An Integrative Single-cell and Spatial Transcriptomic Analysis of Fistulating Crohn's Disease

Colleen Georgette Chantelle McGregor

Wolfson College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2024

I dedicate this thesis to my parents, George and Lucille McGregor,
for their enduring love and support

Acknowledgements

I would like to extend my thanks to Professor Alison Simmons for her supervision and for providing me with the opportunity to pursue this project. I am sincerely grateful for her support, guidance and for fostering a collaborative and supportive laboratory environment. I also wish to acknowledge Professor Nicola Ternette for her secondary supervision and Dr Rebecca Powell Doherty for encouraging me to pursue my academic interests.

My thanks go to the members of the Simmons Group for their insightful discussions, training, and assistance throughout this work. A special thanks to Dr Agne Antanaviciute, with whom I worked closely on the computational methods and analyses. Her collaboration, support and advice have been invaluable.

I am also very appreciative of my collaborators, Dr Eve Fryer, Professor Michael Vieth and Dr Ruchi Tandon, for their histopathological analyses and for providing key research samples. Additionally, I thank my surgical colleagues Mr Bruce George, Mr Mark Bignell, and Miss Kat Baker from Oxford University Hospitals (OUH) for their help in providing research samples. I am also grateful to the ORNID nurses and the Oxford TGU Biobank, led by Simona Fourie and James Chivenga respectively, for their support in sample collection. Importantly, this work would not have been possible without the generous donation of samples from patients at OUH and Bayreuth Hospital, for which I am truly thankful.

I also wish to acknowledge the members of the following facilities for their expertise and collaboration: Craig Waugh and Kevin Clark (WIMM Flow Cytometry Lab), Maria Greco,

Rhea Kujawa and Paola Vargas Gutierrez (WIMM Advanced Single Cell OMICS Facility), Tim Rostron (WIMM Sequencing Laboratory), Jana Koth and Cyril Lai (WIMM Imaging Facility) and Stephanie Jones and Sorayya Moradi (Oxford Centre for Histopathology Research).

I am grateful for the financial support that has made my DPhil work possible, particularly from the Wellcome Trust, Medical Research Council, NIHR and the Oxfordshire Health Services Research Committee.

I would like to express my deepest gratitude to my husband, Ben, my family and close friends. Your unwavering love, support, encouragement and prayers have been invaluable throughout this journey. The University of Oxford's motto, *Dominus illuminatio mea*, meaning 'The Lord is my light,' has been a guiding truth in my life. Above all, I thank Jesus for being my constant light and source of strength.

Attributions

This thesis represents the results of my original work, alongside collaborations and contributions from various researchers during my DPhil. Below is a summary of my chapter-by-chapter contributions:

Chapter 1: All my own work, with the exception of the following figures re-used with permission:

Figure 1.1: Re-used with publisher permission from Hirten et al. (2018) and Panés et al. (2017)

Figure 1.3: My own work, previously published (McGregor CGC et al., 2022)

Figure 1.4: Re-used with publisher permission from Corridoni et al. (2020)

Chapter 2: All content is exclusively my own work, with a review of the computational methods (Methods 2.13.2) provided by Agne Antanaviciute (AA).

Chapter 3: I independently conceived and executed this work, supervised by Alison Simmons (AS). I consented, collected (with ORNID & TGU), processed and stored samples from patients included in this study and assembled the clinical metadata. I received technical and practical advice from Tarun Gupta (TG) regarding epithelial, stromal and immune cell isolation workflows, flow cytometry, and single-cell workflows. I conceived, performed, optimised, and analysed the digestion protocol experiments in discussion with TG. I generated all figures from this data. The pilot LeviCell experiment was conducted by Zoe Christoforidou (ZC), Ana Sousa Geros (ASG), and Paulina Siejka-Zielinska (PSZ) and analysed by AA. I conceived, optimised and performed the subsequent LeviCell experiments on stromal isolations. I independently generated the single-cell dataset (epithelial, stromal and immune) with assistance on cellular isolation and single-cell workflow (until the pre-

loading step) from TG, Zinan Yin (ZY), and Anna Aulicino (AAu). I performed the quality control and library preparation steps independently. All single-cell data generated in this chapter were analysed by AA in collaboration with me.

I conducted all sample quality control experiments (RNA extraction, RNA integrity number and DV200) for spatial transcriptomics (ST) and optimisation of sample processing independently. I conceived and performed the tissue optimisation pilot experiment and fresh frozen ST pilot experiment with assistance from TG. I analysed and generated figures from the tissue optimisation data. I independently conceived, optimised and performed the pilot Visium FFPE ST experiment, collaborating with ZY. I conducted the subsequent quality control and library generation independently. Slide sectioning and preparation for all ST runs (Visium, MERSCOPE, Xenium) were performed in collaboration with Esther Bridges (EB). Maria Greco (MG) and Rhea Kujawa (RK) ran the MERSCOPE workflow in the facility, in discussion with me.

Chapter 4: I conceived the pilot single-cell experiment in collaboration with TG and AA. Subsequent runs were independently conceived and performed by me, in discussion with AA and AS, with assistance on experimental steps requiring more than one person. I prepared all libraries for sequencing. All single-cell data generated in this chapter were analysed by AA, except for the T-cell repertoire analysis, which was conducted by Chloe Hyun-Jung Lee (CHL) in discussion with me, AA and AS. AA and CHL generated all raw figures derived from this dataset. I analysed the FACS data and generated the corresponding figures.

Chapter 5: EB and Marta Jagielowicz (MJ) processed formalin blocks. I sought and obtained ethics to access archived samples from OUH and Bayreuth Hospital. I performed slide sectioning, H&E staining and scanning of fresh frozen and formalin fixed samples with some assistance from EB. Ruchi Tandon (RT), Michael Vieth (MV) and Eve Fryer (EF) provided histopathological analysis in collaboration with me. I independently designed and performed Visium FFPE ST experiments, either solo or with assistance from ZY and prepared all libraries for sequencing. I conceived Xenium

runs in collaboration with AA. I optimised and conducted the Xenium workflow with David Fawkner-Corbett (DFC), Verena Lentsch (VL), TG and MG. I performed and analysed all validation immunohistochemistry experiments with technical advice from EB. I independently conceived, performed and analysed all validation qPCR experiments and generated the corresponding figures. All computational analysis in this chapter was conducted by AA, in discussion with me, who also generated the raw images.

Unless otherwise stated, I generated all results figures, except for the raw figures from ST data and single-cell data, which were generated by AA or CHL. I assembled all the raw figures into thesis figures.

Chapter 6: All content is exclusively my own work.

Abstract

Crohn's disease (CD) often leads to fistula development in approximately 40% of patients, driven by sustained, transmural inflammation of the bowel wall. Despite advanced treatments, one-third of patients experience fistula recurrence, highlighting a significant clinical challenge. The pathogenesis of fistulating CD remains poorly defined and is considered multifactorial, involving complex interactions among immune responses, cellular dynamics, genetic factors and microbial contributions. While processes such as epithelial-to-mesenchymal transition, upregulation of matrix metalloproteinases, pro-inflammatory markers like TNF-alpha and IL-13 and NOD2 polymorphisms have been implicated, comprehensive molecular studies of fistulating CD are lacking, limiting the development of targeted therapies.

This project addresses this gap by investigating the mechanisms underlying diverse CD fistulae presentations using integrative multimodal single cell RNA-sequencing and spatial transcriptomics (ST). Subcellular resolution *in situ* ST, alongside full-transcriptome ST profiling, was applied to over 60 fistulae and more than 6 million imaged cells, revealing common pathological features across various CD fistula phenotypes. The analysis identified novel fibroblast subpopulations at the leading edge of fistulae, defined by patterning-associated transcription factors and driven by dysregulated morphogen networks in both epithelialised and non-epithelialised fistulae. Additionally, a high-resolution view of the re-epithelialisation process within fistulae and the impaired stromal-epithelial crosstalk that hinders epithelial regeneration, was elucidated.

The findings offer a detailed multimodal atlas of fistulating CD at an unprecedented resolution, serving as a key resource for developing disease models and identifying novel therapeutic targets. This work represents a significant advance in understanding the molecular and cellular mechanisms driving fistula formation in CD, with the potential to inform future strategies aimed at improving patient outcomes.

Contents

List of Figures

List of Abbreviations

ADT	Antibody-derived tagged
APPs	Analysis Protocol Packages
ATAC-Seq	Assay for Transposase-Accessible Chromatin with sequencing
BMP	Bone morphogenetic protein
CCL/CCR	C-C motif ligand or receptor
CD	Crohn's disease
CD20	Cluster of differentiation 20 antigen
CD45	Cluster of differentiation 45 antigen
CD68	Cluster of differentiation 68 antigen
CD90	Cluster of differentiation 90 antigen
cDNA	Complementary DNA
CEACAM	Carcinoembryonic Antigen-Related Cell Adhesion Molecule
ChIP-Seq	Chromatin Immunoprecipitation Sequencing
CITE	Cellular Indexing of Transcriptomes and Epitopes
CK	Cytokeratin
CLPF	Colonic lamina propria fibroblast
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CT	Computed Tomography
CXCL/CXCR	C-X-C motif ligand or receptor
DD	Diverticular disease
DEGs	Differentially expressed genes

DKK	Dickkopf-related protein
DSS	Dextran sodium sulphate
DV200	Percentage of RNA fragments longer than 200 nucleotides
ECM	Extracellular matrix
ED	Extra domain
EEC	Enteroendocrine cell
EMT	Epithelial-to-mesenchymal transition
EPCAM	Epithelial cell adhesion molecule
Ets	E-twenty-six
FACS	Fluorescence-activated cell sorting
FAS	Fistula-Associated Stroma
FDR	False discovery rate
FFPE	Formalin-fixed paraffin embedded
FIST	Fistulating
FRC	Fibroblastic reticular cell
FT	Full thickness
FZD	Frizzled
GEX	Gene expression
GO	Gene ontology
GWAS	Genome wide association studies
H&E	Haematoxylin & Eosin
HC	Healthy control
HTO	Hashtag oligonucleotide
IBD	Inflammatory Bowel Disease
IEC	Intestinal epithelial cell
IEL	Intraepithelial lymphocyte
IEO	Intestinal epithelial organoid

IFC	Integrated Fluidic Circuit
IFN	Interferon
IHC	Immunohistochemistry
IL	Interleukin
ILC	Innate lymphoid cell
INF	Inflammatory
ISC	Intestinal stem cell
JAK	Janus kinase
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LGR5	Leucine-rich repeat-containing G-protein coupled receptor 5
MAT	Mesenteric adipose tissue
MDP	Muramyl dipeptide
MKI67	Marker of proliferation Ki-67
MMP	Matrix metalloproteinase
MP	Muscularis propria
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MSCs	Mesenchymal stem cells
MUC	Mucosa
NOD	Nucleotide-binding oligomerization domain-containing protein
OR	Odds ratio
PAMP	Pathogen associated molecular pattern
PCA	Principal component analysis
PDGF	Platelet-derived growth factor
QC	Quality control
qPCR	Quantitative polymerase chain reaction
rRNA	Ribosomal RNA

scRNA-seq	Single cell RNA sequencing
SM	Submucosa
ST	Spatial transcriptomics
STRI	Stricturing
TA	Transit amplifying
TC	Transitional cell
TF	Transcription factor
TGF	Transforming growth factor
Th	T helper
TIMP	Tissue inhibitors of MMP
TNBS	Trinitrobenzene sulfonic acid
TNF	Tumour necrosis factor
tSNE	t-distributed Stochastic Neighbour Embedding
UC	Ulcerative colitis
UMAP	Uniform manifold approximation and projection
UMI	Unique molecular identifier
UST	Ustekinumab
VDZ	Vedolizumab
WNT	Wingless-related integration site

Chapter 1: Introduction

1.1 Fistulating Crohn's disease

Crohn's disease (CD) presents as a chronic inflammatory condition affecting the gastrointestinal tract, believed to result from an immune system dysfunction to environmental triggers in individuals with genetic predispositions. While its highest prevalence is observed in Europe (505 per 100,000) and North America, there is a noticeable increase in incidence across newly industrialised regions of Africa, Asia and South America (1, 2). Moreover, the incidence of CD in Asia has risen more rapidly than that of ulcerative colitis (UC), another form of inflammatory bowel disease (IBD) (3-5). The relapsing-remitting nature of CD, early age of onset and the absence of a definitive cure contribute significantly to its substantial morbidity and mortality rates. Complications such as intestinal strictures, fistulae, malignancies and extraintestinal manifestations further underscore its clinical significance (6).

