Single-Cell Transcriptomic Analysis of the Pediatric Gut in Crohn's Disease Reveals Compositional Remodeling toward an Inflammatory Cellular Landscape

Bingyao Chen
Xi'an Jiaotong-Liverpool University, Suzhou, China

Abstract:Crohn's disease is a chronic inflammatory disorder of the gut. It often begins in childhood, and its cellular basis in young patients is still not fully understood. Bulk gene expression studies report average changes in tissue, but they cannot show which cell types drive disease. In this work we reanalyze a public single-cell RNA sequencing atlas of the pediatric human gut and compare children with Crohn's disease to healthy controls. After quality control we studied 17,965 cells from 15 donors, grouped into 31 cell types that span the epithelial, stromal, vascular, and immune compartments. We used a standard and open pipeline that includes quality control, normalization, batch integration with Harmony, graph based clustering, and marker based validation of cell identity. We then measured how the cell type composition differs between disease and control tissue. Crohn's disease tissue showed a clear shift toward an inflammatory landscape. Inflammatory monocytes, IgG plasma cells, fibroblasts, and blood vessel cells all increased in disease, while mature B cells and absorptive enterocytes decreased. The expansion of IgG plasma cells, monocytes, and stromal cells matches a known pathogenic module that is linked to poor drug response in adult Crohn's disease. The results give a clear, cell level picture of pediatric Crohn's disease and provide a reproducible base for biomarker work and for future spatial and predictive studies.

Keywords: RNA Sequencing; Crohn's Disease; Inflammatory Bowel Disease; Pediatric; Cell Type Composition; Gut Mucosa; Transcriptomics.

1. Introduction

Inflammatory bowel disease, or IBD, is a long term immune disease of the digestive tract. Crohn's disease is one of its two main forms. A global study estimated about 6.8 million cases of IBD worldwide in 2017, and the number keeps rising [1]. A large share of patients first develop disease in childhood or in their teenage years. Pediatric Crohn's disease often runs a more severe course than adult onset disease, with deeper inflammation and a higher need for strong drugs. Children also grow and develop while the disease is active, so the stakes are high. A clear view of the cells that drive disease in young patients is therefore important for better care.

Transcriptomics, the study of gene expression across many genes at once, has become central to IBD research. Early work used bulk RNA sequencing, which measures the average expression of a whole tissue sample. Bulk data are useful, but they mix signals from many cell types. A rise in an inflammatory gene may come from immune cells, from the gut lining, or from connective tissue, and bulk data cannot tell them apart. Single-cell RNA sequencing solves this problem by measuring gene expression in thousands of individual cells. It reveals which cell types change in disease and which new cell states appear [3], [4]. This resolution is the key to understanding a complex tissue such as the inflamed gut.

Several single-cell studies have reshaped the view of IBD in adults. Work on Crohn's disease lesions found a group of cell types that act together and that links to poor response to anti-TNF therapy [3]. Work on ulcerative colitis built a detailed map of the colon and found new disease associated fibroblast and epithelial states [4], [6]. Large atlases have also mapped the healthy and diseased human gut across age and location [5]. Yet the pediatric gut in Crohn's disease has received less attention than adult tissue, and a focused, cell level comparison between young patients and healthy children is still valuable.

In this paper we reanalyze a public single-cell atlas of the pediatric human gut and compare

Crohn's disease with healthy control tissue. We build the analysis on open and widely used tools so that it is easy to reproduce. Our main contributions are as follows.

1)We build a clean, integrated single-cell map of 17,965 cells from the pediatric gut that covers 31 cell types across the epithelial, stromal, vascular, and immune compartments.

2)We validate cell identity with canonical marker genes and then quantify how the cell type composition differs between Crohn's disease and healthy tissue.

3)We show that the disease tissue shifts toward an inflammatory landscape that echoes a known pathogenic module from adult Crohn's disease, and we discuss how this base supports future spatial and machine learning work.

