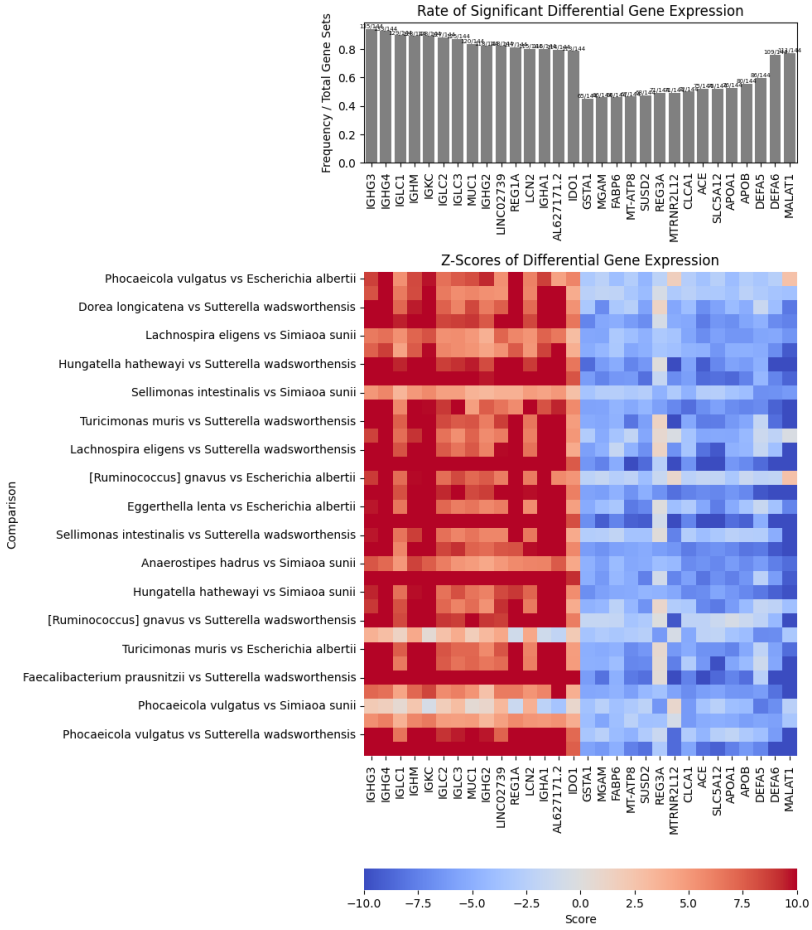

Figure 21. Differential gene expression between beneficial and pathogenic microbiome exposure in all cells. a. Differential expressed gene between cells exposed to beneficial and pathogenic microbiomes were visualized. The top panel shows the rate of significant differential gene expression between beneficial and pathogenic microbiome exposure in all cells in all cells. The heatmap below displays Z-scores of differential gene expression.

Across all cells, beneficial microbiomes increased the expression of immunoglobulin genes,

including IGHG3, IGHG4, IGLC1, IGHM, IGKC, IGLC2, IGHG2, and IGHA1 (Figure 21) (Mikocziova, Greiff, and Sollid 2021). This suggests that beneficial microbiomes may exert their effects through immune cells. In immune cell-rich areas, genes such as REG1A and TNFRSF6B showed increased expression in regions with beneficial microbiomes (Figure 22a). Similarly, in enterocytes, genes including IGKC and REG1A were upregulated in such areas (Figure 22b).

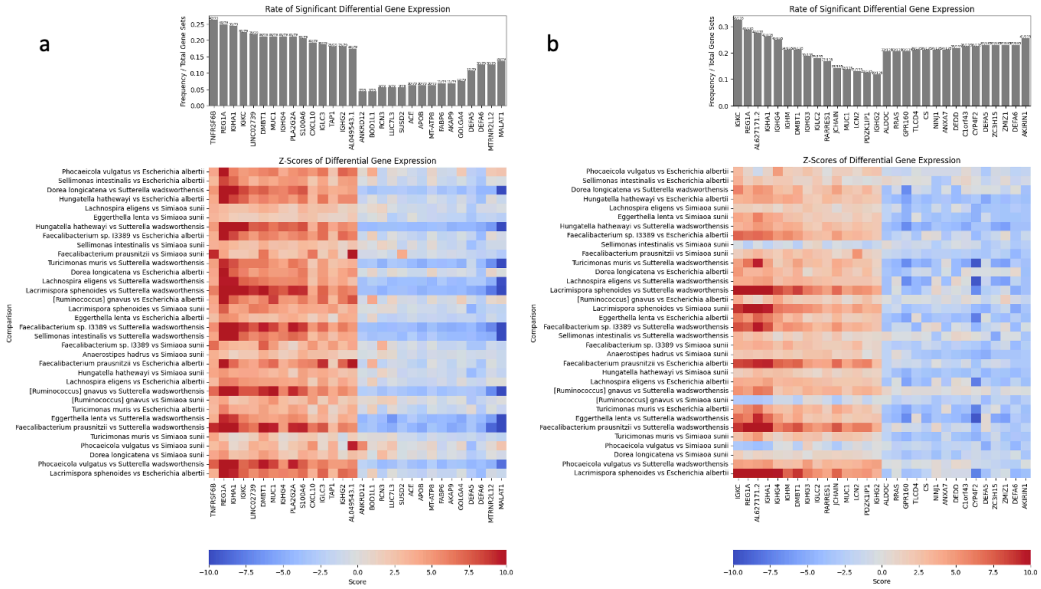

Figure 22. Differential gene expression between beneficial and pathogenic microbiome exposure in immune cell-rich area and enterocyte rich area. a. Differential expressed gene between cells exposed to beneficial and pathogenic microbiomes were visualized in exposure in a. immune cell-rich area and b. enterocyte rich area. The top panel shows the rate of significant differential gene expression between beneficial and pathogenic microbiome exposure in all cells in all cells. The heatmap below displays Z-scores of differential gene expression.

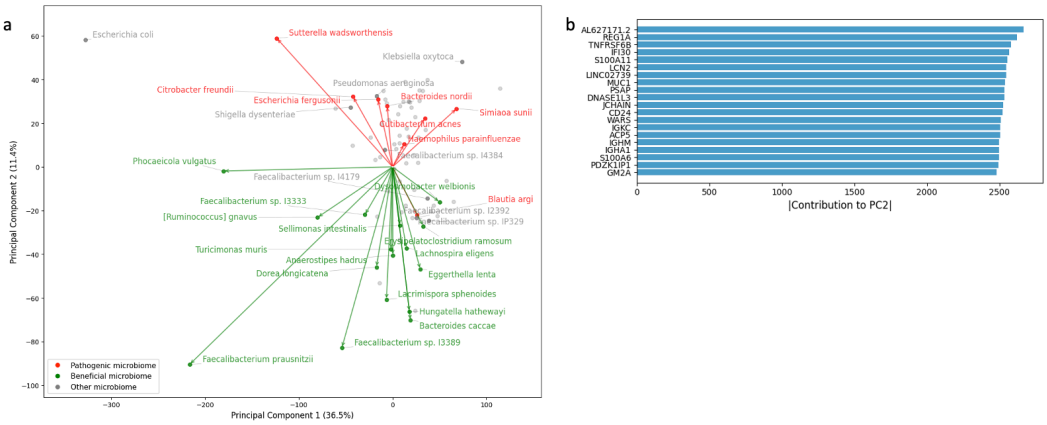

Figure 23. Correlation analysis of microbiome and host transcriptome. a. Principal Component plot of pearson's correlation between bacterial presence and gene expression levels, with pathogenic (red), and beneficial (green) microbiome candidates. b. Bar chart displaying the top 20 gene contributors to Principal Component 2 (PC2). PCA: Principal Component Analysis; PC: Principal Component

To assess species-specific bacterial effects on host gene expression, we calculated Pearson correlations for bacterial presence and gene expression levels. We then applied PCA for dimension reduction of the correlation matrix. The results showed a clear distinction between beneficial and pathogenic microbial species along PC2 (Figure 23a). Furthermore, the expression of genes such as REG1A and TNFRSF6B had substantial contributions in PC2 (Figure 23b).