Definitions

Fistula formation represents a well-established complication and source of morbidity in individuals with CD. The persistent, transmural inflammation affecting the bowel wall in CD can lead to the formation of sinus tracts. Once these tracts breach the serosal layer, they give rise to fistulae, abnormal communications between two epithelialised surfaces. These fistulae can be clinically

categorised anatomically into internal and perianal types. Internal fistulae, in turn, can be subcategorised based on whether they create connections between intestinal segments (e.g., enteroenteric, enterocolic) or between the intestine and adjacent organs (e.g., enterocutaneous, enterovesical or rectovaginal) (7). Perianal fistulae arise from the anorectal canal and course to a cutaneous surface i.e., perianal skin (8) **(Fig. 1.1)**. Perianal fistulae can be subdivided according to Park's classification (9) into two main categories: low (situated below the dentate line) or high (situated above the dentate line) and further delineated as simple or complex (10). The distinction between these classifications is based on various factors such as the presence of pain, multiple external openings, and evidence of rectovaginal involvement or anorectal stricture, which primarily characterise the latter type.

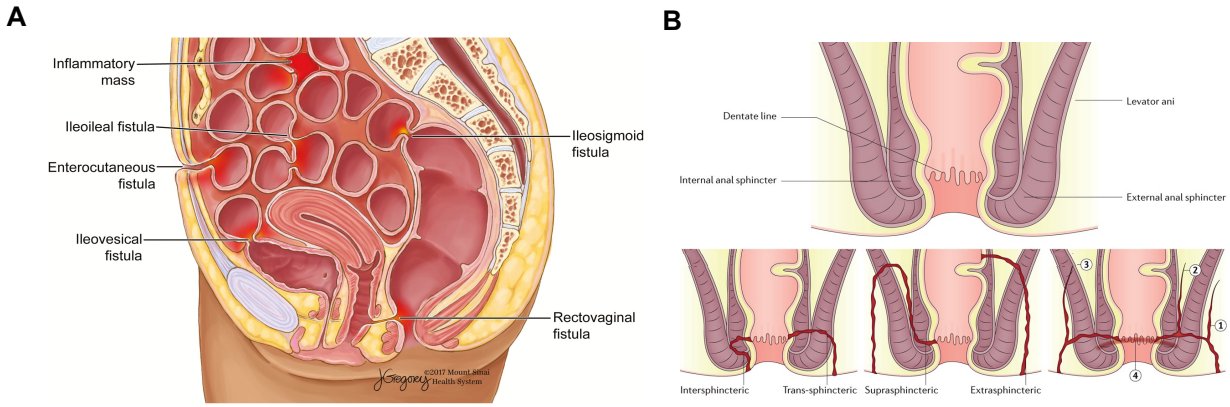


Figure 1.1: Schematic of Different Types of CD-associated Fistulae. Representative schematic of CD-associated fistulae arising internally and terminating externally or internally (A) (11). Schematic of Parks classification of perianal fistula in a coronal plane (B) (12). Illustrations of fistula extensions: extension into roof of ischioanal fossa (infraelevatoric) (1); supraelevatoric extension from apex of trans-sphincteric fistula (2); supraelevatoric extension from trans-sphincteric fistula (3); and horseshoe extension (4). Figures A. and B. from Hirten et al. 2018 and Panés et al. 2017 respectively. Image and adapted caption re-used with permissions.

Epidemiology

Approximately 35-40% of individuals diagnosed with CD will develop at least one fistula throughout their disease course (13-15). Subsequent to this, a population-based study indicated a cumulative incidence of fistulae at 33% within 10 years post-diagnosis and 50% after a span of 20 years (13). While the majority of these fistulae are perianal (54%), internal fistulae, although less common, pose

greater diagnostic and therapeutic challenges (13). The healthcare burden of disease is substantial (16); the majority of fistulating episodes (83%) result in surgery (13). Moreover, the direct economic burden associated with fistulating CD is significant (17,18), with a large epidemiological study demonstrating that fistulating CD incurs the highest annual cost per patient-year compared to any other phenotype ($p < 0.0001$)(19).

Variations in the prevalence of this phenotype are also noted across different ethnic groups (20, 21); South Asian patients with CD exhibit a higher likelihood of developing fistulating disease compared to Caucasians (24.1% *vs.* 8.6%, $p=0.002$) and are more likely to present with perianal disease at the time of diagnosis (22). Additionally, several studies indicate a significantly higher incidence of both perianal and fistulating disease in patients with CD of Afro-Caribbean descent (23-26).

While there seems to be no distinct sex-specific distribution in adult CD overall, gender emerges as a significant factor in perianal fistula incidence. Multiple studies (27, 28) suggest that females are at a lower risk of developing perianal CD fistulae (29). For example, a large IBD cohort study involving 1600 patients with CD reported a relative risk for females of 0.727 ($p = 0.002$) (30).

Natural history

CD is marked by alternating phases of remission and flare-ups of intestinal inflammation, known as 'relapses', which typically occur unpredictably. Approximately 10% of patients achieve sustained and stable remission, as determined through endoscopic evaluation (31). However, within 10 years of CD diagnosis, up to half of patients may require intestinal resection due to complications (32).

While CD can affect any part of the gastrointestinal tract, the terminal ileum and colon are the most commonly affected segments (33). Although the distribution of the disease remains relatively stable (34), its behaviour or phenotype undergoes significant variation over time (14). The Montreal classification broadly categorises CD phenotypes into stricturing, penetrating, or non-stricturing, non-penetrating (**Table 1.1**) (35). The disease phenotype may evolve from non-stricturing, non-

penetrating disease to fibrostenotic stricturing or penetrating disease as chronic inflammation persists, ultimately resulting in intestinal damage (36). For the purpose of this thesis, I will use the terms 'fistulating' and 'inflammatory' to refer to 'penetrating' and 'non-stricturing, non-penetrating' phenotypes respectively.

Table 1.1. Montreal classification for Crohn's disease

Age at diagnosis	A1	<16 y
	A2	17 - 40 y
	A3	>40 y
Location	L1	Ileal
	L2	Colonic
	L3	Ileocolonic
	L4	Isolated upper disease*
Behaviour	B1	Non-stricturing, non-penetrating
	B2	Stricturing
	B3	Penetrating
	p	Perianal disease modifier†

*L4 is a modifier that can be added to L1-L3 when concomitant upper GI disease is present
†p is added to B1-B3 when concomitant perianal disease is present

Phenotypic descriptions are important for prognostication, particularly in population-based studies. While the inflammatory type constitutes the majority of CD cases (81%), nearly a fifth progress to the more complicated fistulating and stricturing phenotypes within 90 days. Additionally, more than half progress within 20 years, especially those presenting with perianal fistulae and ileal involvement at the time of diagnosis (33).

At the time of diagnosis, up to a third of patients exhibit signs of intestinal complications, such as fistulating disease (14). The presence of fistulating complications at diagnosis correlated with worse outcomes, including high rates of surgery and hospitalisation (37). Moreover, perianal fistula formation may precede the formal diagnosis of CD in 30.4% of cases (38). However, as aforementioned, it is recognised that the risk of fistula formation increases with a prolonged disease

duration, which may in part explain the significantly increased risk of fistulating disease associated with earlier age at diagnosis (39).

Colonic involvement in CD, particularly when it extends to the rectum, poses an additional risk factor for perianal fistulating disease, demonstrating a markedly higher incidence compared to patients without colonic involvement (40, 41), (29). Incidentally, a multicentre, retrospective cohort study identified a higher rate of colonic CD among African American patients when compared to Caucasians (OR 1.9, 95% CI: 1.1-3.4, p=0.03) (42).

Although perianal disease activity often correlates with luminal inflammation in CD, approximately 5% of patients present with isolated perianal disease without luminal inflammation (43). In contrast, ileocolonic (L3) or isolated ileal involvement (L1) in CD is more significantly seen co-existing with or associated with the development of internal fistulae ($p<0.03$ and $p=0.009$ respectively) (29, 44).

Beyond the risk for fistulating disease, patients with CD have an increased risk for the development of intestinal and extra-intestinal malignancies (45-48). Specifically, in fistulating disease, the incidence of anal cancer (either adenocarcinoma or squamous cell carcinoma) is estimated at 0.2 per 1000 patient-years, with chronic (>10 years) perianal fistulating disease as an established risk factor (48) (47).

Clinical presentation & management

The symptoms of fistulating disease vary depending on the origin and termination sites of the fistula. Enteroenteric fistulae may present with diarrhoea, while enterovesical or enterourethral fistulae may lead to urinary tract infections and pneumaturia. Enterocutaneous fistulae can cause skin irritation due to the draining of intestinal contents, while rectovaginal fistulae may result in the passage of stool through the vagina or dyspareunia. Perianal fistulae often manifest with perianal pain, purulent discharge and faecal incontinence. Clearly, these symptoms can be debilitating and significantly impact the individual's quality of life (49-51).

Management of established fistulae necessitates meticulous assessment of location, extent and likelihood of other complications (e.g., abscess formation) (13). Utilising cross-sectional imaging techniques such as CT, MR enterography and bowel ultrasonography is essential for thoroughly assessing the extent of fistulating disease. These modalities offer high sensitivity, providing comparable accuracy rates in diagnosing and evaluating fistulating disease (52). If complicated perianal fistulae are suspected, MRI pelvis is the imaging modality of choice to delineate the fistulae and evaluate for complications to guide treatment (13). Uncomplicated perianal fistulae can be managed with seton placement, complicated perianal fistula often require surgical intervention in combination with medical therapy.

With the emergence of biologics, the therapeutic landscape for fistulating CD has evolved, along with treatment targets. It is well established that persistent subclinical intestinal inflammation correlates with adverse outcomes, including fistula formation (36). Therefore, the treatment objective is to achieve deep remission, defined as clinical remission, biomarker remission and mucosal healing (restoration of the intestinal lining, regression or disappearance of endoscopic lesions) (53). Research indicates that mucosal healing not only reduces the risk of fistula-related complications but also decreases the likelihood of surgery compared to patients with severe ulceration (36). Additionally, commencing highly effective biological therapies soon after diagnosis can result in higher rates of achieving clinical remission and mucosal healing (54).

Anti-tumour necrosis factor (anti-TNF) therapies remain the first-line biological treatment for fistulating CD (55-59), with anti-integrin agents like Vedolizumab (VDZ) and anti-IL 12/23 agents like Ustekinumab (UST) also demonstrating efficacy (60-62). McCurdy et al. underscored the effectiveness of anti-TNFs in preventing fistulating CD, highlighting their protective effect against both internal (HR 0.66, 95% CI: 0.55-0.78, $p<0.0001$) and perianal (HR 0.88, 95% CI: 0;80-0.96, $p=0.0045$) fistulae compared to VDZ in a large cohort study (63). Additionally, findings from a population-based cohort study revealed a notable reduction in the cumulative incidence of perianal and rectovaginal fistulae, along with decreased rates of proctectomy, during the biologics era

compared to before the introduction of biologic therapies (32). However, primary non-response to anti-TNF occurs in about one-third of patients, with half experiencing loss of response within the first year (64). Moreover, only 25% of patients with perianal fistulating CD achieve sustained fistula closure with biological therapy (65). Furthermore, although 44% of patients with internal fistulae achieved radiologically confirmed healing at 5-years post anti-TNF therapy, 43% required major surgery during the follow up period and 20% developed a complicating intestinal abscess (59). These findings emphasise the necessity for exploring additional therapeutic options.

Promising data suggests the emergence of small molecule therapy, specifically the JAK-1 inhibitor Filgotinib, as effective in a randomised controlled trial aimed at treating fistulating perianal disease. However, further studies are required to validate its efficacy (66).

Fistulating disease often necessitates surgical intervention, with fistulotomy and seton placement being common approaches for perianal CD fistulae with the objective to achieve fistula resolution while maintaining anal sphincter function. Approximately one-third of patients with treatment-refractory perianal fistulae require proctectomy (67). In the case of internal fistulae, surgery is indicated for complicated disease, unresponsive symptoms, or sepsis. However, surgery can be extensive and associated with considerable morbidity and mortality rates (11) (44), highlighting the importance of collaborative, multidisciplinary approaches to fistulating CD management in specialised centres.

In recent years, stem cell therapy, notably adipose-derived mesenchymal stem cells (MSCs), has emerged as a promising avenue for treating fistulating CD. Early trials demonstrated MSCs potential to heal fistulae, albeit with some instances of recurrence (17.6%) (68, 69). Notably, recent Phase 3 trials (70) have shown significant efficacy of darvadstrocel (MSCs), leading to remission in patients who were unresponsive to conventional treatments. Real-world data have further validated these findings, underlying the therapy's safety profile (71). Additionally, autologous adipose tissue-derived stem cell transplantation has demonstrated superior closure rates when compared to anti-TNF

treatment in cases of fistulating perianal CD, albeit a small cohort (72). With recent approvals, MSC therapy holds promise for patients with treatment refractory fistulating CD.