2. Related Work

2.1 Single-Cell Studies in Inflammatory Bowel Disease

Single-cell studies have changed how researchers view IBD. Martin and colleagues studied Crohn's disease lesions and found a group of cell types that appear together, including IgG plasma cells, inflammatory mononuclear phagocytes, activated T cells, and activated stromal cells [3]. They named this group the GIMATS module and showed that its presence links to poor response to anti-TNF therapy. Smillie and colleagues built a single-cell atlas of the human colon in ulcerative colitis and found inflammation associated fibroblasts and rewired cell signaling [4]. Parikh and colleagues described how the colonic epithelium changes in IBD [6]. Elmentaite and colleagues mapped the human intestinal tract across space and time, including both healthy and Crohn's disease gut in children and adults [5]. More recent work has even reached the preclinical stage, where early immune changes appear before clear symptoms [13]. Our study builds on this line of work and focuses on the pediatric gut.

2.2 Tools for Single-Cell Analysis

A mature set of open tools now supports single-cell analysis. Scanpy provides a scalable framework for quality control, normalization, dimension reduction, and visualization [8]. Harmony corrects batch effects across donors and studies so that the same cell types from different samples group together [9]. The Leiden algorithm finds well connected clusters in the cell neighbor graph [10], and UMAP projects the high dimensional data into two dimensions for inspection [11]. These tools are standard, well documented, and free, which makes a study built on them easy to repeat and to extend.

3. Materials and Methods

3.1 Data Source

We used a public single-cell RNA sequencing dataset of the pediatric human gut, aged four to fourteen years, distributed through the CZ CELLxGENE Discover platform as part of a gut cell atlas [5], [7]. The dataset includes children with Crohn's disease and healthy controls. We downloaded the original annotated file in H5AD format. The data carry donor labels, a disease label for each cell, and curated cell type annotations from the original study. After quality control we analyzed 17,965 cells from 15 donors. The disease label split the cells into 9,430 Crohn's disease cells and 8,535 normal cells, which gave a balanced comparison. Table 1 summarizes the dataset and cohort. We did not generate any new data, and all analysis used public data and open tools.

Table 1. Dataset and Cohort Summary

Property	Value
Tissue and cohort	Pediatric human gut, 4 to 14 years
Data source	CZ CELLxGENE Discover, gut cell atlas
Conditions compared	Crohn's disease vs. normal
Cells after quality control	17,965
Crohn's disease cells	9,430
Normal cells	8,535
Donors (batches)	15
Annotated cell types	31

3.2 Preprocessing and Quality Control

We processed the data with Scanpy [8]. Because the shared matrix stored normalized values, we recovered the raw counts from the stored raw layer before any further step. We mapped gene identifiers to gene symbols so that marker genes could be read by name. We flagged mitochondrial genes by their symbol prefix and computed standard quality metrics for each cell. We removed cells with fewer than 200 detected genes and cells with a mitochondrial read fraction above 20 percent. We also removed cells with an unusually high gene count, above 7,000, since these can be doublets. We removed

genes found in fewer than three cells. These thresholds are common and conservative.

3.3 Normalization and Feature Selection

We stored the raw counts in a separate layer for later use. We normalized each cell to a total count of ten thousand and applied a log transform. We kept the full gene set in a reference layer for marker analysis and plotting. We then selected the 2,000 most highly variable genes and used this set for dimension reduction. This focuses the analysis on the genes that carry the most signal and reduces noise.

3.4 Integration and Clustering

We scaled the selected genes and ran principal component analysis to obtain 50 components. To remove batch effects between donors we applied Harmony on the donor label [9]. We built a neighbor graph from the corrected components and clustered the cells with the Leiden algorithm at a resolution of 1.0 [10]. We projected the cells into two dimensions with UMAP for visualization [11]. We checked that cells from both conditions mixed across the shared clusters, which shows that the integration worked.

E. Cell Type Annotation and Validation

The atlas provides curated cell type labels from the original study, which we used as the main annotation. This gave 31 cell types across the epithelial, stromal, vascular, and immune compartments. To confirm the identity of the broad lineages, we computed marker genes for each Leiden cluster with the Wilcoxon rank sum test and inspected canonical markers in a dot plot. We checked T cell markers such as CD3D and CD3E, B cell markers such as MS4A1 and CD79A, plasma cell markers such as MZB1 and CD27, and myeloid markers such as LYZ, CD14, and S100A8. The clusters matched the expected lineages, which supports the annotation.

3.5 Composition Analysis

To compare the two conditions we computed the percentage of each cell type within each condition. For every cell type we report its share of all cells in Crohn's disease tissue and its share in normal tissue. We then ranked cell types by the size and direction of the change. This composition view is simple and robust, and it directly answers which cell types expand or shrink in disease. We did not perform any new wet lab experiments, and the full pipeline can be repeated on the public data.