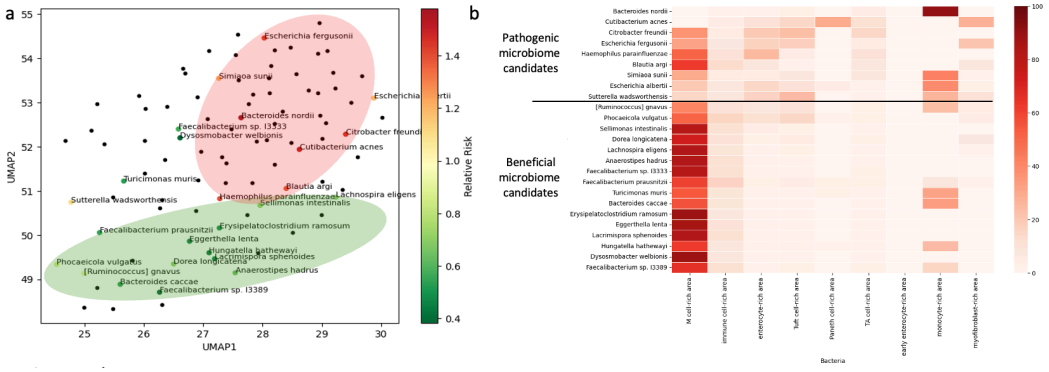

Figure 24. Coexistence of bacteria across cell types a. UMAP plot of coexisting bacteria in tissue

of Crohn's disease patients. Dimension reduction was performed on the spatial information of bacteria in different cell types using UMAP, compressing it into two-dimensional data and visualizing this on a 2D plane. For beneficial microbiome candidates and pathogenic microbiome candidates, the previously determined Relative Risk values were represented using colors. b. Heatmap visualization of bacterial abundance by cell in each bacteria cluster in tissue of Crohn's disease patients. CPM normalized values by bacteria were summed for each cluster.

On the UMAP plot, beneficial microbiome candidates were positioned near each other, as were pathogenic microbiome candidates. Beneficial microbiome candidates were more concentrated in M cell-rich and immune cell-rich areas, while pathogenic candidates were more prevalent in Enterocyte-rich and Tuft cell-rich area.

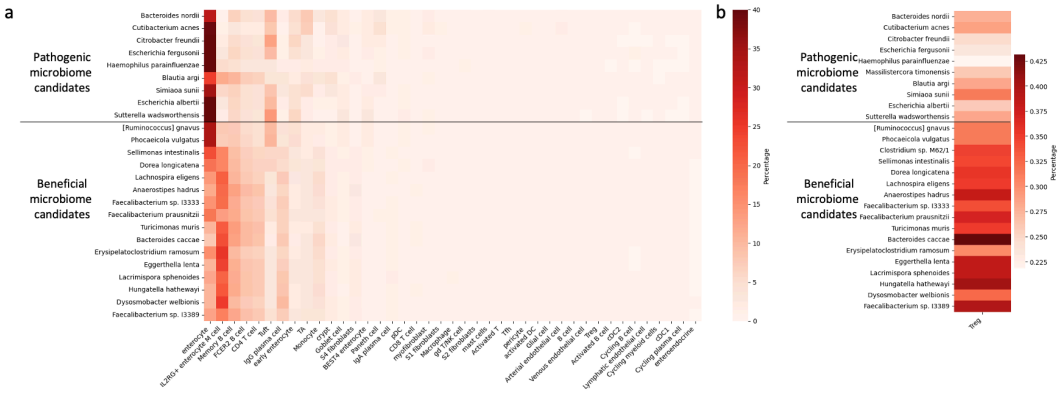

Figure 25. Coexistence of bacteria across specific cell types a. Heatmap visualization of bacterial abundance by specific cell type in each bacteria in tissue of Crohn's disease patients. Estimated cell type proportions using Cell2location were summed for each bacteria. b. Heatmap visualization of bacterial abundance by regulator T cells in each bacteria in tissue of Crohn's disease patients.

Estimation of coexistence frequencies between each bacterial species and 42 cell types revealed distinct patterns between beneficial and pathogenic microbiome candidates. Beneficial microbiome candidates showed higher frequencies particulary with M cells and Memory B cells, while pathogenic candidates predominantly found with enterocytes and tuft cells. Notably, T_{reg} cells, which play a key role in the protective effects of beneficial microbiomes, showed accumulation in regions

containing beneficial microbiome candidates.

3.6. Deep Learning Detect Bacterial Translocation in Histological Images

Our deep learning model, designed to detect bacterial infiltration, demonstrated varying effectiveness across different bacterial categories. For the beneficial microbiome category, it showed the highest discriminative ability with an AUROC of 0.7616 and the best accuracy at 0.7454. The whole bacteria category, on the other hand, exhibited the most balanced performance, achieving the highest F1 score of 0.6582 and the best precision (0.5588) among all categories. In contrast, the pathogenic microbiome category, while having comparable accuracy (0.6978) to the whole bacteria category, showed the lowest F1 score (0.2836).

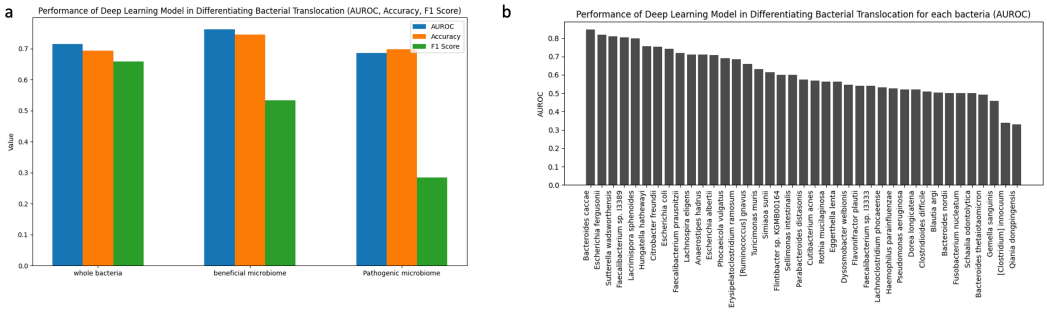

Figure 26. Performance of Deep Learning Model in Differentiating Bacterial Translocation a. AUROC, Accuracy, and F1 Score of the deep learning model for whole bacteria, beneficial microbiome, and pathogenic microbiome categories. b. AUROC values for individual bacterial species detected by the deep learning model. AUROC: Area Under the Receiver Operating Characteristic curve.

To further investigate the relationship between tissue morphology and bacterial presence, we employed a deep learning-based approach to cluster tissue images. Using features extracted from the images, we performed K-Means clustering, resulting in five distinct morphological clusters. Notably, cluster 4, characterized by neutrophil infiltration, demonstrated significant positive correlations with all beneficial microbiome species (mean r = 0.097). These correlations were statistically significant after FDR correction ($p < 0.05$) for all beneficial bacterial species. This finding suggests a potential association between neutrophil-rich tissue environments and the presence of beneficial bacteria.

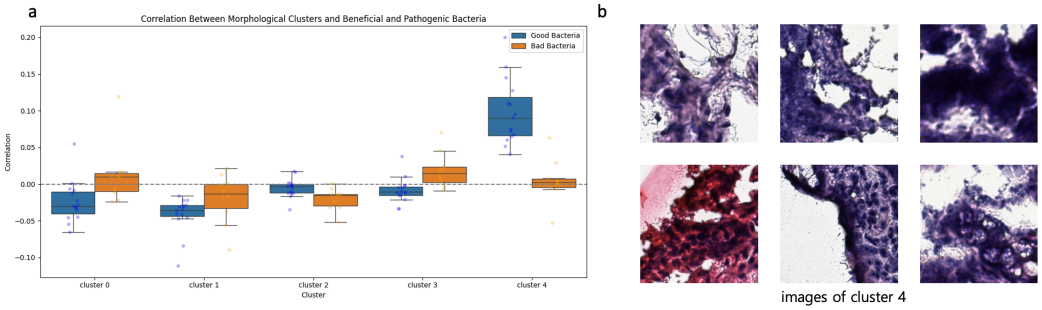

Figure 27. Performance of Deep Learning Model in Differentiating Bacterial Translocation a. AUROC, Accuracy, and F1 Score of the deep learning model for whole bacteria, beneficial microbiome, and pathogenic microbiome categories. b. AUROC values for individual bacterial species detected by the deep learning model. AUROC: Area Under the Receiver Operating Characteristic curve.

3.7. Semi-supervised Learning Discriminates True Bacteria from False Positives

Several bacterial species showd near-complete signal loss following double human read removal compared to single human read removal (Figure 4a). These species showed significantly redued abundance in bulk metagenome sequencing of identical tissue samples, indicating they were likely contaminants from misidentified human reads (Figure 4b). However, we noted that overly stringent human read procedures, while reducing false positives, could potentially diminish genuine biological signals of true pathogen ((Figure 4c).