However, despite conventional medical and/or surgical interventions for fistula healing, one-third of patients will experience a recurrence of fistulae (13), while complications arising from these interventions also pose challenges (11). This underscores the pressing need for innovative therapeutic strategies. Achieving this goal necessitates a comprehensive understanding of the pathogenesis underlying fistulating CD.

1.2 Histopathological characterisation of CD-associated fistula

Histomorphological features of normal intestine

It is crucial to histopathologically define CD-associated fistulae to unravel their pathogenesis. However, before dissecting these aberrations, it is essential to establish an understanding of the normal histomorphological features of the intestine.

The mucosal histomorphology of the small and large intestine represents a dynamic interplay of specialised structures optimised for their respective roles. In both small and large intestine, the mucosa is composed of the epithelium, lamina propria and muscularis mucosa.

In the small intestine, the mucosa is characterised by finger-like projections known as villi, which protrude into the intestinal lumen. These villi are lined with a single layer of columnar epithelial cells called enterocytes, primarily responsible for absorption. Goblet cells, interspersed between enterocytes, secrete mucus. Crypts, extending downward from the mucosal layer, are comprised of straight tubular glands. Intestinal stem cells (ISCs), crucial for epithelial cell differentiation, line the proximal part of the crypts. Paneth cells, located adjacent to intestinal stem cells in crypt bases, secrete antimicrobial peptides.

Conversely, the large intestine mucosa is lined by colonocytes, the predominant epithelial cell. The colonic mucosa lacks villous structures however contain numerous colonic crypts that extend deeply into the mucosa, maximising epithelial surface area. These crypts are arranged in a characteristic parallel pattern, resembling test tubes. The crypt is rich in goblet cells, contributing to mucus secretion, and also house enteroendocrine cells (EECs) in addition to, Paneth cells and stem cells.

The lamina propria surrounds the intestinal crypts and forms the core of the intestinal villi (in the small intestine only). It consists of connective tissue, vasculature, nerves, stromal cells and immune cells, primarily lymphocytes. The muscularis mucosae, consisting of thin strands of smooth muscle fibres, delineates the mucosa from the submucosa. Underlying the mucosa is the submucosa, a thick layer of loose connective tissue containing larger vessels, nerves (Meissner plexus) and thin smooth muscle bundles, along with stromal cells and adipose tissue. In the ileum, the submucosa contains aggregations of lymphoid follicles known as Peyer's patches.

Beneath the submucosa lies the muscularis propria, composed of an inner circular layer and outer longitudinal smooth muscle layer. The myenteric plexus of Auerbach lies between the muscle fibres. In the colon, the outer layer of longitudinal muscle fibres is present only in the teniae coli. The subserosa, a thin layer of fibroadipose tissue, separates the muscle layer from the outermost layer of the intestinal wall, the serosa. The serosa comprises loose connective tissue with elastic and collagen fibres, nerves and vessels, covered by a single layer of flat mesothelial cells. In areas lacking the mesothelial layer, the outermost layer is called adventitia (**Fig. 1.2**). In the rectum, features transition from the sigmoid colon, with thicker mucosa and longer crypts primarily lined by goblet cells, gradually transitioning from columnar to squamous epithelium as it extends into the anal canal.

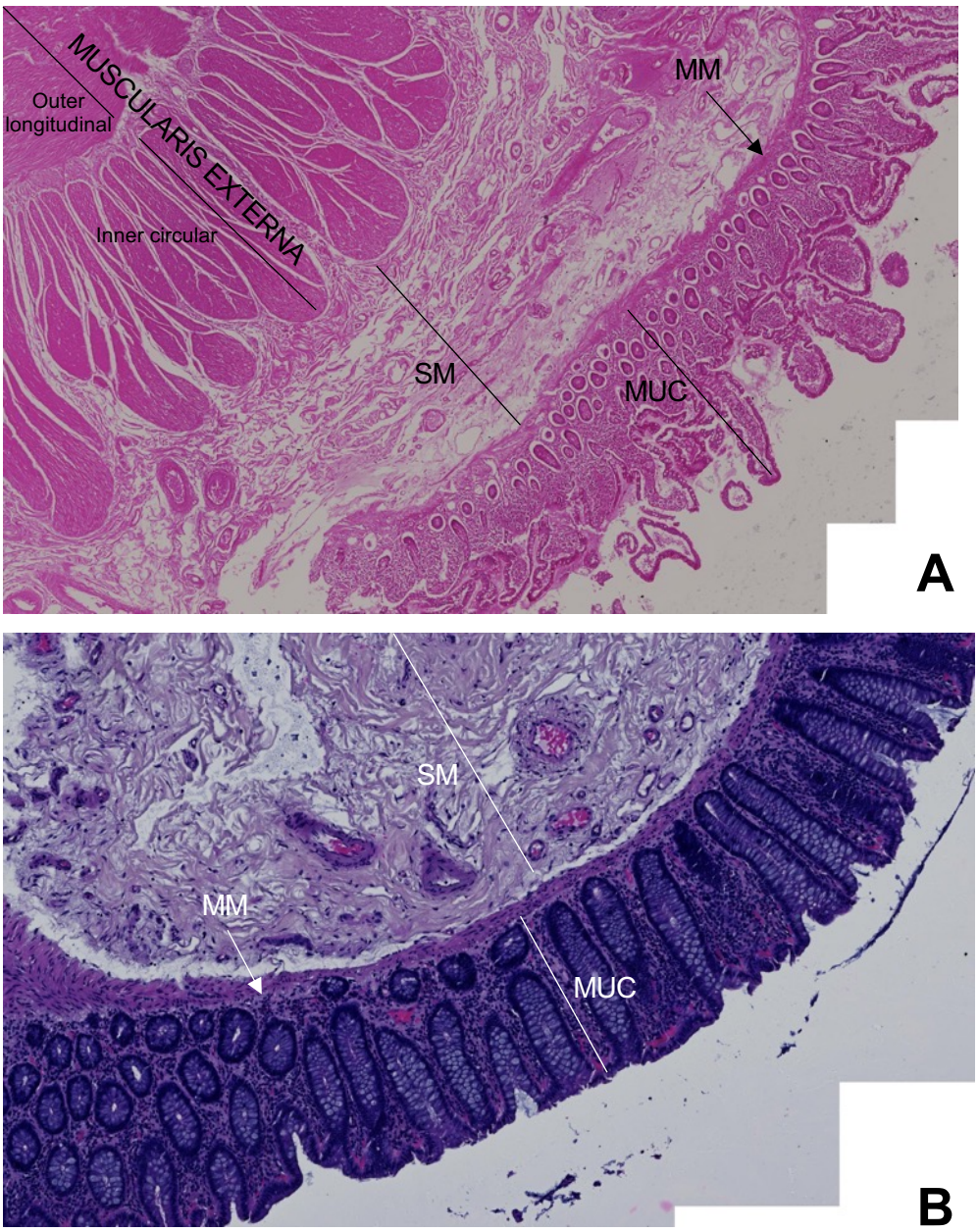


Figure 1.2: Histomorphological Features of Normal Intestine. Human histological haematoxylin & eosin (H&E) staining of human, adult small intestine (A) and large intestine (B) with villi (A) and colonic crypts (B). MUC, mucosa; MM, muscuclaris mucosae; SM, submucosa. [A: Patient GI 9676, B: Patient TIP 571]

Histopathological characterisation of CD-associated fistula

A comprehensive study examined multiple fistula specimens, encompassing both CD-associated and control cases, providing the most thorough histopathological description of fistulae (73). Irrespective of the underlying cause, all fistulae exhibited a central fissure penetrating through the lamina propria and muscularis mucosae, extending into deeper tissue layers. Granulation tissue, characterised by histiocytes and a dense network of capillaries, surrounded the fistulae. Luminal contents varied,

including debris, erythrocytes, or non-specific acute or chronic inflammatory cells (73). While granulomas were present in CD-associated fistulae, they are not pathognomonic; notably, they were often absent in perianal CD-associated fistulae. A multinucleate foreign body-giant cell type reaction can occur in any type of fistula; however, it is considered non-specific (74).

The epithelium lining fistulae typically appeared flattened in the small or large intestine, without goblet cells or composed of narrow squamous epithelial cells in perianal or cutaneous fistulae. However, approximately two-thirds of fistulae lacked an epithelial lining and were instead lined by mesenchymal-like cells known as "transitional cells" (TCs), which retained gap junctions with each other. The area where mucosa changed to TCs is referred to as the transition zone. This unique epithelial transformation to mesenchymal-like cells is characteristic of CD-associated fistulae, distinct from controls. In some instances, a new basement membrane formed with TCs connected by fibronexus, while in others, TCs appeared disordered with fragmented basement membranes. There was no apparent association between epithelialisation of fistulae and the extent of inflammation or depth of penetration (73). Furthermore, in some instances, multiple fistulae were present in the same patient synchronously or metachronously, the epithelialisation status of fistulae were not consistent (73).

In terms of inflammation, the majority of fistula specimens exhibited inflammatory changes, with severe acute inflammation evident in at least 56% of cases (73). Chronic fibrosis often manifested as a feature of CD-associated fistulae, characterised by thickening of the muscularis mucosae, hypertrophy and hyperplasia of the submucosa and muscularis propria. Features of chronic inflammation, with lymphoid aggregates and lymphoid follicles, may co-exist alongside evidence of acute inflammation i.e., extensive neutrophilic infiltrates. Neural hypertrophy could be observed as a secondary reactive change or as a mechanism for intestinal stiffening (75, 76).

Additionally, a prospective study examined forty-two surgical resection specimens from patients with complicated CD (76). In the cases with internal fistulae, the majority (96.3%) had a co-existing

stricture macroscopically; fistulae were found either within the stricture or at its proximal end. The association of fistulae and presence of stricture was significant ($p < 0.02$) and has been found by others (44, 77). However, distinct histopathologic characteristics of stricturing and fistulating CD have been described elsewhere (44); whilst submucosal fibrosis, ulcer and transmural ulceration are seen in both, the presence of pseudopolyps, microabscesses, granuloma, mononuclear cell inflammation and deep fissures were significantly more common in non-perianal fistula compared to stricturing ($p < 0.05$) (44).

With respect to changes to vasculature, in just over a quarter of cases of internal fistulae, a vessel was visualised alongside the fistula traversing the muscularis propria. Vessels demonstrated signs of high-grade reactive vasculitis and were partly destroyed (76). The depth of penetration of the central fissure can give rise to classifying the stage of fistulae histopathologically. Historically, the term "initial" fistulae or "sinus tracts" denotes penetration of granulation tissue, by less than 2-mm depth, into the mesentery or mesocolon (76).

Immunohistochemical studies revealed distinct immune cell populations within CD-associated fistulae, with the inner fistula wall densely infiltrated by $CD45RO^+$ T-cells and a small band of $CD68^+$ macrophages. In contrast, the outer fistula wall is marked by $CD20^+$ B-cells. This pattern differs from non-CD-associated fistulae, where $CD68^+$ macrophages are distributed throughout the entire fistula wall and $CD45RO^+$ T-cells predominantly occupy the outer two thirds, with fewer $CD20^+$ B-cells. Immune cell distribution appears to be independent of the fistula location or depth of penetration. Perianal fistulae in the control group often displayed an isolated infiltration of monocytes, surrounded by fibrosis, resembling advanced chronic fistulae (73). The study concluded that the histopathologic features were fairly similar between CD-associated and non-CD associated fistulae however they were markedly different immunohistochemically.

Maggi and colleagues conducted phenotypic and functional analysis of T-cells retrieved from CD perianal fistulae tissue and circulating T-cells from peripheral blood (78). They found that

$CD4^+CD161^+$ T-cells with Th17, Th1, and Th17/1 phenotype significantly infiltrate fistula tissue compared to peripheral blood. Local administration of anti-TNF (Adalimumab) resulted in clinical resolution of fistulae and a significant reduction in the frequency of $CD161^+$ T-cells within fistula tissue.

Aberrations in immune cell populations have been described in CD fistulae. More specifically, alterations in T-cell distributions; $CD4^+$ and $CD8^+$ cells were shown to populate the CD fistula tracts compared with the surrounding tissue whereas more circulating $CD3^+CD8^-$ T-cells were present in the peripheral blood in patients with perianal CD fistulae compared to healthy controls. Both T-cell subsets secreted high amounts of TNF-α and interleukin-13 (IL-13) in co-culture experiments, highlighting their potential role in the pathogenesis of fistulating CD (79).