4. Results

4.1 A Single-Cell Map of the Pediatric Gut

After quality control and integration, the 17,965 cells formed a clear map with well separated groups. The 31 annotated cell types covered the main parts of the gut. The epithelial compartment included enterocytes, goblet cells, tuft cells, enteroendocrine cells, crypt stem cells, transit amplifying cells, and M cells. The stromal compartment included fibroblasts, myofibroblasts, and glial cells. The vascular compartment included artery, vein, and lymphatic endothelial cells and pericytes. The immune compartment was rich and included several T cell states, B cells and memory B cells, plasma cells split by antibody class, monocytes, dendritic cells, and mast cells. Figure 1 shows the map colored by cell type.

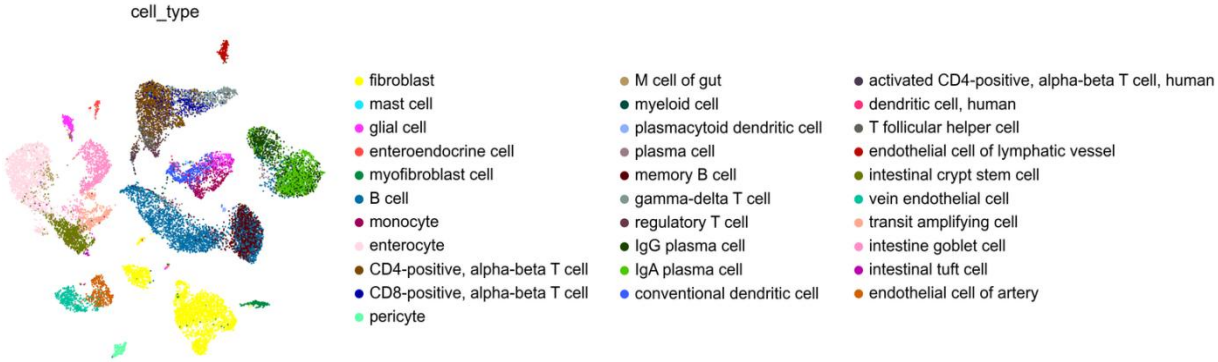

Figure 1. UMAP Embedding of 17,965 Pediatric Gut Cells after Batch Integration, Colored by the 31 Annotated Cell Types. The Map Covers Epithelial, Stromal, Vascular, and Immune Compartments.

Figure 2 shows the same cells colored by their Leiden cluster. Clustering at a resolution of 1.0 produced 29 clusters. The graph based clustering recovered the major lineage groups, which we read together with the curated cell type labels from the atlas. The clusters were compact and

well separated, which is a sign of clean preprocessing and integration.

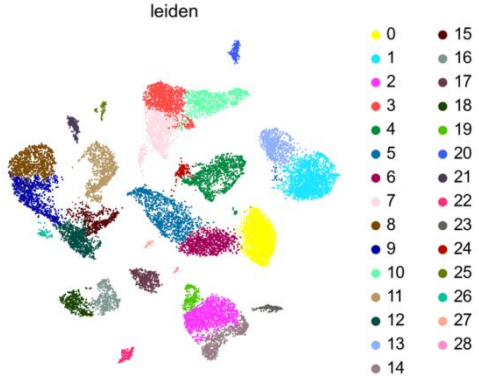

Figure 2.The Same UMAP Colored by Leiden Cluster. Clustering Produced 29 Compact and Well Separated Groups that Align with the Major Cell Lineages.

Figure 3 shows the same map colored by condition. Cells from Crohn's disease and from healthy gut mixed across most clusters. This mixing shows that the batch integration removed donor specific effects, so that the same cell type from different donors grouped together. At the same time, some clusters leaned toward one condition, which is an early sign of the composition changes that we measure below.

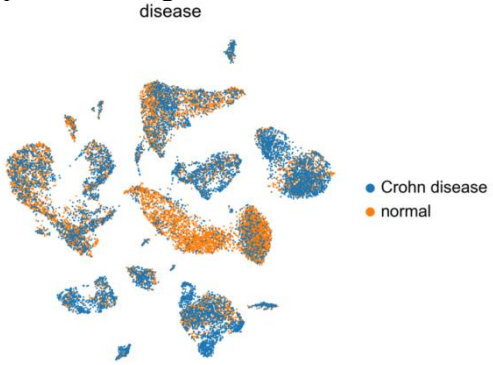

Figure 3. The Same UMAP Colored by Condition. Crohn Disease and Normal Cells Mix Across Most Clusters, Which Confirms That Batch Integration Worked, While Some Regions Are Skewed Toward One Condition.