Based on these findings, we established comprehensive criteria for bacterial identification by comparing single and double human read removal process with bulk metagenome sequencing results. However, conducting bulk metagenome sequencing in addition to spatial transcriptomics analysis, as done in this study, may not be feasible in other research settings. Considering situations where metagenome sequencing for validation is not available, we classified species into three categories based solely on comparing single and double human read removal processes: 54 definitive true bacteria that retained substantial signals after double human read removal, 43 definitive contaminant species that showed complete signal loss after double huamn read removal, and remaining unlabeled

species.

Among the 81 bacterial species validated by both single-double human read removal comparison and metagenome sequencing, our semi-supervised deep learning model achived an AUC of 1.00 for labeled species and 0.84 for the whole species. The model classified the unlabeled species with an accuracy of 0.875, demonstrating its utility for bacterial identification in settings where metagenome validationa is not available.

Metrics	Total dataset	Labeled dataset	Unlabeled dataset
Accuracy	0.68	0.98	0.66
AUROC	0.84	1.00	0.79

Table 2. Performance of Semi-Supervised Deep Learning Model in Differentiating True Bacteria and False positive species. AUROC: Area Under the Receiver Operating Characteristic curve.

4. DISCUSSION

The critical roles of bacteria-host interactions in CD are increasingly acknowledged. However, the intricate relationships between various cell types and bacterial species at the cellular level is still poorly understood (Neurath 2019; Graham and Xavier 2020; Haberman et al. 2014). Therefore, we developed a novel spatial host-microbiome profiling approach that, to the best of our knowledge, is the first to enable simultaneous species-level identification of bacteria and host transcriptomics. Using this method, we demonstrated increased bacterial translocation in CD, with a significant association between the extent of translocation and disease prognosis, while also revealing distinct host transcriptome alterations in response to translocation of various bacterial species. Furthermore, we identified and characterized potentially beneficial and pathogenic microbial species associated with CD, including several newly discovered risk-modulating bacterial species. Our spatial host-microbiome profiling approach not only provides profound insights regarding CD, but also offers potential applications for studying various microbiome-associated diseases such as gastrointestinal cancer and infectious diseases. It reveals intricate interactions between translocating gut bacteria and various host cell types in CD pathophysiology at the cellular level. Moreover, the identification of beneficial and pathogenic microbiome members enables the development of microbiome-based

therapeutic strategies.

Our results highlighted the importance of bacterial translocation in CD, revealing that CD tissues exhibited increased bacterial presence compared with that in controls. Furthermore, within CD samples, compared with non-inflamed tissues, inflamed tissues showed increased bacterial translocation. These findings suggest a more active bacterial invasion in CD, potentially due to impaired host defense mechanisms, which aligns with the findings of Sun, D, et al., who also reported an increase in bacterial translocation in inflamed tissues compared with non-inflamed tissues in CD.(Sun et al. 2021) Moreover, the extent of bacterial translocation at diagnosis not only predicted disease prognosis in patients with CD but also showed a strong association with the severity of endoscopic findings in the ileum. This indicates that the level of bacterial infiltration in intestinal tissues could serve as a potential prognostic marker for predicting disease course in pediatric CD.

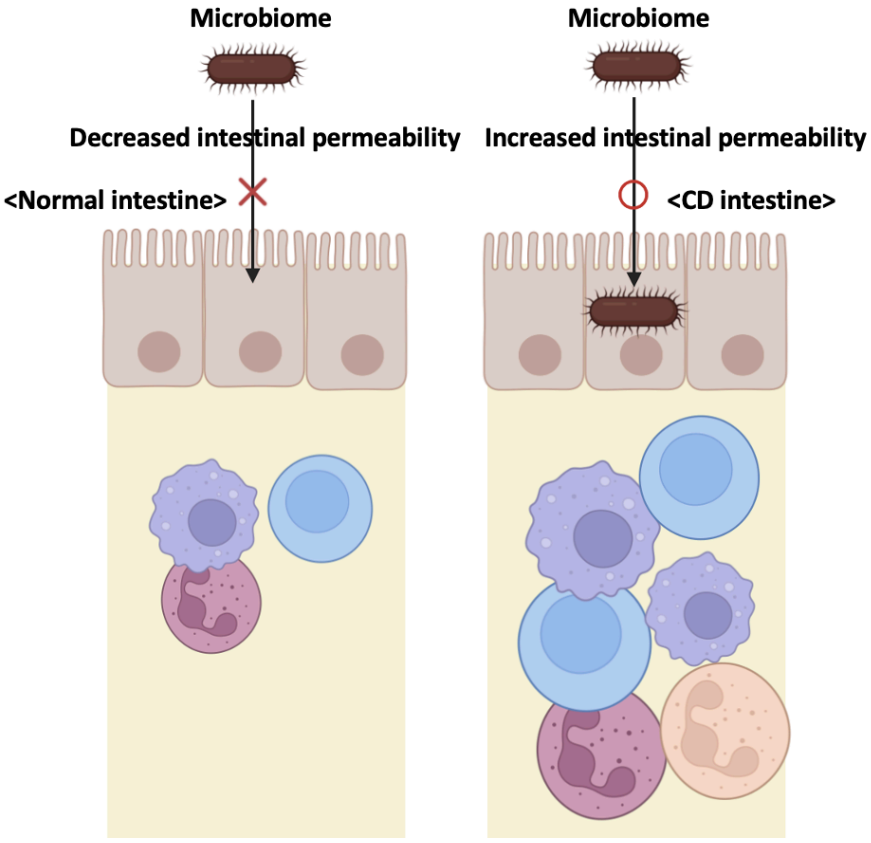

Figure 28. Schematic overview of increased bacterial infiltration in Crohn's disease Schematic comparison of bacterial infiltration in normal (left) and Crohn's disease (CD) intestines (right). CD intestines exhibit increased permeability, leading to enhanced bacterial infiltration compared to normal intestines.

Analysis of differentially expressed genes in response to the presence of bacterial species revealed a specific and targeted response to bacterial infiltration. The consistent upregulation of genes encoding immunoglobulin components appears to be an effect of bacterial translocation on host gene expression, which was further confirmed by gene enrichment analysis revealing the involvement of these upregulated genes in bacterial defense functions. Notably, the higher degree of consistency in the differential expression of genes in immune cell-rich areas compared with enterocyte-rich areas shows the uniform and robust transcriptional response of immune cells to bacterial presence, highlighting their crucial role in orchestrating an effective defense against bacterial translocation. Furthermore, our data showed the upregulation of lncRNA, such as LINC02739, in response to bacterial infiltration, suggesting their potential involvement in the host-microbe interaction and disease pathogenesis. While the specific functions of these lncRNAs remain unclear, their differential expression warrants further investigation into their roles in modulating the immune response and disease progression in CD (Atianand, Caffrey, and Fitzgerald 2017).

To identify potentially beneficial and pathogenic microbes in CD, we assessed the impact of specific bacterial species on intestinal barrier integrity by determining the RRs of reduced cell viability upon exposure to these bacteria (Table 3). Our analysis identified 16 beneficial and nine pathogenic microbiome candidates. Among these, six were previously reported to have a confirmed impact on CD at the species level, while eight were known to affect only at the genus level (Gevers et al. 2014; Kansal et al. 2019). Interestingly, the majority of the beneficial microbiome members we identified (13 out of 16 species) belong to Bacillota (Firmicutes), with a substantial number (11 species) belonging to Clostridium clusters XIVa and IV, which have been previously associated with a reduced risk of CD (Andoh et al. 2011; Liu et al. 2008; Seo et al. 2016; Taras et al. 2002; Mahowald et al. 2009; Allen-Vercoe et al. 2012; Schoch et al. 2020; Haas and Blanchard 2020; Kaur et al. 2014; Le Roy et al. 2020). In contrast, five out of nine pathogenic microbiome members, including Citrobacter freundii, Escherichia fergusonii, H. parainfluenzae, Escherichia albertii, and Sutterella wadsworthensis, belong to Pseudomonadota (Proteobaccteria). This aligns with previous studies

suggesting an association between increased abundance of Pseudomonadota (Proteobaccteria) and CD (Haberman et al. 2014; Vester-Andersen et al. 2019). Notably, we identified five beneficial microbiome members that were not reported in previous studies. Four of these newly identified beneficial microbes belong to Bacillota (Firmicutes), showing similar phylogeny to other known beneficial microbes. Interestingly, three of these newly identified species, Sellimonas intestinalis, Turicimonas muris, and Dysosmobacter welbionis, have been reclassified or newly discovered after the publication of the reference studies we used (Gevers et al., 2014; Kansal et al., 2019) (Gevers et al. 2014; Kansal et al. 2019; Seo et al. 2016; Le Roy et al. 2020; Lagkouvardos et al. 2016). These findings demonstrate that our newly identified beneficial microbes are aligned with those from previous reports and highlight how our understanding of the microbiome deepens as bacterial databases expand and taxonomic classifications become more precise.