1.3 Pathogenesis of CD-associated fistula

1.3.1 Immune and cellular biology

1.3.1.1 Epithelial-to-mesenchymal transition

The initial step in fistula formation is thought to be due to a compromised gut barrier allowing pathological molecules to enter in, triggering two responses in the intestinal mucosa – one for repair and another causing intestinal inflammation. Both responses are thought to trigger the process of epithelial-to-mesenchymal transition (EMT) in intestinal epithelial cells (IEC). EMT stands as a crucial process wherein epithelial cells undergo differentiation into mesenchymal-like cells, assuming their characteristics and properties. This transformation involves a loss of epithelial traits such as polarity, adhesion, alongside the acquisition of a mesenchymal phenotype marked by reduced cell-to-cell adhesion and enhanced migratory potential (80, 81).

In physiological contexts, EMT plays a pivotal role in embryogenesis and organ development (82) (83), facilitating the migration and infiltration of cells into adjacent tissue layers. However, in pathological states, EMT's involvement is noted in conditions like cancer and fibrosis (84-86). Within

tissue damage scenarios, EMT serves as a mechanism for tissue remodelling and repair, allowing the recruitment of mesenchymal cells from epithelial cells. While epithelial cells gain the ability to migrate to injury sites through EMT, excessive extracellular matrix (ECM) deposition by mesenchymal-like cells can lead to tissue fibrosis (87). The transition from epithelial to mesenchymal cells is orchestrated by a spectrum of cellular changes, including the expression of transcription factors, cytokines and regulatory proteins (88).

A body of evidence supports the proposition that EMT drives the formation of CD-associated fistulae (73, 89, 90). Chronic inflammation results in diminished epithelial repair and impaired migratory capabilities of colonic lamina propria fibroblasts (CLPF), contributing to suboptimal wound healing in fistulating CD (91, 92). Pro-inflammatory cytokines, notably TNF-α and IFN, are implicated in inducing this impaired migratory potential of CLPF. Remarkably, this alteration appears to a be an irreversible process, as the functional status of CLPF does not revert even after withdrawal of cytokines (91). Furthermore, fibronectin, a crucial ECM protein, is believed to stimulate the migration of CLPF to epithelial defects during wound healing. Notably, analysis of CD fistula tissue revealed a striking absence of the ED-A and ED-B isoforms of fibronectin (93), underscoring the critical role of fibronectin in governing the migratory potential of CLPFs in CD fistulae.

To compensate for defective CLPF, IEC within fistulae transform into TCs, forming a thin monolayer lining the fistula (73, 94) enabling defects in the intestinal barrier to be repaired. TC, uniquely found in CD-associated fistulae, express epithelial markers (cytokeratin 8 and 20, reflecting their epithelial origins) and mesenchymal markers (vimentin and smooth muscle actin) while downregulating epithelial adhesion markers (E-cadherin and β-catenin), facilitating their migration towards defects in the intestinal barrier – a hallmark of EMT (89).

Immunohistochemical studies of CD-associated fistulae have revealed the presence of potent inducers and markers of EMT (95) (89). β6-integrin, a pivotal protein predominantly confined to epithelial cells during development and organogenesis, exhibits heightened expression within the

transitional zone of TC in CD-associated fistulae (89). This aberrant overexpression of $\beta 6$-integrin has been directly correlated with increased cell invasiveness, implicating its role in the pathogenesis of fistula formation in CD.

Severe intestinal inflammation leads to the secretion of cytokines such as TNF-α, IL-13 and TGF-β, with TGF-β being the most potent inducer of EMT. TGF-β induces the expression of transcription factors associated with EMT, including *SNAIL1*, *SNAIL2 (SLUG)* and Ets-1, with the former two strongly repressing E-cadherin and enhancing cell invasion. *SNAIL1* exhibits significant expression within the nuclei of TC lining CD-associated fistulae, whereas *SNAIL2* predominantly appears in the cells surrounding the fistula tract (96). Upregulation of TGFβ-1 and TGFβ-2 expression, along with increased expression of $\beta 6$-integrin, is observed in TC lining CD-associated fistula tracts when compared to normal IECs (89, 97). In addition, there is notable upregulation of other molecules in CD-associated fistulae including Ets-1(90, 98), IL-13 and the IL-13 receptor.

In summary, CD-associated fistulae exhibit prominent features indicative of EMT, including reduced E-cadherin and β-catenin expression, upregulation of TGF-β, induction of EMT transcription factors, and overexpression of $\beta 6$-integrin in TC, culminating in enhanced migratory potential and increased cell invasiveness, irrespective of fistula location (81).

Additionally, Dickkopf-homolog 1 (*DKK-1*), strongly upregulated in TC lining CD-associated fistula tracts, is proposed to regulate EMT in IEC by inhibiting WNT signalling and influencing TGF-β-stimulated IL-13 secretion (99, 100). Conversely, Ortiz-Masiá and colleagues implicate succinate and its receptor SUCNR1 in inducing EMT in IEC through the activation of WNT signalling pathways. Previous work conducted by the same authors revealed that the heightened induction of EMT in CD-associated fistula tissue stems from an increased interaction between the WNT ligand, WNT2b, and its receptor FDZ4, as compared to control samples (99). The maintenance of the intestinal stem cell niche relies on the precise regulating of various signalling pathways, among which

the WNT signalling pathway holds prominence. This pathway governs epithelial stem cell self-renewal, proliferation and differentiation. Therefore, it is reasonable to infer that aberrations in the WNT pathway and its downstream targets might contribute to the initiation of EMT.

1.3.1.2 Cytokine profile

Various published studies have aimed to delineate the cytokine profile of CD-associated fistula tracts, shedding light on its immunopathogenesis. TNF emerges as a key player in inducing EMT in IEC and promoting the expression of TGF-β (98, 101, 102). Notably, TNF and its receptor, TNF Receptor 1 (TNF-R1), exhibit strong expression in TCs lining fistula tracts, as well as in IEC of adjacent crypts in patients with CD (96). Moreover, TNF induces the expression of β6-integrin and the Ets-1 transcription factor – both pivotal mediators of EMT – in both IEC and CLPF (98).

Serum levels of TNF-α and IL-6 significantly correlate with the presence of active CD perianal fistulae. TNF-α induces the production of other inflammatory cytokines, including IL-1β and IL-6. The expression of IL-1β and IL-6 was significantly higher in the rectal mucosa of perianal CD fistulae compared to in those with luminal CD or healthy controls, implicating TNF-α in the pathogenesis of CD fistulae (103).

Similar to TNF, IL-13 and its receptor, IL-13R1, are prominently expressed in TCs lining fistula tracts and adjacent crypts, a phenomenon unique to CD-associated fistulae. IL-13 upregulates pro-inflammatory cytokines such as TNF-α, IL-12 and IL-6 in fistulating CD tissue (90). Moreover, IL-13 functionally promotes the expression of genes associated with cell invasion (β6-integrin) and EMT (*SNAIL2*) in IEC and *in vitro* models (90).

In addition to TCs, sources of IL-13 include Th2 cells, ILC2s, natural killer cells, innate immune cells and non-immune cells such as IECs and intestinal fibroblasts. The function of IL-13 extends to

enhancing epithelial permeability, with its production potentiated by the presence of IL-33 and IL-25, which are derived from IECs (104). In healthy intestine, as well as in cases of UC and non-fistulating CD, IL-13 expression is notably absent, regardless of the presence of inflammation. However, the primary source of IL-13 in CD-associated fistulae remains to be elucidated (90).

A comparative study examining cytokine concentrations in fistula tracts of idiopathic versus perianal CD fistulae revealed significantly higher IL-12 concentrations and a lower IL-1RA/IL-1β ratio in the CD group (105).

TGF-β, a central mediator of EMT, localises to the TC lining fistula tracts and induces the expression of *SNAIL1* and IL-13 in *in vitro* models. Primary human CLPF derived from patients with fistulating CD exhibit altered function when treated with TGF-β. The synergistic effect of TGF-β and IL-13 in the pathogenesis of fistulae is also proposed (90).

The abundance of these cytokines in the lining of fistula tracts, adjacent tissue, and peripheral blood underscores their involvement in fistulating CD. Moreover, the clinical efficacy of anti-cytokine biological agents, such as infliximab (anti-TNF), in treating fistulating CD further substantiates the significance of cytokines in this context (106).

1.3.1.3 MMPs & TIMPs

The ECM constitutes a network of structural proteins essential for maintaining the integrity and architecture of the intestine. This dynamic matrix undergoes continuous remodelling, primarily orchestrated by various enzymes responsible for its degradation. Among these enzymes, matrix metalloproteinases (MMPs) hold a pivotal role in ECM turnover and remodelling processes. Dysregulated activity of MMPs can lead to aberrant breakdown of ECM components, which has been implicated in various pathological conditions, including cancer and IBD (107).

Increased MMP activity is associated with immune mediated tissue injury and is notably observed in CD (108). Studies using murine models of DSS-colitis underscore the importance of MMPs, as selective deletion of MMP-9 has been shown to confer a protective effect (109-112). Tissue inhibitors of MMP (TIMPs) act as natural inhibitors of MMPs and are secreted by MMP producing cells (113).

Kirkegaard and colleagues identified robust expression of MMP-3 in CD-associated fistula tissue compared to controls, regardless of the inflammatory state. MMP-3 were predominantly localised to mononuclear cells and fibroblasts, while MMP-9 was primarily found in granulocytes and only in fistulae exhibiting active inflammation (114). Additionally, MMP-3 and MMP-9 have been identified in idiopathic fistulae. MMP-13 protein expression is detectable in CD-associated fistulae but nearly absent in non-fistula CD tissue. Conversely, protein levels of inhibitory molecules such as TIMP-1, TIMP-2 and TIMP-3 are notably low in CD-associated fistulae tissue (114).

Further, previous studies have shown induction of MMPs by IL-1β, IL-12, IL-13, IL-22, TNF-α and TGF-β, all of which are cytokines proven to be associated with CD-associated fistulae (115). These findings provide compelling evidence for the involvement of MMPs in the pathogenesis of CD-associated fistulae through aberrant ECM degradation (**Fig. 1.3**).

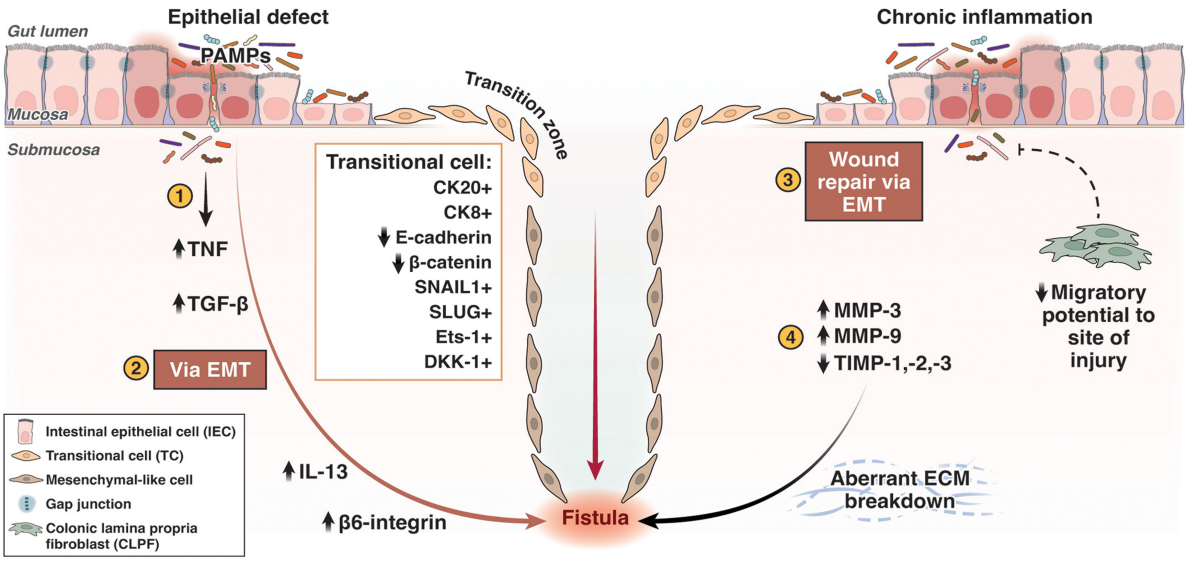


Figure 1.3: Schematic of the Pathogenesis of CD-associated Fistula Formation. (1) Intestinal epithelial cells (IECs) undergo EMT and transform into transitional cells (TCs) in response to epithelial barrier defects. This allows pathogen-associated molecular patterns to infiltrate the gut mucosa, triggering an immune response. (2) TNF is upregulated, which acts as a strong inducer of TGF-β, driving a cascade of pro-inflammatory cytokines (such as IL-13) and invasive molecules (like β6-integrin). This process causes TCs to exhibit traits of invasive, mesenchymal-like cells. Although TCs retain their epithelial origin markers (CK20/8^{+}), they lose epithelial adhesion molecules (E-cadherin, β-catenin) and upregulate EMT-inducing transcription factors (*SNAIL1*, *SLUG*, Ets-1) as well as migratory molecules like DKK-1. Additionally, TCs may exhibit more disordered structures, with loss of gap junctions and fragmentation of basement membranes. (3) Chronic inflammation impairs epithelial repair and decreases the migratory capacity of colonic lamina propria fibroblasts (CLPFs), leading to poor wound healing in fistulating CD. In response, IECs undergo EMT to compensate by attempting to restore the barrier. (4) Matrix metalloproteinases (MMPs), such as MMP-3 and MMP-9, are overexpressed in fistulating CD, while their inhibitors (TIMPs) are reduced. This imbalance leads to excessive degradation of the extracellular matrix and tissue remodeling, contributing to fistula development. [CK, Cytokeratin; ECM, extracellular matrix, MMP, matrix metalloproteinases; PAMP, pathogen-associated molecular pattern]. Figure re-used with permissions (116).