4.2 Marker Based Validation of Cell Identity

To confirm the identity of the major lineages we computed marker genes for each of the 29 Leiden clusters with the Wilcoxon rank sum test. Figure 4 shows the top ranked genes for each cluster. For every cluster, the strongest markers were genes that are specific to one lineage. T cell clusters were marked by CD3D and CD3E. The B cell cluster was marked by MS4A1, CD79A, and CD19. Plasma cell clusters were marked by MZB1 and CD27. Myeloid clusters were marked by LYZ, CD14, FCN1, S100A8, and S100A9. Each cluster matched a known lineage, which gives confidence in the annotation used for the composition analysis. Table II lists the canonical markers used for the major lineages.

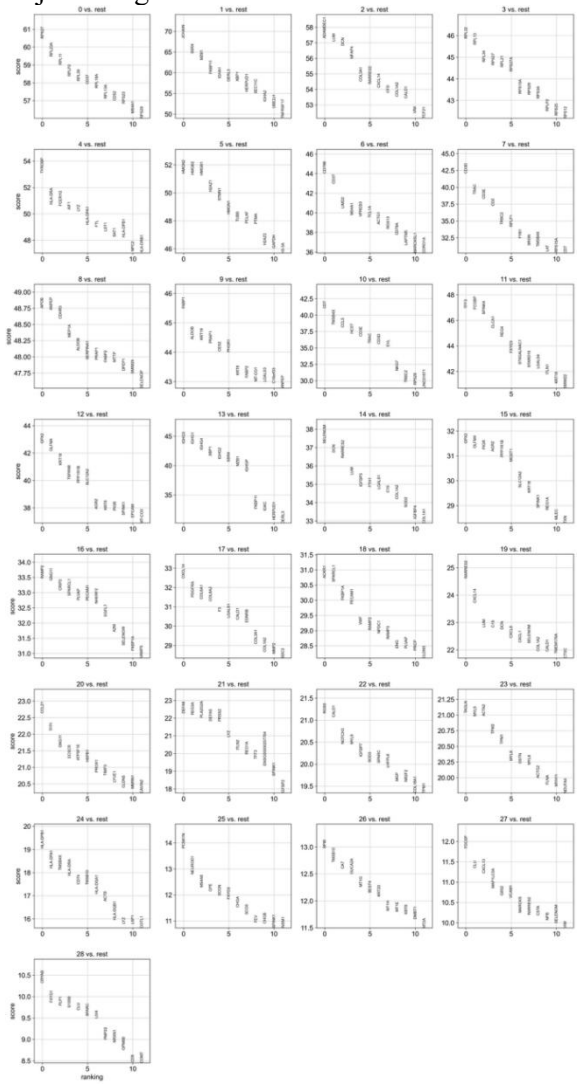

Figure 4. Top Ranked Marker Genes for Each of the 29 Leiden Clusters, Found with the Wilcoxon Rank Sum Test. Each Panel Shows the Genes Most Specific to One Cluster, and the Strongest Markers Are Lineage Specific.

Table 2. Canonical Markers Used for Lineage Validation

Lineage	Marker genes
Epithelial	KRT8, KRT18
Enterocyte	APOA1
BEST4+ enterocyte	BEST4
Inflammation-assoc. fibroblast	TNFSF14
T cell	CD3D, CD3E
B cell	MS4A1, CD79A, CD19

Lineage	Marker genes
Plasma cell	MZB1, CD27
Myeloid / Monocyte	LYZ, CD14, FCN1, S100A8, S100A9

4.3 Compositional Remodeling in Crohn's Disease

The cell type composition differed clearly between Crohn's disease and healthy gut. Figure 5 shows the share of each cell type in each condition. Several immune, stromal, and vascular cell types expanded in disease, while mature B cells and absorptive epithelial cells shrank.