Bacteria	Relative Risk	Phylum	Clostridium cluster
Bacteroides nordii	1.58	Bacteroidota	
Cutibacterium acnes	1.497	Actinomycetota	
Citrobacter freundii	1.452	Pseudomonadota	
Escherichia fergusonii	1.426	Pseudomonadota	
Haemophilus parainfluenzae	1.397	Pseudomonadota	
Blautia argi	1.368	Bacillota	cluster XIVa
Simiaoa sunii	1.202	Bacillota	
Escherichia albertii	1.202	Pseudomonadota	
Sutterella wadsworthensis	1.106	Pseudomonadota	
[Ruminococcus] gnavus	0.908	Bacillota	cluster XIVa
Phocaeicola vulgatus	0.854	Bacteroidota	
Sellimonas intestinalis	0.715	Bacillota	cluster XIVa
Dorea longicatena	0.701	Bacillota	cluster XIVa
Lachnospira eligens	0.699	Bacillota	cluster XIVa
Anaerostipes hadrus	0.642	Bacillota	cluster XIVa
Faecalibacterium sp. I3333	0.574	Bacillota	cluster IV

Faecalibacterium prausnitzii	0.54	Bacillota	cluster IV
Turicimonas muris	0.531	Bacillota	
Bacteroides caccae	0.529	Bacteroidota	
Erysipelatoclostridium ramosum	0.489	Bacillota	cluster XVIII
Eggerthella lenta	0.448	Actinomycetota	
Lacrimispora sphenoides	0.445	Bacillota	cluster XIVa
Hungatella hathewayi	0.431	Bacillota	cluster XIVa
Dysosmobacter welbionis	0.387	Bacillota	cluster IV
Faecalibacterium sp. I3389	0.383	Bacillota	cluster IV

Table 3. Phylum and Clostridium Cluster Classification of Beneficial and Pathogenic Microbiomes in Crohn's Disease List of bacterial species identified as beneficial or pathogenic microbiomes in our Crohn's Disease (CD) study. The table presents the phylum and Clostridium cluster classification of each bacterium when relevant.

The mechanisms underlying the protective effects of gut microbiota in inflammatory bowel disease are not yet fully understood, but evidence points to their interaction with T_{reg} cells as a key immunological pathway(Sefik et al. 2015; Ohnmacht et al. 2015; Atarashi et al. 2011; 2013; Vignali, Collison, and Workman 2008). Atarashi et al. demonstrated that Clostridium clusters IV and XIVa promote Treg cell accumulation in the colon, playing a critical role in maintaining immune homeostasis (Atarashi et al. 2011). Their subsequent study confirmed that mixtures of Clostridia strains isolated from human microbiota could effectively induce accumulation of T_{reg} cells and reduce colitis severity (Atarashi et al. 2013). In our study, beneficial microbiome candidates not only inhibited the reduction in cell viability in pediatric Crohn's disease tissue samples but also demonstrated T_{reg} cell accumulation in human intestinal tissue in situ.

Our study highlights the impact of gut microbes on CD pathogenesis is not uniform, even within the same genus, as demonstrated by the Faecalibacterium genus. While F. prausnitzii, Faecalibacterium sp. I3389, and Faecalibacterium sp. I3333 were identified as beneficial microbes in this study, Faecalibacterium sp. I2392, Faecalibacterium sp. I4179, Faecalibacterium sp. I4384, and Faecalibacterium sp. IP329 did not have statistically significant RR values. Additionally, our spatial microbiome analysis at the strain level confirmed that E. coli, which have various strains ranging from commensal to pathogenic, showed varying associations with CD risk (Johnson et al.

2007; Riley et al. 1983; Kao et al. 1997; Kaper, Nataro, and Mobley 2004).

Comparing tissue-specific gene expression between cells exposed to beneficial and pathogenic microbiomes, we observed increased expression of genes such as REG1A and TNFRSF6B in the presence of beneficial microbiomes. According to Mao et al., REG1A plays a crucial role in tissue regeneration and the repair of intestinal epithelial damage (Mao et al. 2021). In their study, inducing REG1A expression in a DSS colitis mouse model promoted the recovery of the intestinal barrier. TNFRSF6B, also known as DcR3, is a well-known inhibitor of FASL and LIGHT, which are essential for cell apoptosis (Su et al. 2023). The upregulation of TNFRSF6B in the presence of beneficial bacteria suggests that these microbes may be associated with the maintenance of intestinal epithelial integrity, potentially through their influence on host cell apoptosis pathways. These findings provide insights into the potential mechanisms by which beneficial microbiomes may provide protection against CD pathogenesis, emphasizing the need for further research to understand the complex interactions between gut microbes and host cellular processes.

Our spatial host-microbiome sequencing approach offers several advantages over traditional methods. Unlike dissociation-based single-cell RNA sequencing methods such as Chromium, our approach allows for the accurate identification of host cells exposed to bacteria, as it avoids the potential dissociation of bacteria and host cells during sample preparation (Zheng et al. 2017). Additionally, our method minimizes the risk of contamination by ambient RNA, a common issue in droplet-based methods, resulting from cell lysis and RNA release within the microfluidic droplets, ensuring that the detected bacterial reads accurately reflect the true bacterial distribution within the tissue (Caglayan, Liu, and Konopka 2022). Furthermore, our approach enables the analysis of damaged cells, which are typically removed in dead cell removal processes prior to single-cell RNA sequencing studies (Ilicic et al. 2016). This is particularly considerable in the context of CD, where removing damaged cells may inadvertently exclude the more inflamed portion of the tissue, leading to a potential bias in the results.

Previous studies using 16s rRNA targeting approaches identified bacterial distributions only at the genus level (Durazzi et al. 2021). In contrast, the Kraken2-based shotgun metagenomic profiling approach employed in our study allows for the detection of even small amounts of bacteria and enables differentiation at the strain level (Wood, Lu, and Langmead 2019). Moreover, our approach eliminates the need for creating probes, allowing for the evaluation of a wide range of bacteria, including those less frequently reported in the literature, some of which may have potential as

therapeutic targets (Galeano Niño et al. 2022; Saarenpää et al. 2023; Lötstedt et al. 2023). Unlike probe-based methods that restrict analysis to pre-selected bacterial targets, our shotgun metagenomic approach enables comprehensive identification of all bacterial species present in the sample. This is achieved through a simple sample preparation step with the addition of yeast poly A polymerase, eliminating the need for the cumbersome preparation of new probes for specific bacteria.

This study has several limitations. First, the small sample size and limited age range may introduce potential biases, although the total number of cells analyzed was substantial (13,876). Future studies should include more samples from diverse ethnic backgrounds to improve the generalizability of the findings. Second, the spatial transcriptomic technology lacks single-cell resolution, resulting in the mixing of various cell types within spots (Kleshchevnikov et al. 2022). Although we employed Cell2location to classify areas based on cell combinations, heterogeneity within areas of the same cell-type may still exist, meaning that there could be slight differences in cellular composition among areas classified as the same type. This heterogeneity could potentially influence the comparisons made within these areas, such as those for differentially expressed genes in immune cell-rich areas due to bacterial exposure. However, the emergence of high-resolution single-cell spatial transcriptomics techniques offers opportunities to apply our algorithm to more refined spatial data in future investigations, which could help reduce the impact of cellular heterogeneity on our analyses (Chen et al. 2023).

5. CONCLUSION

In conclusion, our study introduces a novel spatial host-microbiome profiling approach that enables the simultaneous profiling of the host transcriptome and bacterial species at a high taxonomic resolution in the ileal tissues of pediatric patients with CD. This approach allowed us to identify increased bacterial translocation in CD tissues, as well as the potential prognostic value of assessing bacterial infiltration in intestinal tissues. We also discovered specific beneficial and pathogenic microbiomes associated with CD pathogenesis and suggested potential mechanisms by which these microbes may influence disease progression, such as the modulation of host cell apoptosis pathways. The identification of several newly discovered beneficial microbiomes provides promising candidates for the development of novel microbiome-based therapeutics for CD. Our spatial host-microbiome sequencing approach offers a valuable method to understand the intricate interactions between gut microbes and host cells in the context of CD pathogenesis.