1.3.2 Genetic contribution

The interplay between genetics and the environment is well-established in CD. Extensive genome-wide association studies (GWAS) have identified approximately 320 genes linked to disease pathogenesis or risk (117-119) (120). Genetic contributions to specific phenotypes, including fistulating disease (29, 121), have also been elucidated.

Nucleotide oligomerization domain 2 (*NOD2*) stands out as the strongest genetic predictor of CD susceptibility and phenotype, including fistulating disease (122-124). Various studies have demonstrated association between specific genetic variants and different fistula phenotypes. For non-

perianal CD fistulating phenotype, the presence of certain alleles at rs12704036 (OR 1.74, 95% CI: 1.20-2.54, p=0.004), any *NOD2* variant (OR 1.47, 95% CI: 1.01-2.15, p=0.04), and *IRGM* rs4958847 G-allele (OR 9.22, 95% CI: 1.17–72.94, p=0.04) have been independently associated (29). Conversely, perianal fistulating phenotypes have been linked to the presence of a C-allele at *CDKALI* rs6908425, homozygosity for the A-allele at *HERC2* rs916977 and the absence of any *NOD2* variant (p=0.008, p=0.05 and p=0.002 respectively) (29). Moreover, *NOD2* mutation rs72796353 has been strongly associated with perianal fistula development (OR 5.27, 95% CI: 2.75-10.12, $p=2.78 \times 10^{-7}$) when compared to *NOD2* wild-type carriers (125). The function of *NOD2* rs72796353 remains to be characterised.

Beyond *NOD2* polymorphisms, the likelihood of developing internal fistulating CD is notably linked with carrying a variant allele of *PRDM1* (PR domain containing 1, with ZNF domain) rs7746082, LOC441108, and IL23R. Specifically, the presence of *ATG16L1* (autophagy-related 16-like 1) and *PRDM1* is independently correlated with an earlier onset of internal fistulating disease, whereas IL23R rs11465804 is associated with a delayed onset (122). Whereas polymorphisms in the gene *PUS10* provides a protective effect against perianal fistulating disease (122).

OCTN variants have also shown associations with penetrating disease behaviour in both Korean (OR 4.239, p=0.042) (126) and Belgian (OR 1.47, 95% CI: 1.028–2.114, p=0.035) (127) populations. Additionally, a study of Chinese individuals with CD identified the following, *IRGM* rs72553867, *NKX2-3* rs4409764, *AOX1* rs3731772 as well as SNPs located in *TNFSF15*, as susceptibility factors for perianal fistulating CD (128).

Recent epigenetic analysis (129) has revealed distinct DNA methylation signatures within the intestinal mucosal lesions of fistulating CD, characterised by upregulation of apoptotic processes and positive regulation of IL-8 production compared to normal intestinal tissue. IL-8, a potent chemokine, is recognised for its role in recruiting and activating neutrophils in inflammatory regions (130); its secretion in CD is significantly increased (131) and therefore used as a biomarker in some

studies (132). Additionally, pathway analysis of differential DNA methylation signatures between fistulating and non-fistulating tissues, revealed an enrichment of signalling pathways linked to ECM-receptor interactions, further supporting the role of ECM remodelling in fistula formation.

It is important to note that genetic contributions to fistulating CD phenotypes differ between perianal and internal fistulae. However universally, risk alleles associated with fistulating CD encode proteins involved in regulating the ileal microbiota, adaptive immunity, autophagy or maintaining intestinal epithelial barrier integrity.

1.3.3 Microbial contribution

The interaction between the host and microbes in CD has been thoroughly researched (119, 133). It is well established that intestinal inflammation in CD results from the dysregulated immune responses to microbiota within the intestinal mucosa (134), however the role of the intestinal microbiome in the development and persistence of CD-associated fistulae is not well understood. Antibiotics and faecal diversion have been effective in managing perianal CD fistulae, suggesting a potential pathogenic bacterial contribution in their development.

Some studies have looked into the microbial composition of CD-associated fistula tracts to identify potential causative organisms. For instance, West et al., found that gram positive organisms are predominant in perianal fistulae (135), while another study reported a higher abundance of gram-positive bacterium *Corynebacterium* and gram-negative *Achromobacter* in perianal fistula tracts relative to paired stool samples via 16s rRNA gene profiling (136). Conversely, a study examining idiopathic and perianal CD fistula tracts found no mucosa-associated bacteria despite mucosal inflammation. While sampling error could be a factor, the absence of bacterial colonisation in established fistula tracts suggests bacteria may not be crucial for fistula persistence (137).

Bacterial cell wall muramyl dipeptide (MDP), a ligand for *NOD2*, induces the expression of molecules associated with EMT (TNF-*a*, TGF-*β*, SNAIL1, IL-13 and Ets-1) within IEC and CLPF from fistulae (98). Additionally, certain enteric pathogens like *Citrobacter rodentium* and *E. coli* have been shown to induce EMT through signalling pathways, suggesting luminal bacteria might contribute to fistula development via EMT pathways (138, 139).

Recent studies have begun to uncover the role of the mycobiome in CD. Jain et al. found that the fungus Debaryomyces hansenii is enriched in inflamed intestinal CD tissue and can disrupt mucosal healing (140), potentially contributing to CD-associated fistula pathogenesis. Disruption of the natural competition between fungi and bacteria within the gut due to antibiotic use might lead to impaired mucosal healing and enrichment of fungi, potentially explaining the persistence of some perianal fistulae despite antibiotic treatment.

Advancements in technology are needed to further refine and expand our understanding of the microbiome, including the mycobiome, in CD-associated fistulae.

1.3.4 Mesenteric contribution

Recent insights highlighting the role of the mesentery in CD pathogenesis are of interest (141). Macroscopic alterations in the mesentery adjacent to affected intestinal segments serves as a hallmark of CD. Notably, the migration of mesenteric adipose tissue (MAT) toward inflammatory CD lesions in the intestine, termed 'creeping fat', manifests as hyperplastic MAT encircling sites of intestinal inflammation, predominantly within the small bowel (142). While the precise mechanisms for this process remain unclear, Ha et al. provided evidence implicating the gut microbiome in triggering creeping fat formation through the translocation of gut bacteria to MAT (142).

In addition to adipose tissue, the mesentery, encompasses lymphatics and nerves, possessing metabolic and immune properties that may influence fistulating CD (142). MAT plays an important

role in intestinal inflammation by producing and secreting adipokines and cytokines. Among these, adiponectin, an anti-inflammatory adipokine synthesised by MAT, has gained attention. Studies have revealed significantly lower concentrations of adiponectin in patients with internal CD fistulae compared to those without. Moreover, diminished adiponectin levels may exacerbate intestinal inflammation in CD and heighten the risk of internal fistula development. Further exploration of the mesentery's role in CD fistula pathogenesis is warranted to deepen our understanding of these mechanisms.

1.4 Existing models of CD-associated fistula

Current models of CD fistula have historically focused more on perianal manifestations due to easier accessibility to tissue. However, future research directions should prioritise obtaining lesional tissue from CD-associated internal fistulae. Using novel approaches, researchers must first characterise human lesional tissue and define the microenvironment of CD-associated fistulae, which will guide the development and exploration of *in vivo* models. Additionally, leveraging existing knowledge of the phenotypic and molecular characteristics of idiopathic fistulae (114, 143, 144) will aid in understanding the distinct and shared pathways of CD-associated fistula pathogenesis.

In vitro models

In vitro models have primarily relied on intestinal epithelial cell lines and descriptive histopathological and immunohistochemical data to study fistulating CD pathology (79, 89, 99). Notably, Meier and colleagues developed a 3D matrix model to study the behaviour of CD-CLPFs isolated from CD-associated fistulae and strictures. Fistula-derived CLPFs were found to have significantly reduced migratory potential (94). Advances in intestinal epithelial organoids (IEOs) now offer a more accurate representation of intestinal epithelial structure and function (145, 146). Organoid based disease models have provided further evidence for EMT in CD (147) (148-151) . Pivotal work from Hahn et al. demonstrated the capability of TNF-α and TGF-β to induce EMT in IEOs, with discernible

mesenchymal phenotypic changes evident in TGF-*β1* stimulated IEOs. This study proposed a novel method for the examination of EMT in the context of intestinal fibrosis (147).

Future work should focus on developing reproducible intestinal-derived organoid models to explore the role of EMT in CD-associated fistula formation. Moreover, advances in sequencing technology now allows for the identification of somatic mutational burden within epithelia of patient-derived organoid cultures (152).

In vivo models

The lack of a suitability sufficient *in vivo* model to study fistulating CD remains a major limitation of the field. Whilst reliable and reproducible murine models for IBD do exist, neither of these models develop internal fistulae. The sub strain of SAMP1/YitFc mice, developed by Cominelli's group, shows promise as an animal model for fistulating CD with its spontaneous perianal fistula formation and co-existing terminal ileitis (153). The mice exhibit mucosal ulceration of the anal canal characterised by fissure formation and dense inflammatory infiltrate, leading to perirectal abscess formation and anocutaneous fistulae. However, the relatively low incidence (4.8%) of perianal fistula, limits its widespread use as a model. Murine models of non-intestinal fistulae such as trachea-oesophageal (154), and vascular fistulae (155), may also help inform development of future CD models.

Another novel model involves human foetal gut xenografts in SCID mice, which develop spontaneous enterocutaneous fistulae (156). Histopathological features showed remarkable overlap with CD-associated fistulae including evidence for EMT immunohistochemically (157). Although promising, this model requires further validation. Another surgically created model of fistulating disease, in rats, was developed by Flacs and colleagues shows promise (158). Proctitis was initially induced in rats by administering a Trinitrobenzene sulfonic acid (TNBS) enema, followed by the creation of a surgically constructed transsphincteric fistula tract using a surgical thread one week later. This was followed by repeated TNBS enemas until thread removal on day 28. Radiological

examination confirmed the presence of fistula tracts, while histopathological analysis of the fistula tracts and adjacent mucosa revealed features of inflammation (acute and chronic), granulation tissue, epithelialisation and fibrosis; similar to the characteristics of perianal fistulae in CD.

While TNBS-induced colitis shares some similarities with CD, such as transmural inflammation. and a Th1/Th17-type immune response involving pro-inflammatory cytokines, TNF-α, IL-12 and IL-23, it does not fully replicate the complex interplay of immunological mechanisms involved in CD.

Much of the evidence surrounding the pathogenesis of fistulating CD revolves around the dynamic interplay of cellular changes within the intestinal epithelium and stroma, along with their reciprocal interactions. Further, the remodelling of ECM components and characterisation of the changes in the immune milieu require more granular exploration. Advancing technologies, such as single-cell RNA-sequencing (scRNA-seq), now allow for further interrogation of complex diseases, including CD, and may pave the way for the development of functional studies using suitable animal models.

1.5 Single-cell RNA sequencing (scRNA-seq)

1.5.1 Introduction to scRNA-seq

In recent years, scRNA-seq technology has emerged as a transformative tool in molecular biology, enabling granular analysis of individual cells. Traditional bulk RNA sequencing methods provide valuable insights into average gene expression profiles within a population of cells however they fail to capture the heterogeneity present among individual cells. scRNA-seq overcomes this limitation by allowing researchers to analyse the full transcriptomes of thousands to millions of single cells simultaneously, allowing insights into cellular heterogeneity, developmental trajectories and disease pathogenesis through computational analysis (159-161).

Since the inception of scRNA-seq (162), multiple methods have been developed over the last decade. The droplet-based scRNA-seq approach (**Fig. 1.4**) begins by isolating and encapsulating individual cells within microscopic droplets. Each droplet contains gel beads, each tagged with a unique barcode. As the cells are lysed within these droplets, their mRNA is released and captured by the gel beads, along with a unique barcode. This barcode becomes incorporated into cDNA during reverse transcription, retaining information about gene expression. The cDNA is then labelled with another barcode, known as a unique molecular identifier (UMI), to identify their cell of origin. All the cDNA molecules from the droplets are then pooled together and sequenced using high-throughput sequencing machines (163).