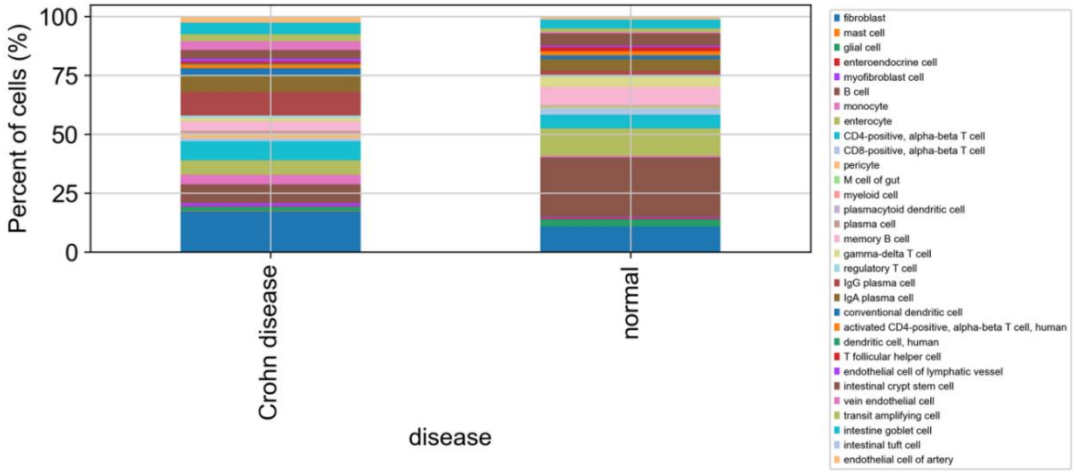

Figure 5. Cell Type Composition in Crohn Disease Versus Normal Gut. Each Bar Shows the Percentage of Each Cell Type Within a Condition. Inflammatory and Stromal Cells Expand in Disease While Mature B cells and Enterocytes Contract.

Among immune cells, the change was striking. Inflammatory monocytes rose from 0.6 percent of all cells in normal tissue to 3.9 percent in Crohn's disease, an increase of about six fold. IgG plasma cells rose from 2.3 percent to 10.0 percent, an increase of more than four fold. Conventional dendritic cells and plasmacytoid dendritic cells also rose, and regulatory T cells increased from 0.4 percent to 1.1 percent. In contrast, mature B cells fell sharply from 25.4 percent in normal tissue to 8.1 percent in disease, and memory B cells fell from 7.8 percent to 4.0 percent. Gamma-delta T cells also dropped. The pattern points to a shift away from resting B cells and toward antibody secreting plasma cells and active myeloid cells.

The stromal and vascular compartments also expanded in disease. Fibroblasts rose from 10.9 percent to 17.3 percent, and myofibroblasts and pericytes increased as well. Blood vessel cells showed a clear gain, with vein endothelial cells rising from 0.5 percent to 3.8 percent and artery endothelial cells rising from 1.3 percent to 2.6 percent. This vascular gain fits the growth of new blood vessels that is common in chronically inflamed tissue. In the epithelial compartment, absorptive enterocytes fell from 11.6 percent to 6.2 percent and crypt stem cells fell from 5.1 percent to 3.7 percent, while transit amplifying cells and goblet cells rose modestly. The drop in mature enterocytes together with shifts in progenitor states suggests a disturbed and remodeling epithelium under constant immune pressure.

Table III lists the full composition for all 31 cell types, sorted by the size of the change. The largest single shift was the loss of mature B cells, followed by the gain of IgG plasma cells and fibroblasts. The table makes the overall direction clear, with inflammatory, stromal, and vascular cells gaining ground while resting B cells and mature epithelial cells lose ground.

Table 3. Cell Type Composition, Crohn's Disease vs. Normal

Cell type	Crohn (%)	Normal (%)	Change (pp)
B cell	8.06	25.37	-17.31
IgG plasma cell	9.98	2.30	+7.68
fibroblast	17.32	10.86	+6.46
enterocyte	6.21	11.63	-5.42
memory B cell	3.97	7.79	-3.82
vein endothelial cell	3.81	0.48	+3.33
monocyte	3.92	0.61	+3.31
gamma-delta T cell	1.42	4.17	-2.75
CD4-positive, alpha-beta T cell	8.09	5.76	+2.33
IgA plasma cell	7.20	4.93	+2.27
intestinal crypt stem cell	3.65	5.14	-1.49
CD8-positive, alpha-beta	1.24	2.61	-1.37