REFERENCES

1. Allen-Vercoe, Emma, Michelle Daigneault, Aaron White, Remo Panaccione, Sylvia H. Duncan, Harry J. Flint, Lindsey O'Neal, and Paul A. Lawson. 2012. "Anaerostipes Hadrus Comb. Nov., a Dominant Species within the Human Colonic Microbiota; Reclassification of Eubacterium Hadrum Moore et al. 1976." *Anaerobe* 18 (5): 523–29. https://doi.org/10.1016/j.anaerobe.2012.09.002.

2. Andoh, Akira, Hirotsugu Imaeda, Tomoki Aomatsu, Osamu Inatomi, Shigeki Bamba, Masaya Sasaki, Yasuharu Saito, Tomoyuki Tsujikawa, and Yoshihide Fujiyama. 2011. "Comparison of the Fecal Microbiota Profiles between Ulcerative Colitis and Crohn's Disease Using Terminal Restriction Fragment Length Polymorphism Analysis." *Journal of Gastroenterology* 46 (4): 479–86. https://doi.org/10.1007/s00535-010-0368-4.

3. Atarashi, Koji, Takeshi Tanoue, Kenshiro Oshima, Wataru Suda, Yuji Nagano, Hiroyoshi Nishikawa, Shinji Fukuda, et al. 2013. "Treg Induction by a Rationally Selected Mixture of Clostridia Strains from the Human Microbiota." *Nature* 500 (7461): 232–36. https://doi.org/10.1038/nature12331.

4. Atarashi, Koji, Takeshi Tanoue, Tatsuichiro Shima, Akemi Imaoka, Tomomi Kuwahara, Yoshika Momose, Genhong Cheng, et al. 2011. "Induction of Colonic Regulatory T Cells by Indigenous Clostridium Species." *Science* 331 (6015): 337–41. https://doi.org/10.1126/science.1198469.

5. Atianand, Maninjay K., Daniel R. Caffrey, and Katherine A. Fitzgerald. 2017. "Immunobiology of Long Noncoding RNAs." *Annual Review of Immunology* 35 (April):177–98. https://doi.org/10.1146/annurev-immunol-041015-055459.

6. Balish, Edward, and Thomas Warner. 2002. "Enterococcus Faecalis Induces Inflammatory Bowel Disease in Interleukin-10 Knockout Mice." *The American Journal of Pathology* 160 (6): 2253–57. https://doi.org/10.1016/S0002-9440(10)61172-8.

7. Boyapati, Ray K., David A. Dorward, Arina Tamborska, Rahul Kalla, Nicholas T. Ventham, Mary K. Doherty, Philip D. Whitfield, et al. 2018. "Mitochondrial DNA Is a Pro-Inflammatory Damage-Associated Molecular Pattern Released During Active IBD." *Inflammatory Bowel Diseases* 24 (10): 2113–22. https://doi.org/10.1093/ibd/izy095.

8. Caglayan, Emre, Yuxiang Liu, and Genevieve Konopka. 2022. "Neuronal Ambient RNA Contamination Causes Misinterpreted and Masked Cell Types in Brain Single-Nuclei Datasets." *Neuron* 110 (24): 4043-4056.e5. https://doi.org/10.1016/j.neuron.2022.09.010.

9. Chen, Ao, Yidi Sun, Ying Lei, Chao Li, Sha Liao, Juan Meng, Yiqin Bai, et al. 2023. "Single-Cell Spatial Transcriptome Reveals Cell-Type Organization in the Macaque Cortex." *Cell* 186 (17): 3726-3743.e24. https://doi.org/10.1016/j.cell.2023.06.009.

10. Darabi, Sajad, Shayan Fazeli, Ali Pazoki, Sriram Sankararaman, and Majid Sarrafzadeh. 2021. “Contrastive Mixup: Self- and Semi-Supervised Learning for Tabular Domain.” arXiv. http://arxiv.org/abs/2108.12296.

11. Dillon, Andrea, and David D. Lo. 2019. “M Cells: Intelligent Engineering of Mucosal Immune Surveillance.” *Frontiers in Immunology* 10 (July). https://doi.org/10.3389/fimmu.2019.01499.

12. D’incà, R., V. Annese, V. Di Leo, A. Latiano, V. Quaino, C. Abazia, M. G. Vettorato, and G. C. Sturniolo. 2006. “Increased Intestinal Permeability and NOD2 Variants in Familial and Sporadic Crohn’s Disease.” *Alimentary Pharmacology & Therapeutics* 23 (10): 1455–61. https://doi.org/10.1111/j.1365-2036.2006.02916.x.

13. Durazzi, Francesco, Claudia Sala, Gastone Castellani, Gerardo Manfreda, Daniel Remondini, and Alessandra De Cesare. 2021. “Comparison between 16S rRNA and Shotgun Sequencing Data for the Taxonomic Characterization of the Gut Microbiota.” *Scientific Reports* 11 (1): 3030. https://doi.org/10.1038/s41598-021-82726-y.

14. Elmentaite, Rasa, Natsuhiko Kumasaka, Kenny Roberts, Aaron Fleming, Emma Dann, Hamish W. King, Vitalii Kleshchevnikov, et al. 2021. “Cells of the Human Intestinal Tract Mapped across Space and Time.” *Nature* 597 (7875): 250–55. https://doi.org/10.1038/s41586-021-03852-1.

15. Elmentaite, Rasa, Alexander D. B. Ross, Kenny Roberts, Kylie R. James, Daniel Ortmann, Tomás Gomes, Komal Nayak, et al. 2020. “Single-Cell Sequencing of Developing Human Gut Reveals Transcriptional Links to Childhood Crohn’s Disease.” *Developmental Cell* 55 (6): 771-783.e5. https://doi.org/10.1016/j.devcel.2020.11.010.

16. Galeano Niño, Jorge Luis, Hanrui Wu, Kaitlyn D. LaCourse, Andrew G. Kempchinsky, Alexander Baryiames, Brittany Barber, Neal Futran, et al. 2022. “Effect of the Intratumoral Microbiota on Spatial and Cellular Heterogeneity in Cancer.” *Nature* 611 (7937): 810–17. https://doi.org/10.1038/s41586-022-05435-0.

17. Garrido-Sanz, Lidia, Miquel Àngel Senar, and Josep Piñol. 2022. “Drastic Reduction of False Positive Species in Samples of Insects by Intersecting the Default Output of Two Popular Metagenomic Classifiers.” *PLOS ONE* 17 (10): e0275790. https://doi.org/10.1371/journal.pone.0275790.

18. Gevers, Dirk, Subra Kugathasan, Lee A. Denson, Yoshiki Vázquez-Baeza, Will Van Treuren, Boyu Ren, Emma Schwager, et al. 2014. “The Treatment-Naive Microbiome in New-Onset Crohn’s Disease.” *Cell Host & Microbe* 15 (3): 382–92. https://doi.org/10.1016/j.chom.2014.02.005.

19. Gihawi, Abraham, Colin S. Cooper, and Daniel S. Brewer. 2023. “Caution Regarding the Specificities of Pan-Cancer Microbial Structure.” *Microbial Genomics* 9 (8): 001088. https://doi.org/10.1099/mgen.0.001088.

20. Gihawi, Abraham, Yuchen Ge, Jennifer Lu, Daniela Puiu, Amanda Xu, Colin S. Cooper, Daniel S. Brewer, Mihaela Pertea, and Steven L. Salzberg. 2023. “Major Data Analysis Errors Invalidate Cancer Microbiome Findings.” *mBio* 14 (5): e01607-23. https://doi.org/10.1128/mbio.01607-23.

21. Graham, Daniel B., and Ramnik J. Xavier. 2020. “Pathway Paradigms Revealed from the Genetics of Inflammatory Bowel Disease.” *Nature* 578 (7796): 527–39. https://doi.org/10.1038/s41586-020-2025-2.

22. Ha, Connie W. Y., Anthony Martin, Gregory D. Sepich-Poore, Baochen Shi, Yizhou Wang, Kenneth Gouin, Gregory Humphrey, et al. 2020. “Translocation of Viable Gut Microbiota to Mesenteric Adipose Drives Formation of Creeping Fat in Humans.” *Cell* 183 (3): 666-683.e17. https://doi.org/10.1016/j.cell.2020.09.009.

23. Haas, Kelly N., and Jeffrey L. Blanchard. 2020. “Reclassification of the Clostridium Clostridioforme and Clostridium Sphenoides Clades as Enterocloster Gen. Nov. and Lacrimispora Gen. Nov., Including Reclassification of 15 Taxa.” *International Journal of Systematic and Evolutionary Microbiology* 70 (1): 23–34. https://doi.org/10.1099/ijsem.0.003698.