Computational analysis entails decoding barcodes to decipher gene expression profiles of individual cells based on a reference genome. Conducting supervised and unsupervised clustering analyses enables the categorisation of cells with similar transcriptome profiles, forming clusters that represent distinct cell types or states. Furthermore, through differential expression and differential abundance analyses, researchers can identify alterations in cell type composition and transcriptomic variances within identical cell types and among samples of interest (163, 164).

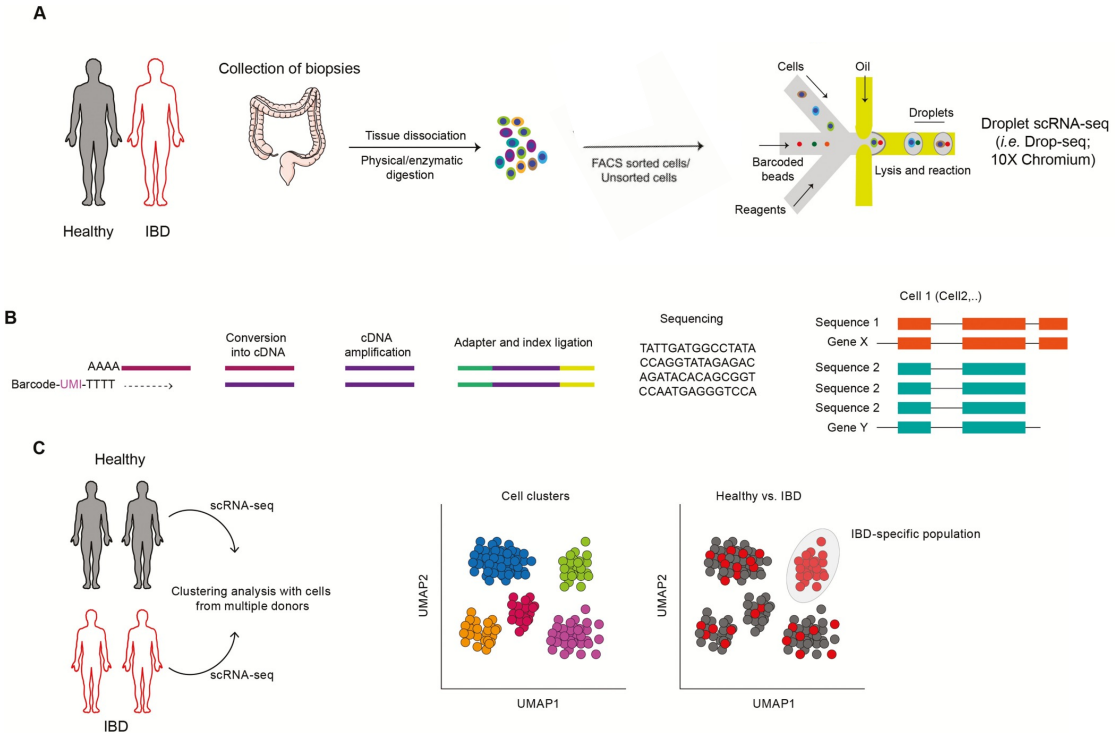


Figure 1.4: Cellular Mapping in Intestinal Mucosa. Intestinal biopsies from healthy individuals and patients with IBD are dissociated to create cell suspensions, which are then further sorted into specific subsets (e.g., via FACS) and processed using droplet-based scRNA-seq (A). The RNA from individual cells is converted into cDNA, amplified, tagged with adapters, and pooled for sequencing on high-throughput platforms (B). Computational analysis of the scRNA-seq data enables unbiased identification of novel cell subsets, helping to discover rare cell populations and pinpoint disease-specific subpopulations (C). Adapted with permission from Corridoni et al., 2020. (164) with permissions.

A primary advantage of scRNA-seq is its ability to discover previously unrecognised cell types and subpopulations within complex tissues in a highly sensitive and unbiased fashion. Moreover, by capturing gene expression profiles from individual cells, scRNA-seq enables researchers to characterise transcriptional states, regulatory networks and signalling pathways within specific cell populations. Furthermore, scRNA-seq offers insights into lineage relationships and cell fate decisions during development and tissue regeneration through the analysis of gene expression profiles across different cell types and developmental stages (165).

However, whilst scRNA-seq offers numerous benefits, it is not without its limitations. Technical challenges such as cell capture efficiency, RNA amplification biases, and dropout events may introduce noise (166) and artefacts into scRNA-seq datasets, thereby complicating data interpretation

and analysis (164). Additionally, the high costs associated with scRNA-seq experiments may be prohibitive, limiting accessibility to this technology.

1.5.2 Rationale for the utility of scRNA-seq in this project

Despite some limitations, scRNA-seq technologies has significantly advanced our understanding of the pathophysiology of IBD (167-171). Within our group, we have utilised scRNA-seq to identify pathogenic remodelling of colonic mesenchymal cells in UC (168) and to characterise the functional interrelationships of colonic $CD8^+$ T-cells in UC compared to health (170). Additionally, we identified a novel population of absorptive colonocytes (BEST4/OTOP2 cells) characterised functionally by luminal pH sensing, which were found to be depleted in UC (167).

CD is characterised by significant cellular heterogeneity within affected intestinal tissues, whereby immune dysregulation, epithelial barrier dysfunction and stromal remodelling contribute to disease progression; the ability to capture this heterogeneity at high resolution has led to some important discoveries. Martin et al., published one of the first scRNA-seq papers in CD, contributing insights into anti-TNF therapy resistance by identifying a distinctive cellular module in affected ileum (consisting of IgG plasma cells, mononuclear phagocytes, activated T-cells and stromal cells) associated with therapy response (172). Jaeger et al. further dissected the T-cell composition of Crohn's terminal ileum, highlighting the involvement of specific T-cell subsets, including Th17 cells, in disease pathogenesis (173). Further, Yokoi et al. classified tissue-resident memory T-cells in CD, revealing a distinct $CD4^+$ subset implicated in tissue-specific immune responses (174). Lastly and most recently, Xavier and colleagues unveiled extensive transcriptomic alterations in colonic tissue of individuals with CD, particularly in stromal and immune compartments, implicating key regulatory genes such as *CHMP1A*, *TBX3* and *RNF168*. These pivotal studies in CD shed light on the intricate cellular landscape underlying the pathology of CD (175).

It is important to highlight that the majority of these studies relied on biopsy tissue, which typically sample the mucosa and superficial submucosa primarily, potentially overlooking pathological cell populations residing deeper, such as the submucosa and muscularis propria. In addition, there are inherent problems in the impact of cellular isolation in the generation of intestinal scRNA-seq data. The enzymatic digestion and chelation processes employed to liberate epithelial crypts as well as immune and non-immune cells from the lamina propria are necessary but may inadvertently influence the viability and transcriptome of the isolated cells. As such, optimisation of cellular dissociation protocols is imperative to ensure the acquisition of high-quality single-cell data.

At the inception of this project and up to the present moment, there exists a notable absence of data at a single-cell resolution elucidating fistulating CD. The analysis of fistulating CD tissue is particularly well-suited to single-cell approaches. Profiling these highly heterogeneous samples will facilitate the discovery of rare cell types or states, elucidate transcriptional regulatory networks and unravel cell-to-cell interactions that uniquely contribute to pathology.

1.6 Spatial transcriptomics (ST)

1.6.1 Introduction to ST

While scRNA-seq methods provide invaluable insights into the gene expression of individual cells, they are limited in their ability to provide spatial context, which makes deciphering the spatial organisation of gene expression patterns and microenvironmental influences within complex tissues challenging. ST overcomes this limitation by enabling researchers to simultaneously characterise gene expression data whilst retaining information of the spatial tissue context using cutting-edge technology (176).

ST, based on spatial barcoding technologies, involves capturing the spatial distribution of RNA molecules within tissue sections by first fixing the tissue onto a surface, spatially barcoding the tissue

with UMIs and performing RNA sequencing on the generated cDNA. Decoding of barcoding and alignment with tissue morphology allow for the identification of gene expression patterns while retaining spatial information (177). Most protocols are compatible with both fresh frozen tissue and formalin-fixed paraffin embedded (FFPE) samples. A recent significant advance is the advent of subcellular ST (178), which enables the precise mapping of molecules at subcellular resolution within intact cells. This approach typically involves fluorescently labelled oligonucleotide probes to target RNA within the tissue, followed by specialised imaging techniques to visualise and quantify RNA expression with subcellular resolution.

ST facilitates the unbiased mapping of gene expression profiles to specific anatomical structures and cellular components, revealing spatially localised molecular signatures associated with specific cell types, microenvironments and disease states. Furthermore, ST enables the identification of spatially coordinated gene expression patterns and cellular interactions within tissues. By analysing gene expression data across neighbouring cells and tissue regions, one can infer cell-to-cell communication networks, spatially resolved signalling gradients and spatially restricted molecular interactions (177).

However, despite its advantages, ST has its limitations. Technical issues with tissue samples such as tissue fixation, sectioning, tissue artefacts or tissue RNA quality may impact the yield of RNA recovered and the quality of ST data. Moreover, the large volumes of data generated and the complexity of spatial data integration required necessitate specialised computational methods.

1.6.2 Rationale for the utility of ST in this project

The method of ST has been widely adopted in various fields including cancer, immunology, developmental biology and neuroscience (179-185). However, its application in the field of IBD remains in its infancy, with limited studies conducted thus far (186, 187). For instance, Garrido-Trigo et al. utilised subcellular ST to demonstrate the co-localisation of inflammatory fibroblasts with inflammation-dependent activated macrophages in the inflamed colons of patients with CD (186).

Spatially resolved information regarding fistulating CD would significantly enhance our understanding of the complex alterations associated with the pathology, including stromal remodelling, inflammatory infiltrates and epithelial damage. This is especially critical in the case of fistulae, a pathology that traverses every layer of the bowel wall and, therefore, lends itself well to examination with spatial integrity intact. By applying ST to fistulating CD tissue within this project, one can potentially reveal spatially localised molecular signatures associated with disease pathology and progression.

1.7 Project aims

Fistulating CD presents a complex clinical entity, carrying significant morbidity, mortality and economic burden to healthcare systems. The development of targeted treatments for fistulae relies upon a deeper understanding of their pathogenesis. However, understanding their pathogenesis remains challenging due to limited access to human tissue from affected areas. Additionally, the absence of accepted *in vitro* models that recapitulate the cellular lesions associated with fistulae, along the scarce *in vivo* models to study this phenomenon, further complicates elucidating its underlying mechanism.

Existing literature suggests potential pathways involved in the pathogenesis of CD-associated fistulae, including the transition of IECs to mesenchymal-like cells, upregulation of MMPs, and overexpression of invasive molecules. However, current data on EMT primarily relies on techniques such as RT-PCR, co-culturing and immunohistochemistry. The genetic basis for this phenotype remains elusive, except in cases of early-onset monogenic IBD, where some insights have been gained (88, 188). However, the role of somatic mutations in this aspect of CD remains unclear. Furthermore, the contribution of microbiota or environmental factors to fistula development remains unclear.

Employing multimodal transcriptomic analysis holds promise in elucidating the cellular circuits driving CD-associated fistulae. In this project, I propose to define the distinct molecular mechanisms involved in the development of fistulating CD using complementary scRNA-seq and ST technologies. In doing so, I aim to expand our understanding of the fistulating CD pathogenesis and identify and pathways which may serve as therapeutic targets.

To achieve this, the project aims are as follows:

1. *Unbiased Characterisation of Cellular Landscapes*

The primary objective is to comprehensively characterise the cellular landscapes of human fistulating CD tissue using scRNA-seq. By employing multi-modal single-cell profiling, this aim seeks to analyse resected full-thickness intestinal tissue from individuals with fistulating CD alongside tissue from non-fistulating CD and healthy controls. Through comparative analysis against controls, the aim is to discern unique pathways associated with fistulating CD, providing insights into the underlying molecular mechanisms driving fistula formation.

2. *Spatial Characterisation of Gene Expression Profiles*

Utilising multimodal ST approaches, the aim is to spatially characterise distinct gene expression profiles within human fistulating CD tissue. By integrating scRNA-seq data, this objective intends to map and localise specific pathways in fistulating CD, elucidating the cellular crosstalk between the epithelium and stroma. This spatial characterisation will offer insights into the spatial organisation and molecular alterations associated with fistula formation and progression.