T cell			
endothelial cell of artery	2.60	1.28	+1.32
glial cell	1.75	3.06	-1.31
transit amplifying cell	2.92	1.61	+1.31
conventional dendritic cell	2.80	1.53	+1.27
intestine goblet cell	4.92	3.68	+1.24
pericyte	1.40	0.29	+1.11
T follicular helper cell	0.87	1.76	-0.89
regulatory T cell	1.09	0.40	+0.69
myofibroblast cell	1.27	0.59	+0.68
endothelial cell of lymphatic vessel	1.26	0.63	+0.63
plasma cell	0.67	0.18	+0.49
dendritic cell, human	0.41	0.05	+0.36
plasmacytoid dendritic cell	0.21	0.04	+0.17
activated CD4-positive, alpha-beta T cell, human	1.60	1.76	-0.16
intestinal tuft cell	0.05	0.20	-0.15
mast cell	0.12	0.01	+0.11
enteroendocrine cell	0.30	0.36	-0.06
M cell of gut	0.66	0.71	-0.05
myeloid cell	0.23	0.21	+0.02

Change is Crohn minus Normal in percentage points (pp). Rows are sorted by absolute change.

4.4 An Inflammatory Stromal, Myeloid, and Plasma Cell Signature

Taken together, the three cell types that expanded most in disease were IgG plasma cells, inflammatory monocytes, and fibroblasts, joined by a clear gain in blood vessel cells. This combination is notable. The pairing of IgG plasma cells, inflammatory mononuclear phagocytes, and activated stromal cells closely matches the GIMATS module that was defined in adult Crohn's disease and that was linked to poor response to anti-TNF therapy [3]. Our independent reanalysis of pediatric tissue recovers the same broad signature from a different cohort and a different age group. This agreement strengthens the view that a stromal, myeloid, and plasma cell axis is a core feature of active Crohn's disease, and it suggests that the same axis is already present in children.

The plasma cell change is also informative on its own. In healthy tissue, IgA plasma cells were the main antibody secreting cells, which fits the normal role of IgA at the gut surface. In disease, IgG plasma cells expanded far more than IgA plasma cells, so the balance shifted toward IgG. IgG plasma cells reached 10.0 percent of all cells in disease and were the single largest plasma cell group, while IgA plasma cells rose only modestly. A move from IgA toward IgG in the gut is a known sign of mucosal inflammation, since IgG is linked to stronger and more damaging immune responses. The single-cell view makes this shift easy to see and to measure.

5. Discussion

Our results give a clear, cell level picture of pediatric Crohn's disease. Single-cell data show that the inflamed gut is not simply a healthy gut with more of every cell. Instead, the tissue is remodeled. Inflammatory monocytes, IgG plasma cells, fibroblasts, and blood vessel cells expand, while mature B cells and absorptive enterocytes contract. A bulk measurement would blur these opposite movements into a weak average and would hide the true structure of the change.

The findings agree with and connect several lines of prior work. The expansion of an IgG plasma cell, monocyte, and stromal axis matches the GIMATS module from adult Crohn's disease [3]. The growth of fibroblasts fits reports of disease associated fibroblasts in IBD [4]. The reduction in mature enterocytes fits known epithelial disruption in IBD [6]. The vascular gain fits the angiogenesis that often accompanies chronic inflammation. The fact that these patterns appear in a pediatric cohort, through an independent reanalysis, adds weight to the idea that they are general features of active disease rather than features of one cohort.

This study has limits that should be stated plainly. It is a reanalysis of one public dataset, so the results depend on the choices and the sampling of the original study. The comparison is based on cell type composition, which is descriptive and associative. It shows which cell types differ between groups, but it does not prove cause, and it does not by itself identify the genes that drive each change. The data are cross sectional, so we cannot follow how the tissue changes over time or in response to treatment. The annotation uses the labels from the source atlas, which we validated for the major lineages but did not redefine. Finally, the cohort is pediatric, so the results may not transfer directly to adult disease.

Several clear next steps follow from this work. Differential expression within each expanded cell type would move the analysis from composition to mechanism and would point to candidate genes and pathways. Cell to cell

communication analysis would show how the expanded stromal, myeloid, and plasma cells signal to one another. Spatial transcriptomics would place these cell states back into the tissue and show where the inflammatory niches form [15]. Machine learning on the same features could build a compact and non-invasive marker panel and could group patients into subtypes for treatment planning [14], [16]. The clean and reproducible base built here makes each of these steps practical.