24. Haberman, Yael, Timothy L. Tickle, Phillip J. Dexheimer, Mi-Ok Kim, Dora Tang, Rebekah Karns, Robert N. Baldassano, et al. 2014. “Pediatric Crohn Disease Patients Exhibit Specific Ileal Transcriptome and Microbiome Signature.” *The Journal of Clinical Investigation* 124 (8): 3617–33. https://doi.org/10.1172/JCI75436.

25. Hansen, Richard, Richard K. Russell, Caroline Reiff, Petra Louis, Freda McIntosh, Susan H. Berry, Indrani Mukhopadhya, et al. 2012. “Microbiota ofDe-NovoPediatric IBD: IncreasedFaecalibacterium Prausnitziiand Reduced Bacterial Diversity in Crohn’s But Not in Ulcerative Colitis.” *Official Journal of the American College of Gastroenterology | ACG* 107 (12): 1913. https://doi.org/10.1038/ajg.2012.335.

26. Huang, Ruili, Ivan Grishagin, Yuhong Wang, Tongan Zhao, Jon Greene, John C. Obenauer, Deborah Ngan, et al. 2019. “The NCATS BioPlanet – An Integrated Platform for Exploring the Universe of Cellular Signaling Pathways for Toxicology, Systems Biology, and Chemical Genomics.” *Frontiers in Pharmacology* 10 (April). https://doi.org/10.3389/fphar.2019.00445.

27. Hyams, J S, G D Ferry, F S Mandel, J D Gryboski, P M Kibort, B S Kirschner, A M Griffiths, A J Katz, R J Grand, and J T Boyle. 1991. “Development and Validation of a Pediatric Crohn’s Disease Activity Index.” *Journal of Pediatric Gastroenterology and Nutrition* 12 (4): 439–47.

28. Ilicic, Tomislav, Jong Kyoung Kim, Aleksandra A. Kolodziejczyk, Frederik Otzen Bagger, Davis James McCarthy, John C. Marioni, and Sarah A. Teichmann. 2016. “Classification of Low Quality Cells from Single-Cell RNA-Seq Data.” *Genome Biology* 17 (1): 29. https://doi.org/10.1186/s13059-016-0888-1.

29. Jin, Suoqin, Christian F. Guerrero-Juarez, Lihua Zhang, Ivan Chang, Raul Ramos, Chen-Hsiang Kuan, Peggy Myung, Maksim V. Plikus, and Qing Nie. 2021. “Inference and Analysis of Cell-Cell Communication Using CellChat.” *Nature Communications* 12 (1): 1088. https://doi.org/10.1038/s41467-021-21246-9.

30. Johnson, Timothy J., Subhashinie Kariyawasam, Yvonne Wannemuehler, Paul Mangiamele, Sara J. Johnson, Curt Doetkott, Jerod A. Skyberg, Aaron M. Lynne, James R. Johnson, and Lisa K. Nolan. 2007. “The Genome Sequence of Avian Pathogenic Escherichia Coli Strain O1:K1:H7 Shares Strong Similarities with Human Extraintestinal Pathogenic E. Coli Genomes.” *Journal of Bacteriology* 189 (8): 3228–36. https://doi.org/10.1128/JB.01726-06.

31. Kansal, Shivani, Anthony G Catto-Smith, Karen Boniface, Sarah Thomas, Donald J Cameron, Mark Oliver, George Alex, Carl D Kirkwood, and Josef Wagner. 2019. “The Microbiome in Paediatric Crohn’s Disease—A Longitudinal, Prospective, Single-Centre Study.” *Journal of Crohn’s and Colitis* 13 (8): 1044–54. https://doi.org/10.1093/ecco-jcc/jjz016.

32. Kao, J. S., D. M. Stucker, J. W. Warren, and H. L. Mobley. 1997. “Pathogenicity Island Sequences of Pyelonephritogenic Escherichia Coli CFT073 Are Associated with Virulent Uropathogenic Strains.” *Infection and Immunity* 65 (7): 2812–20. https://doi.org/10.1128/iai.65.7.2812-2820.1997.

33. Kaper, James B., James P. Nataro, and Harry L. T. Mobley. 2004. “Pathogenic Escherichia Coli.” *Nature Reviews Microbiology* 2 (2): 123–40. https://doi.org/10.1038/nrmicro818.

34. Kaur, Sukhpreet, Mir Yawar, P. Anil Kumar, and K. Suresh. 2014. “Hungatella Effluvii Gen. Nov., Sp. Nov., an Obligately Anaerobic Bacterium Isolated from an Effluent Treatment Plant, and Reclassification of Clostridium Hathewayi as Hungatella Hathewayi Gen. Nov., Comb. Nov.” *International Journal of Systematic and Evolutionary Microbiology* 64 (Pt_3): 710–18. https://doi.org/10.1099/ijs.0.056986-0.

35. Kim, Sandra C., Susan L. Tonkonogy, Carol A. Albright, Julia Tsang, Edward J. Balish, Jonathon Braun, Mark M. Huycke, and R. Balfour Sartor. 2005. “Variable Phenotypes of Enterocolitis in Interleukin 10-Deficient Mice Monoassociated with Two Different Commensal Bacteria.” *Gastroenterology* 128 (4): 891–906. https://doi.org/10.1053/j.gastro.2005.02.009.

36. Kleshchevnikov, Vitalii, Artem Shmatko, Emma Dann, Alexander Aivazidis, Hamish W. King, Tong Li, Rasa Elmentaite, et al. 2022. “Cell2location Maps Fine-Grained Cell Types in

Spatial Transcriptomics." *Nature Biotechnology* 40 (5): 661–71. https://doi.org/10.1038/s41587-021-01139-4.

37. Lagkouvardos, Ilias, Rüdiger Pukall, Birte Abt, Bärbel U. Foesel, Jan P. Meier-Kolthoff, Neeraj Kumar, Anne Bresciani, et al. 2016. "The Mouse Intestinal Bacterial Collection (miBC) Provides Host-Specific Insight into Cultured Diversity and Functional Potential of the Gut Microbiota." *Nature Microbiology* 1 (10): 16131. https://doi.org/10.1038/nmicrobiol.2016.131.

38. Langmead, Ben, and Steven L Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

39. Le Roy, Tiphaine, Patrick Van der Smissen, Adrien Paquot, Nathalie Delzenne, Giulio G. Muccioli, Jean-François Collet, and Patrice D. Cani. 2020. "Dysosmobacter Welbionis Gen. Nov., Sp. Nov., Isolated from Human Faeces and Emended Description of the Genus Oscillibacter." *International Journal of Systematic and Evolutionary Microbiology* 70 (9): 4851–58. https://doi.org/10.1099/ijsem.0.003547.

40. Li, Heng, and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform." *Bioinformatics* 25 (14): 1754–60. https://doi.org/10.1093/bioinformatics/btp324.

41. Liu, Chengxu, Sydney M. Finegold, Yuli Song, and Paul A. Lawson. 2008. "Reclassification of Clostridium Coccoides, Ruminococcus Hansenii, Ruminococcus Hydrogenotrophicus, Ruminococcus Luti, Ruminococcus Productus and Ruminococcus Schinkii as Blautia Coccoides Gen. Nov., Comb. Nov., Blautia Hansenii Comb. Nov., Blautia Hydrogenotrophica Comb. Nov., Blautia Luti Comb. Nov., Blautia Producta Comb. Nov., Blautia Schinkii Comb. Nov. and Description of Blautia Wexlerae Sp. Nov., Isolated from Human Faeces." *International Journal of Systematic and Evolutionary Microbiology* 58 (Pt 8): 1896–1902. https://doi.org/10.1099/ijs.0.65208-0.

42. Lötstedt, Britta, Martin Stražar, Ramnik Xavier, Aviv Regev, and Sanja Vickovic. 2023. "Spatial Host–Microbiome Sequencing Reveals Niches in the Mouse Gut." *Nature Biotechnology*, November, 1–10. https://doi.org/10.1038/s41587-023-01988-1.

43. Mahowald, Michael A., Federico E. Rey, Henning Seedorf, Peter J. Turnbaugh, Robert S. Fulton, Aye Wollam, Neha Shah, et al. 2009. "Characterizing a Model Human Gut Microbiota Composed of Members of Its Two Dominant Bacterial Phyla." *Proceedings of the National Academy of Sciences of the United States of America* 106 (14): 5859–64. https://doi.org/10.1073/pnas.0901529106.