3. *Novel Classification of Fistula Subtypes*

The goal is to establish a novel classification of fistula subtypes based on underling molecular mechanisms. Through the generation of an integrated transcriptomic atlas of fistulating CD pathology, the aim is to identify and characterise diverse mechanistic subtypes of fistulae. This classification will provide a deeper understanding of molecular heterogeneity within fistulating CD, potentially enabling more precise diagnosis and treatment strategies tailored to specific subtypes.

Chapter 2: Methods

2.1 Human tissue samples

Informed written consent was obtained (Research Ethics Committees (REC) references: 18/WM/0237 Targeting Immune Pathways (TIP); 16/YH/0247 GI Illness and 09/H1204/30 IBD Biobank) from adult patients undergoing elective and emergent IBD surgery with resection. 'Healthy', non-IBD, control tissue was obtained from patients undergoing elective colorectal cancer (uninvolved tissue) or stoma reversal surgery. Patient metadata are summarised in Appendix (**Appx. Table A.3**).

Archived, diagnostic FFPE blocks were also obtained for spatial transcriptomics and validation experiments. Blocks were sourced from Oxford University Hospitals NHS Foundation Trust under Oxford Research Biobank (ORB) (REC:19/SC/0173) via the Oxford Centre for Histopathology Research (OCHRe). Blocks were additionally sourced from collaborators from the Institute of Pathology, Friedrich-Alexander University Erlangen-Nuremberg, Klinikum Bayreuth, Germany (Ethics number: 23-131 bp).

2.2 Tissue handling

Full thickness (FT) surgical tissue samples were collected by the operating surgeon from tissue that exceeded clinical histopathological needs. In each case, the operating surgeon would macroscopically

identify involved (diseased) and uninvolved tissue for sampling. Pathology status was subsequently confirmed additionally upon review of the formal clinical histopathology report.

Tissue samples, stored in DMEM (Gibco) supplemented with 10% Fetal Calf Serum (FCS) (Sigma-Aldrich), 100U/ml of Penicillin and 100U/ml of Streptomycin (Sigma-Aldrich), were transported on ice to the lab.

2.3 Tissue processing and storage

Samples were washed with Phosphate Buffered Saline (PBS, Gibco). Once the anatomical landmarks were defined, samples then underwent blunt dissection on ice to preserve samples as FT, Submucosa (SM) and Mucosa (MUC) layers. Each corresponding layer was then processed for cryopreservation, formalin fixation and as cryosections for spatial transcriptomics (ST). Once FT, SM, MUC were dissected, samples were immediately frozen down in CryoStor® CS10 preservation media (Sigma-Aldrich) and placed in a liquid-nitrogen storage tank for long-term storage. For FFPE, samples were orientated in FFPE cassettes and placed in 10% buffered formaldehyde (VWR International) for forty-eight hours and then transferred to 70% ethanol followed by in-house wax embedding [EB & MJ]. A different fixation protocol was observed for FFPE blocks received from OCHRe and collaborators in Germany. FFPE cassettes were then placed at -20^{o}C for long term storage until required for histological staining, immunohistochemistry or spatial experiments. For cryosections, the desired section was orientated in a cryomould containing OCT sectioning media (CellPath), and then placed in a bath slurry of isopentane (2-methylbutane, Sigma-Aldrich) and dry ice (-80^{o}C). Once solid, the embedded sections were placed in -80^{o}C for storage until required for spatial transcriptomics experiments. For the 'ISO' method, desired samples were placed in a cryovial of isopentane, the sealed cryovial was then flash frozen in liquid nitrogen. The sealed cryovial was then placed in -80^{o}C for long term storage until required for spatial experiments.

2.4 Human intestinal epithelial crypt chelation protocol

The epithelial cell dissociation protocol was performed by adapting previously described protocols (167, 189). In brief, cryopreserved FT or MUC tissue samples were thawed for a minute in a 37°C waterbath then washed in 15mls of cold, supplemented DMEM (DMEM (High glucose, Gibco) with Penicillin and Streptomycin, HEPES (Gibco), FCS, NEAA (Gibco) and Sodium Pyruvate (Sigma-Aldrich)) by centrifugation. All samples were cut into small fragments followed by further centrifugation. The sample was transferred into pre-warmed (37°C) chelation medium (Hank's balanced salt solution (HBSS, Lonza) supplemented with Penicillin and Streptomycin, HEPES, 5mM Ethylenediaminetetraacetic (EDTA, Invitrogen), 2mM Dithiothreitol (DTT, Thermo Fisher) and FCS) and vortexed. The sample was then incubated at 37°C for a total of twenty minutes with agitation by vortexing every three minutes. After incubation, the tissue pieces were removed to leave only the epithelial crypt-containing supernatant. Supplemented DMEM was added to the supernatant before centrifugation. The crypt pellet was re-suspended in DMEM. This chelation protocol was completed a total of two times on the same tissue pieces to increase yield of isolated crypts. The pooled epithelial crypt suspension was then processed into a single cell suspension with TrypLE Express (Gibco) and DNAse (Sigma) for thirty minutes at 37°C. The obtained single cell suspension was then filtered through a 70uM filter, washed and resuspended in supplemented DMEM. Where necessary, additional filtering through a 40um filter was required to improve viability. The sample was then counted at least twice for viability (Countess II, Thermo Fisher). Between 0.5-1 million, highly viable, cells per sample were taken forward for cell staining (see **Methods 2.6**).

2.5 Human intestinal lamina propria (LP) isolation

Digestion of intestinal tissue for lamina propria (LP) isolation was performed by adapting the Miltenyi Biotech LP dissociation kit protocol. I conducted several experiments to optimise this protocol for FT ileal tissue (both diseased and healthy) (See **Chapter 3.2**). Namely the optimisations included

increasing the 37°C incubation duration and adjusting the proprietary enzyme concentrations. The digestion enzyme cocktail was optimised for 300mg of tissue.

Cryopreserved FT tissue samples were thawed for 1 minute in 37°C water bath then washed in 20mls of cold, supplemented DMEM (see above) by centrifugation. After discarding the supernatant, all samples were cut into small fragments. The weight of the tissue was deduced by weighing the tissue containing 50ml falcon and subtracting the weight of an empty 50ml falcon. The number of digest fractions required was determined by the total weight (mg) divided by 300mg. The tissue fragments were then resuspended in Buffer L (DMEM (High glucose), FCS and HEPES). The volume of Buffer L for resuspension was dependent on the total number of digest fractions required (500ul Buffer L per digest fraction). The digestion enzyme mix (50ul of Enzyme D, 25ul of Enzyme R and 12.5ul of Enzyme A per 200ul of pre-warmed Buffer L) was then added to 500ul Buffer L plus tissue sample and placed in one well of a 24-well plate. The pre-warmed (37°C) digestion plate was then placed in an incubator (37°C) for two hours. Every twenty minutes the digest was mechanically dissociated using a blunt needle and 2.5ml syringe by aspirating and expelling the digest sample repeatedly. The needle was washed with Buffer L pre- and post-mechanical dissociation to remove any adherent cells. Post two-hour digest, the sample was filtered through a 100uM filter and washed with DMEM (High glucose) followed by centrifugation. The single cell stromal suspension was then counted twice for high viability by Countess II with Trypan Blue staining (Themo Fisher). Up to 2 million cells were then transferred into one well of a 96-well plate for cell staining (**Methods 2.6**).

2.6 Cell staining, flow cytometry and sorting

Flow cytometry was used for analysis of digestion efficiency and sorting cells into a single cell suspension of immune ($CD45^+/EpCAM^-$) and stromal ($CD45^-/EpCAM^-$) cells. **For digestion efficiency experiments**: cells were stained with CD45 (FITC, H130), CD90 (PE, 5E10), CD3 (APC, OKT3), EpCAM (BV785, 9C4), Live/Dead (Zombie Violet) (BioLegend, see **Methods**

2.16). Compensation controls were generated using single colour controls and fluorescence minus one (FMO) samples for accuracy of gating using the BD LSR II (BD Biosciences).

For epithelial scRNA-seq staining: 0.5-1million cells per sample were suspended in DMEM (supplemented) and incubated with TruStain FcX (BioLegend) for 10 minutes at 4^{o}C degrees, to block non-specific binding. Cells were then stained with a Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-Seq) antibody mix (CD4, CD8, CD45RO, CD45RA, PD1, CD103, CD56 and CD3), to identify and exclude any contaminating immune subsets or intraepithelial lymphocytes (IELs) and also stained with hashtag-oligo (HTO) antibodies (BioLegend, **Methods 2.16)** for 30 minutes at 4^{o}C. Hashing allows for multiplexing of samples; the number of hashing antibodies required depended on the number of samples used per run. After staining, cells were immediately spun down, counted, and resuspended at the desired pooling concentrations and loaded on the 10x Chromium platform (**Methods 2.8**).

For stromal scRNA-seq staining: 0.5-1million cells per sample were suspended in DMEM (supplemented) Trustain FcX for 10 minutes at 4^{o}C degrees, to block non-specific binding. Cells were then stained with an antibody mix of CD45, EpCAM, CITE-Seq (CD4, CD8, CD45RO, CD45RA, PD1, CD103, CD56 and CD3) and required HTO antibodies for 30 minutes at 4^{o}C with a subsequent DAPI stain (1:1000, BD Biosciences Cat#564907) just prior to FACS sorting (BD FACSAria III and BD FACSAria Fusion (BD FACS Diva Software), BD Biosciences). Compensation controls were generated using single colour controls and FMO samples were used for sorting gates. After gating, live cells were sorted into two populations: immune (CD45^{+}/EpCAM^{-}) and stromal (CD45^{-}/EpCAM^{-}), and collected in Eppendorfs containing FCS. Cells were immediately spun down, counted, resuspended at the desired pooling concentrations and loaded on the 10x Chromium single-cell platform (**Methods 2.8**).

Flow cytometry data was analysed using FlowJo software (v10.8.1).

2.7 LeviCell platform for stromal cell isolation

LeviCell (LevitasBio™) platform provides an alternative cell sorting strategy for more complex tissue samples. This strategy was employed as an alternative method for stromal cell enrichment. In brief, FT tissue samples underwent both epithelial crypt chelation and LP digestion as previously described (**Methods 2.4, 2.5**). Remaining LP cells were then stained with respective cell hashing antibodies, counted and pooled in a one-to-one ratio. Pooled cells were then resuspended in proprietary Levitation Buffer and CD45 depletion beads (magnetic nanospheres conjugated with an anti-human CD45 antibody) as per the LeviSelect Human CD45 Depletion Kit protocol (PN 1004001, LevitasBioTM). The cell suspension was then loaded on to the LeviCell-1.0 platform which employs the use of paramagnetic fields to separate live cells from dead or dying cells. $CD45^+$ cells in suspension remain bound to the LeviCell cartridge, meanwhile the $CD45^-$ population undergoes live cell enrichment, with live cells in the suspension collected in the top output well of the cartridge. The bottom output well comprises primarily dead $CD45^-$ cells as the $CD45^+$ cells remain immobilised inside the cartridge. The isolated stromal fraction is therefore defined as live, $CD45^-$ and epithelial crypt-deplete, digested lamina propria cells. The stromal fraction isolated by LeviCell was then loaded directly on to the 10x Chromium single-cell platform.

2.8 Droplet based scRNA-sequencing

After cell dissociation, staining and sorting, cells were pooled in a one-to-one ratio (where possible) and suspended in sorting buffer (PBS supplemented with 0.04% BSA) and loaded on the 10x Chromium single-cell platform at a final concentration of 1,000 cells/ul as per protocol (10x Genomics, Single cell 5' kit, v1.1, CG000208 Rev F). A total of 30-40,000 cells were loaded per pool. The process involved capturing individual cells within gel beads in emulsion (GEMs), barcoding, GEM reverse transcription, cleaning up the reverse-transcribed complementary DNA (cDNA), amplifying cDNA, and constructing libraries for gene expression (GEX) and antibody-derived tagged (ADT) data. This entire workflow, including the generation of T-cell receptor (TCR) libraries from

CD45+ (immune) and epithelial pools, was performed using 10x Genomics library kits according to the manufacturer's instructions. Library quality was subsequently assessed to ensure compliance with downstream sequencing requirements, employing Agilent's Bioanalyzer TapeStation for this purpose.

2.8.1 Jumpcode technology

GEX libraries underwent CRISPR RNA depletion using the Cas9/single guide RNA ribonucleoprotein complex provided within the CRISPRclean Single Cell RNA Boost Kit (KIT1018 v1.0, Jumpcode Genomics). The CRISPRclean workflow aims to deplete unaligned reads, highly expressed protein-coding ribosomal and mitochondrial genes, and non-variable genes, redistributing 50% of sequencing reads to biologically relevant transcripts. This process maximizes gene and UMI sensitivity.