The clinical value of this kind of analysis is practical. A cell level map of disease can guide the search for blood or tissue markers, because a marker is more trustworthy when it can be traced to a defined cell type that expands in disease. The pediatric focus matters because children are underrepresented in many studies, yet they carry a heavy disease burden and need treatment choices that fit their biology. Because our pipeline uses only public data and open tools, other groups can repeat it, apply it to their own cohorts, and compare results directly. Reproducibility of this kind is important for building trust in single-cell findings and for moving them toward the clinic.

6. Conclusion and Future Work

We presented a single-cell transcriptomic analysis of the pediatric gut in Crohn's disease, based on a public atlas and built with open tools. After quality control and batch integration we studied 17,965 cells across 31 cell types and compared disease with healthy tissue. The disease tissue showed a clear shift toward an inflammatory landscape, with strong expansion of inflammatory monocytes, IgG plasma cells, fibroblasts, and blood vessel cells, and with loss of mature B cells and absorptive enterocytes. This signature recovers, in children and from independent data, a pathogenic module that was first defined in adult Crohn's disease. The analysis is reproducible and provides a base for future work on differential expression, cell communication, spatial mapping, and machine learning. We hope it helps move pediatric inflammatory disease research toward a more precise and cell aware understanding.

References

[1] GBD 2017 Inflammatory Bowel Disease Collaborators, "The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017," Lancet Gastroenterol. Hepatol., vol. 5, no. 1, pp. 17-30, 2020.

[2] D. B. Graham and R. J. Xavier, "Pathway paradigms revealed from the genetics of inflammatory bowel disease," Nature, vol. 578, no. 7796, pp. 527-539, 2020.

[3] J. C. Martin et al., "Single-cell analysis of Crohn's disease lesions identifies a pathogenic cellular module associated with resistance to anti-TNF therapy," Cell, vol. 178, no. 6, pp. 1493-1508, 2019.

[4] C. S. Smillie et al., "Intra- and inter-cellular rewiring of the human colon during ulcerative colitis," Cell, vol. 178, no. 3, pp. 714-730, 2019.

[5] R. Elmentaite et al., "Cells of the human intestinal tract mapped across space and time," Nature, vol. 597, no. 7875, pp. 250-255, 2021.

[6] K. Parikh et al., "Colonic epithelial cell diversity in health and inflammatory bowel disease," Nature, vol. 567, no. 7746, pp. 49-55, 2019.

[7] CZ CELLxGENE Discover, single-cell data platform, Chan Zuckerberg Initiative. [Online]. Available: https://cellxgene.cziscience.com

[8] F. A. Wolf, P. Angerer, and F. J. Theis, "SCANPY: large-scale single-cell gene expression data analysis," Genome Biol., vol. 19, art. 15, 2018.

[9] I. Korsunsky et al., "Fast, sensitive and accurate integration of single-cell data with Harmony," Nat. Methods, vol. 16, pp. 1289-1296, 2019.

[10] V. A. Traag, L. Waltman, and N. J. van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," Sci. Rep., vol. 9, art. 5233, 2019.

[11] L. McInnes, J. Healy, and J. Melville, "UMAP: uniform manifold approximation and projection for dimension reduction," arXiv:1802.03426, 2018.

[12] T. Stuart et al., "Comprehensive integration of single-cell data," Cell, vol. 177, no. 7, pp. 1888-1902, 2019.

[13] D. Kioroglou et al., "Single-cell analysis reveals significant transcriptomic alterations in preclinical Crohn's disease," Front. Immunol., vol. 16, art. 1539830, 2025.

[14] I. S. Stafford et al., "A systematic review of artificial intelligence and machine learning applications to inflammatory bowel disease," Inflamm. Bowel Dis., vol. 28, no. 10, pp.

1573-1583, 2022.
[15] Y. Miyamoto et al., "Spatial transcriptomics in autoimmune rheumatic disease: potential clinical applications and perspectives," Inflammation and Regeneration, vol. 45, art. 6, 2025.
[16] M. H. D. Nazari, M. Ghorbaninejad, S. Shahrokh, and A. Meyfour, "Biomarker discovery for non-invasive diagnosis of inflammatory bowel disease using blood transcriptomics," Front. Immunol., vol. 16, art. 1570374, 2025.
[17] F. Zhang et al., "Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry," Nat. Immunol., vol. 20, no. 7, pp. 928-942, 2019.