44. Mao, Hongli, Jinlin Jia, Jinxiu Sheng, Shanfeng Zhang, Kaida Huang, Hongle Li, and Fucheng He. 2021. "Protective and Anti-Inflammatory Role of REG1A in Inflammatory

Bowel Disease Induced by JAK/STAT3 Signaling Axis." *International Immunopharmacology* 92 (March):107304. https://doi.org/10.1016/j.intimp.2020.107304.

45. McKellar, David W., Madhav Mantri, Meleana M. Hinchman, John S. L. Parker, Praveen Sethupathy, Benjamin D. Cosgrove, and Iwijn De Vlaminck. 2022. "Spatial Mapping of the Total Transcriptome by in Situ Polyadenylation." *Nature Biotechnology*, November, 1–8. https://doi.org/10.1038/s41587-022-01517-6.

46. Mikocziova, Ivana, Victor Greiff, and Ludvig M. Sollid. 2021. "Immunoglobulin Germline Gene Variation and Its Impact on Human Disease." *Genes & Immunity* 22 (4): 205–17. https://doi.org/10.1038/s41435-021-00145-5.

47. Mohebali, Nooshin, Markus Weigel, Torsten Hain, Mona Sütel, Jana Bull, Bernd Kreikemeyer, and Anne Breitrück. 2023. "Faecalibacterium Prausnitzii, Bacteroides Faecis and Roseburia Intestinalis Attenuate Clinical Symptoms of Experimental Colitis by Regulating Treg/Th17 Cell Balance and Intestinal Barrier Integrity." *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie* 167 (November):115568. https://doi.org/10.1016/j.biopha.2023.115568.

48. Neurath, Markus F. 2019. "Targeting Immune Cell Circuits and Trafficking in Inflammatory Bowel Disease." *Nature Immunology* 20 (8): 970–79. https://doi.org/10.1038/s41590-019-0415-0.

49. Ohnmacht, Caspar, Joo-Hong Park, Sascha Cording, James B. Wing, Koji Atarashi, Yuuki Obata, Valérie Gaboriau-Routhiau, et al. 2015. "The Microbiota Regulates Type 2 Immunity through RORγt+ T Cells." *Science* 349 (6251): 989–93. https://doi.org/10.1126/science.aac4263.

50. Riley, L. W., R. S. Remis, S. D. Helgerson, H. B. McGee, J. G. Wells, B. R. Davis, R. J. Hebert, et al. 1983. "Hemorrhagic Colitis Associated with a Rare Escherichia Coli Serotype." *The New England Journal of Medicine* 308 (12): 681–85. https://doi.org/10.1056/NEJM198303243081203.

51. Rutgeerts, P., K. Geboes, G. Vantrappen, J. Beyls, R. Kerremans, and M. Hiele. 1990. "Predictability of the Postoperative Course of Crohn's Disease." *Gastroenterology* 99 (4): 956–63. https://doi.org/10.1016/0016-5085(90)90613-6.

52. Saarenpää, Sami, Or Shalev, Haim Ashkenazy, Vanessa Carlos, Derek Severi Lundberg, Detlef Weigel, and Stefania Giacomello. 2023. "Spatial Metatranscriptomics Resolves Host–Bacteria–Fungi Interactomes." *Nature Biotechnology*, November, 1–10. https://doi.org/10.1038/s41587-023-01979-2.

53. Schoch, Conrad L., Stacy Ciufo, Mikhail Domrachev, Carol L. Hotton, Sivakumar Kannan, Rogneda Khovanskaya, Detlef Leipe, et al. 2020. "NCBI Taxonomy: A Comprehensive

Update on Curation, Resources and Tools." *Database: The Journal of Biological Databases and Curation* 2020 (January):baaa062. https://doi.org/10.1093/database/baaa062.

54. Sefik, Esen, Naama Geva-Zatorsky, Sungwhan Oh, Liza Konnikova, David Zemmour, Abigail Manson McGuire, Dalia Burzyn, et al. 2015. "Individual Intestinal Symbionts Induce a Distinct Population of RORγ+ Regulatory T Cells." *Science* 349 (6251): 993–97. https://doi.org/10.1126/science.aaa9420.

55. Seo, Boram, Ju Eun Yoo, Yung Mi Lee, and GwangPyo Ko. 2016. "Sellimonas Intestinalis Gen. Nov., Sp. Nov., Isolated from Human Faeces." *International Journal of Systematic and Evolutionary Microbiology* 66 (2): 951–56. https://doi.org/10.1099/ijsem.0.000817.

56. Smith, Tom, Andreas Heger, and Ian Sudbery. 2017. "UMI-Tools: Modeling Sequencing Errors in Unique Molecular Identifiers to Improve Quantification Accuracy." *Genome Research* 27 (3): 491–99. https://doi.org/10.1101/gr.209601.116.

57. Sorbara, Matthew T., and Eric G. Pamer. 2022. "Microbiome-Based Therapeutics." *Nature Reviews Microbiology* 20 (6): 365–80. https://doi.org/10.1038/s41579-021-00667-9.

58. Ståhl, Patrik L., Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, et al. 2016. "Visualization and Analysis of Gene Expression in Tissue Sections by Spatial Transcriptomics." *Science* 353 (6294): 78–82. https://doi.org/10.1126/science.aaf2403.

59. Su, Jingqian, Zhiyong Tong, Shun Wu, Fen Zhou, and Qi Chen. 2023. "Research Progress of DcR3 in the Diagnosis and Treatment of Sepsis." *International Journal of Molecular Sciences* 24 (August):12916. https://doi.org/10.3390/ijms241612916.

60. Sun, Desen, Xiaolong Ge, Shasha Tang, Yaxin Liu, Jun Sun, Yuping Zhou, Liang Luo, Zhengping Xu, Wei Zhou, and Jinghao Sheng. 2021. "Bacterial Characteristics of Intestinal Tissues From Patients With Crohn's Disease." *Frontiers in Cellular and Infection Microbiology* 11 (November):711680. https://doi.org/10.3389/fcimb.2021.711680.

61. Taras, David, Rainer Simmering, Matthew D. Collins, Paul A. Lawson, and Michael Blaut. 2002. "Reclassification of Eubacterium Formicigenerans Holdeman and Moore 1974 as Dorea Formicigenerans Gen. Nov., Comb. Nov., and Description of Dorea Longicatena Sp. Nov., Isolated from Human Faeces." *International Journal of Systematic and Evolutionary Microbiology* 52 (Pt 2): 423–28. https://doi.org/10.1099/00207713-52-2-423.

62. Tian, Luyi, Fei Chen, and Evan Z. Macosko. 2023. "The Expanding Vistas of Spatial Transcriptomics." *Nature Biotechnology* 41 (6): 773–82. https://doi.org/10.1038/s41587-022-01448-2.

63. Vester-Andersen, M. K., H. C. Mirsepasi-Lauridsen, M. V. Prosberg, C. O. Mortensen, C. Träger, K. Skovsen, T. Thorkilgaard, et al. 2019. "Increased Abundance of Proteobacteria in

Aggressive Crohn's Disease Seven Years after Diagnosis." *Scientific Reports* 9 (September):13473. https://doi.org/10.1038/s41598-019-49833-3.

64. Vignali, Dario A. A., Lauren W. Collison, and Creg J. Workman. 2008. "How Regulatory T Cells Work." *Nature Reviews Immunology* 8 (7): 523–32. https://doi.org/10.1038/nri2343.

65. Wolf, F. Alexander, Philipp Angerer, and Fabian J. Theis. 2018. "SCANPY: Large-Scale Single-Cell Gene Expression Data Analysis." *Genome Biology* 19 (1): 15. https://doi.org/10.1186/s13059-017-1382-0.

66. Wood, Derrick E., Jennifer Lu, and Ben Langmead. 2019. "Improved Metagenomic Analysis with Kraken 2." *Genome Biology* 20 (1): 257. https://doi.org/10.1186/s13059-019-1891-0.

67. Wright, Emily K., Michael A. Kamm, Shu Mei Teo, Michael Inouye, Josef Wagner, and Carl D. Kirkwood. 2015. "Recent Advances in Characterizing the Gastrointestinal Microbiome in Crohn's Disease: A Systematic Review." *Inflammatory Bowel Diseases* 21 (6): 1219–28. https://doi.org/10.1097/MIB.0000000000000382.

68. Zheng, Grace X. Y., Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, et al. 2017. "Massively Parallel Digital Transcriptional Profiling of Single Cells." *Nature Communications* 8 (1): 14049. https://doi.org/10.1038/ncomms14049.

APPENDICES

Supplementary Table 1. Results of prior research on the identified bacterial species and Crohn's disease risk.

Bacteria	Gevers et al. 2014	Kansal et al. 2019
Aggregatibacter aphrophilus		
Anaerobutyricum hallii		
Anaerostipes caccae	Decreased CD risk (genus-level confirmed)	
Anaerostipes hadrus	Decreased CD risk (genus-level confirmed)	
Bacteroides caccae		Decreased CD risk (species-level confirmed)
Bacteroides fragilis		
Bacteroides nordii		
Bacteroides ovatus		
Bacteroides sp. PHL 2737		
Bacteroides thetaiotaomicron		
Bacteroides uniformis	Decreased CD risk (species-level confirmed)	
Bacteroides xylanisolvens		
Bifidobacterium bifidum	Decreased CD risk (genus-level confirmed)	
Bifidobacterium longum	Decreased CD risk (genus-level confirmed)	
Bifidobacterium pseudocatenulatum	Decreased CD risk (genus-level confirmed)	
Blautia argi	discovered in 2018	discovered in 2018
Blautia hansenii	Decreased CD risk (genus-level confirmed)	
Blautia massiliensis	Decreased CD risk (genus-level confirmed)	
Blautia producta	Decreased CD risk (genus-level confirmed)	
Blautia sp. NBRC 113351	Decreased CD risk (genus-level confirmed)	
Blautia sp. SC05B48	Decreased CD risk (genus-level confirmed)	
Citrobacter freundii		

Citrobacter sp. RHBSTW-00986		
Clostridioides difficile		
Clostridium sp. M62/1	Decreased CD risk (genus-level confirmed)	
Coprococcus catus	Decreased CD risk (genus-level confirmed)	
Coprococcus comes	Decreased CD risk (genus-level confirmed)	Decreased CD risk (species-level confirmed)
Coprococcus sp. ART55/1	Decreased CD risk (genus-level confirmed)	
Cutibacterium acnes		Increased CD risk (species-level confirmed)
Dorea longicatena		
Dysosmobacter welbionis	discovered in 2019	discovered in 2019
Eggerthella lenta		
Enterocloster bolteae		
Enterocloster clostridioformis		
Enterococcus casseliflavus		
Enterococcus faecium		
Erysipelatoclostridium ramosum	Decreased CD risk (genus-level confirmed)	Paradoxical treds(EXCEL, SIMPER)
Escherichia coli		
Escherichia fergusonii		
Escherichia marmotae		
Escherichia sp. E4742		
Faecalibacterium prausnitzii	Decreased CD risk (species-level confirmed)	
Faecalibacterium sp. I2392	Decreased CD risk (genus-level confirmed)	
Faecalibacterium sp. I3333	Decreased CD risk (genus-level confirmed)	
Faecalibacterium sp. I3389	Decreased CD risk (genus-level confirmed)	
Faecalibacterium sp. I4179	Decreased CD risk (genus-level confirmed)	
Faecalibacterium sp. I4384	Decreased CD risk (genus-level confirmed)	
Faecalibacterium sp. IP329	Decreased CD risk (genus-level confirmed)	

Faecalitalea cylindroides		
Flintibacter sp. KGMB00164		
Fusobacterium nucleatum		Increased CD risk (genus-level confirmed)
Fusobacterium ulcerans		Increased CD risk (genus-level confirmed)
Gemella morbillorum		
Haemophilus parainfluenzae	Increased CD risk (species-level confirmed)	Increased CD risk (species-level confirmed)
Hungatella hathewayi	Decreased CD risk (genus-level confirmed)	
Klebsiella michiganensis		
Klebsiella oxytoca		
Lachnoclostridium phocaeense		
Lachnoclostridium sp. YL32		
Lachnospira eligens	Decreased CD risk (genus-level confirmed)	
Lacrimispora saccharolytica		
Lacrimispora sphenoides		
Massilistercora timonensis		
Parabacteroides distasonis	Decreased CD risk (genus-level confirmed)	
Phocaeicola dorei		
Phocaeicola vulgatus		Decreased CD risk (species-level confirmed)
Pseudomonas aeruginosa		
Qiania dongpingensis		
Roseburia hominis	Decreased CD risk (genus-level confirmed)	Decreased CD risk (species-level confirmed)
Roseburia intestinalis	Decreased CD risk (genus-level confirmed)	Decreased CD risk (genus-level confirmed)
Roseburia sp. NSJ-69	Decreased CD risk (genus-level confirmed)	Decreased CD risk (genus-level confirmed)
Rothia mucilaginosa	Increased CD risk (genus-level confirmed)	
Ruthenibacterium lactatiformans		
Sellimonas intestinalis		
Shigella dysenteriae		

Simiaoa sunii	discovered in 2021	discovered in 2021
Streptococcus intermedius		
Sutterella wadsworthensis		Increased CD risk (species-level confirmed)
Turicimonas muris	discovered in 2016	discovered in 2016
Veillonella atypica	Increased CD risk (genus-level confirmed)	Increased CD risk (species-level confirmed)
Veillonella nakazawae	Increased CD risk (genus-level confirmed)	
Veillonella parvula	Increased CD risk (genus-level confirmed)	Increased CD risk (species-level confirmed)
Wansuia hejianensis		
[Clostridium] innocuum	Decreased CD risk (genus-level confirmed)	
[Clostridium] scindens	Decreased CD risk (genus-level confirmed)	
[Ruminococcus] gnavus	Increased CD risk (species-level confirmed)	

Abstract in Korean

소아 크론병에서 공간 전체 RNA 시퀀싱 기반 멀티모달 검증을 이용한 예후 마커 개발

크론병은 장내 미생물군과 숙주 면역 체계 간 상호작용으로 발생하는 것으로 여겨지는 만성 염증성 장질환의 한 유형이다. 다수의 연구에서 크론병 환자의 장내 미생물 구성이 확인되었으며, 고처리량 전사체학 접근법의 발전으로 다양한 질병 상황에서의 세포 이질성에 대한 이해가 향상되었다. 그러나 크론병에서 조직 상주 세균과 숙주 면역 네트워크 간의 복잡한 관계는 여전히 불분명하며, 기존의 공간 전사체학 방법들은 세균 RNA 식별에 한계가 있었다. 이러한 제한점을 극복하고자 우리는 새로운 공간 숙주-미생물군 프로파일링 접근법을 개발했다. 이 방법은 공간 전사체학과 in situ 폴리아데닐화를 결합하여 숙주와 세균 RNA를 동시에 검출한다. 우리는 이를 통해 크론병 조직에서 대조군에 비해 세균 침투가 증가했음을 확인했으며, 이는 질병 예후와 연관됨을 보여준다. 또한 우리의 방법은 공간 전사체학 데이터에서 고분류학적 해상도로 세균을 검출하여 균주 수준의 식별을 가능케 한다. 이를 통해 크론병과 관련된 유익 및 병원성 미생물군을 식별했으며, 크론병에서 숙주 세포와 세균 간의 상호작용을 통한 세포 수준에서 크론병 병인에 대한 통찰을 제공한다. 본 연구는 다양한 미생물군 관련 질병에서 숙주-미생물군 상호작용을 조사하는 새로운 접근법을 제시하며, 잠재적으로 미생물군 기반 치료법과 예후 마커 개발의 새로운 전략을 제시할 수 있다.

핵심되는 말 : 장내미생물군, 멀티오믹스, 박테리아 침투, 숙주-미생물 상호작용

PUBLICATION LIST

Sooyoung Jang, Younjuong Kim, Changjun Lee, Bomi Kwon, Jihye Noh, Jai J. Jee, Sang Sun Yoon, Hong Koh, and Sowon Park. 2021. “The Effect of Formula-Based Nutritional Treatment on Colitis in a Murine Model.” Journal of Korean Medical Science 36 (50). https://doi.org/10.3346/jkms.2021.36.e342.

Sooyoung Jang, Tai hui Sun, Seunghyun Shin, Heon-Jeong Lee, Yu-Bin Shin, Ji Won Yeom, Yu Rang Park and Chul-Hyun Cho. 2024. A digital phenotyping dataset for impending panic symptoms: a prospective longitudinal study. Scientific Data 11, 1264. https://doi.org/10.1038/s41597-024-04147-6

Sooyoung Jang, JaeYong Yu, Sowon Park, Hyeji Lim, Hong Koh, Yu Rang Park. 2024. Development of Time-Aggregated Machine Learning Model for Relapse Prediction in Pediatric Crohn’s Disease. Clinical and Translational Gastroenterology. https://doi.org/10.14309/ctg.0000000000000794.