The CRISPRclean protocol was integrated into the 10x Chromium Single Cell 5' kit, v1.1 protocol. During Step 6.4 – Post Ligation Cleanup, the sample was eluted into 16 ul of Nuclease-Free water instead of Buffer EB. Subsequently, the CRISPRclean protocol initiated, involving ribonucleoprotein complex formation, CRISPR digestion, and size selection. Once the CRISPRclean protocol concluded, the Chromium Single Cell 5' kit, v1.1 protocol resumed at Step 6.5 – Sample Index PCR and continued until the completion of GEX library construction.

2.9 ST fresh frozen – 10x Genomics Visium protocol

RNA was extracted from representative cryosections (10µm thickness) using the RNeasy® Plus Micro kit (Qiagen) as per manufacturer's instructions. The RNA Integrity Number (RIN) was assessed using a high sensitivity RNA ScreenTape assay in a 4200 TapeStation (Agilent). The RNA 6000 Pico assay (Agilent) was used for samples with lower RNA concentrations (quantitative range: 50-5000 pg/ul) as identified at Nanodrop. All selected samples for ST had both a good representative section at Haematoxylin & Eosin (H&E) (**Methods 2.14.2**) and RIN ≥7.8.

Tissue permeabilisation was optimised on uninvolved FT ileal sections stored in ISO using the Tissue Optimisation workflow (10x Genomics Visium TO, CG000238 Rev D). Briefly, 10 µm thick serial sections were placed on a TO slide and incubated with a proprietary permeabilisation enzyme from a range of 3 to 36 minutes (TO Run 1) and 12 to 60 minutes (TO Run 2), determining an optimum digestion timepoint of 12 minutes.

For ST, tissue samples were cryopreserved in OCT or ISO, as previously detailed (**Methods 1.3**). The samples were then processed for the full ST experiment according to the protocol (10x Genomics Visium Spatial Gene Expression, CG000239 Rev D). Briefly, 10 µm sections were placed on the Visium slide (PN 20000233), which contained four capture areas (6.5mm x 6.5mm per square), each with 5000 gene expression spots. Sections were then subjected to methanol fixation followed by H&E staining and imaging. The tissue then underwent digestion at the optimised time point of 12 minutes to release cellular mRNA captured by primers on the gene expression spots. Reverse transcription and cDNA synthesis occurred on the slide with a corresponding spatial barcode. cDNA quantification, determined by qPCR (KAPA SYBR FAST), informed DNA amplification cycles. Libraries were then constructed from the cDNA as per the protocol at a resolution of 55 µm per spot. Samples were pooled at a 4nM concentration, with the proportion of each sample representative of the percentage tissue coverage area (ratio of section area divided by capture area). Spatial libraries were then sequenced on an Illumina NextSeq 500 platform.

Slides for TO and ST were imaged on a Leica DM18 inverted microscope, for fluorescence, and Zeiss AxioScan for matched H&E information at 10X and 20X magnifications.

2.10 ST FFPE – 10x Genomics Visium protocol

Five to ten scrolls (5µm thickness) of each FFPE block were collected and stored at -80°C for subsequent RNA quality checks. Total RNA extraction followed the Qiagen RNeasy FFPE kit (PN

73504) protocol according to the manufacturer's instructions. The RNA Pico assay (Agilent) quantified DV200 values, with samples exceeding 50% deemed acceptable for spatial analysis.

All FFPE blocks underwent H&E staining to assess tissue morphology and pathology representation. Cases meeting the criteria of acceptable DV200 values and desirable tissue morphology were selected for spatial analysis.

The sample preparation workflow for 10x Genomics Visium Spatial Gene Expression (FFPE, CG000408, and CG000409 Rev D) followed the manufacturer's instructions with optimisations. Deviations from the protocol included storing FFPE blocks at -20°C, extending slide drying time to 48 hours at room temperature, and modifying deparaffinization and alcoholic rehydration series to reduce the risk of tissue detachment.

For the ST experiment, 10 µm sections on Visium FFPE spatial GEX slides (with the same capture area and number of gene expression spots as the FF Visium slide) underwent deparaffinization, H&E staining (imaged on a Zeiss AxioScan at 10X and 20X magnifications), and decrosslinking. Probe hybridization targeting specific genes and cDNA generation followed the protocol outlined in the 10x Genomics Visium Spatial Gene Expression for FFPE User Guide (CG000407). Subsequently, library construction, quality control, and pooling were carried out as previously described in the FF workflow (**Methods 2.9**). The resulting spatial libraries underwent sequencing on an Illumina NextSeq 500/550 platform (High Output v2.5, 150 cycles) with the following sequencing parameters: Sequencing Read 1 (28 cycles), i7 index (10 cycles), i5 index (10 cycles), and Read 2S (50 cycles).

2.11 Subcellular ST for FFPE tissue samples – MERFISH (Vizgen)

2.11.1 Gene selection for panel

We designed a customised panel of 500 genes to identify cell populations of interest with MERFISH. The full list of genes can be found in the Appendix (**Appx. Table A.2**). The 500-plex panel was

curated by using existing scRNA-seq datasets generated by the Simmons group including the scRNA-seq data derived for this work. The customised panel was designed to incorporate gastrointestinal markers capturing key physiological and disease states manifest in the intestine. Initially, a 300-gene core pan-gut panel was derived including broad lineage markers (e.g., EPCAM, LGR5 – epithelial; BMP4, WNT5A – stromal; and CD45 – immune) as well as very rare cell subtypes (e.g., CHGA – enterochromaffin cells, TNFRSF11B – M-cells). Secondly, we curated literature sets for cells which cannot be adequately captured by droplet-based scRNA-Seq – e.g., Paneth cells, neutrophils, neurons. In addition to the 300-gene core, gene markers of interest for key intestinal physiological and disease states as identified by scRNA-seq analysis were included. We also designed with in-built redundancy of 5-10 specific markers to enhance the reliability and robustness of transcriptomic analyses. To ensure the effectiveness of this gene panel for MERSCOPE imaging, it underwent a thorough evaluation via Vizgen's MERSCOPE Gene Panel Design Portal. This evaluation sought to confirm that each gene possessed adequate length to accommodate the necessary encoding probes for binding, while also ensuring that the overall panel adhered to abundance thresholds, preventing optical crowding during MERSCOPE imaging sessions. This stringent evaluation process culminated in a final selection of 500 genes.

2.11.2 FFPE tissue MERSCOPE workflow

FFPE blocks with DV200 >40% and representative tissue morphology and pathology were selected. FFPE blocks were sectioned at 5 µm thickness using the Leica RM2135 Microtome and positioned onto a MERSCOPE FFPE slide (PN 10500102) following the guidelines outlined in Vizgen's MERSCOPE User Guide for FFPE samples (Vizgen MERSCOPE User Guide, Formalin-Fixed Paraffin-Embedded Tissue Sample Preparation, 91600112, Rev B). The tissue sections were dried at 55°C for 15 minutes and subsequently for 2-3 hours at room temperature.

To begin the preparation process, the tissue sections underwent dual deparaffinization using Deparaffinization Buffer (PN 20300112) at 55°C for 5 minutes, followed by three washes in 100% ethanol for 2 minutes each, succeeded by 2 minutes each in 90% ethanol and a 2-minute rehydration

step in 70% ethanol. Post-rehydration, the tissue sections were treated with Decrosslinking Buffer (PN 20300115) at 90°C for 15 minutes and then allowed to cool at room temperature for 5 minutes.

Following decrosslinking, the sections were incubated with Conditioning Buffer (PN 20300116) at 37°C for 30 minutes, and subsequently with Pre-Anchoring Reaction Buffer (PN 20300117) for 2 hours at 37°C. After the anchoring pre-treatment, the tissue sections were stained for the cell boundary utilising Vizgen's Cell Boundary Kit (PN 10400118) according to the provided user guide. Subsequently, the samples were washed using Formamide Wash Buffer (PN 20300002) at 37°C for 30 minutes and incubated with Anchoring Buffer (PN 20300117) at 37°C for 18 hours.

Post incubation, the samples underwent a wash prior to gel embedding, and clearing. Samples were set in a gel formed by combining Gel Embedding Premix (PN. 20300118), 10% ammonium persulfate (Millipore-Sigma), and N,N,N′,N′-tetramethylethylenediamine (Millipore-Sigma). Gel coverslips were treated with RNaseZap (Thermo Fisher) and 70% ethanol, then coated with Gel Slick. The samples were incubated with the gel solution, then allowed to polymerise for 1 hour at 30°C after excess solution removal by automatic aspiration. After removing the gel coverslip, the samples were subjected to a digestion solution composed of Digestion premix (PN 20300005) and RNase Inhibitor (NEB) for 6 hours at 37°C. Following digestion, samples were incubated with a Clearing solution (Protease K, NEB and Clearing Premix, PN 20300114) at varying temperatures overnight.

After tissue clearing, the samples were treated with MERSCOPE Photobleacher (PN 10100003) for 6 hours, washed again using Formamide Wash Buffer at 37°C for 30 minutes, and subsequently incubated with our custom designed MERSCOPE 500-Gene Panel Mix (Gene ID: CP0835, PN 10400003) for 2 days at 37°C.

2.11.3 Sample imaging

Post probe hybridisation, samples were incubated with Formamide wash buffer for 30 minutes at 47°C twice and finally washed with Sample Prep Wash Buffer (PN 20300001). Afterward, they were exposed to a mixture of DAPI and polyT Reagent (PN 2030021) for 15 minutes at room temperature, followed by a 10-minute wash with Formamide Wash Buffer and transferred to Sample Prep Wash Buffer. The Imaging Activation Mix was created by combining Imaging Buffer Activator (PN 20300022) and RNase inhibitor (NEB). The MERSCOPE system (PN 10000001) was used to load the imaging reagents and processed samples. Initially, a low-resolution DAPI mosaic at ×10 magnification was conducted, and subsequently, specific areas of interest were chosen for high-resolution imaging at ×60 (a total area of $1cm^2$) in accordance with the MERSCOPE Instrument User Guide (Vizgen MERSCOPE Instrument User Guide, 91600001, Rev G). The resulting data were employed for cell segmentation and analysis (**Methods 2.13.2.3**).

2.12 Subcellular ST for FFPE tissue samples – Xenium (10x Genomics)

2.12.1 Gene selection for panel

The curated list of genes chosen for the Xenium panel was identical to that of the MERSCOPE 500-plex panel (refer to **Section 2.11.1**), with the exception of 20 genes. These exclusions were made to accommodate the space limitations of the Xenium panel, which is a 480-plex panel (**Appx. Table A.2**). The final gene list underwent evaluation through the Xenium Panel Designer tool portal.

2.12.2 FFPE tissue Xenium workflow

Sample preparation

FFPE blocks representative tissue morphology and pathology were selected. FFPE tissue sections were prepared in accordance with manufacturer's instructions (10x Genomics, Xenium Tissue Preparation FFPE Demonstrated Protocol, CG000578, Rev C). Briefly, FFPE blocks were sectioned

at 5 µm thickness and placed on Xenium slides (PN 1000460) within a capture area measuring 235 mm^2. Sections were dried with a fan at room temperature for 30 minutes before being baked at 42°C for 3 hours. Sections were dried for 48-72 hours with desiccant beads at room temperature prior to processing.

Experimental workflow

Sections were then deparaffinized and decrosslinked as per protocol (10x Genomics, Xenium, Sample Preparation FFPE Demonstrated Protocol, CG000580, Rev C) to release sequestered mRNA within the tissue. The sections were then processed as per assay workflow, firstly "Probe Hybridization, Ligation & Amplification" (10x Genomics, Xenium, Probe Hybridization, Ligation & Amplification, User Guide, CG000582, Rev E) followed by imaging. Briefly, sections were incubated with custom designed Human Colon gene probes (P8GA2K_hColon_480g) at 50°C overnight. Subsequent washing (with PBS-T and Post Hybridization Wash Buffer (PN 2000395)) was conducted to eliminate unhybridized gene probes, followed by probe ligation and annealing of rolling circle amplification primers at 37°C for 2-hours. The resultant circularised probes underwent enzymatic amplification for 2-hours at 30°C. Sequential washing steps aimed to quench background fluorescence chemically. Finally, nuclei are stained with DAPI (Xenium Nuclei Staining Buffer) prior to imaging.

Imaging and decoding

The tissue sections were imaged utilising the automated Xenium Analyzer instrument (10x Genomics, Xenium Analyzer User Guide CG000584). The imaging process encompassed 15 cycles. During each cycle, fluorescently-labelled oligonucleotides were bound to amplified barcodes, and their intensity in the four Xenium colour channels was measured. An optical signature unique to each gene was generated, aiding the identification of target genes by a process of decoding. Decoded transcripts possessing a quality value (Q-score) >20 were retained for further analysis. Z-stacks, with a 0.75 µm step size covering the tissue's entire thickness, were meticulously captured and subsequently reconstructed into a spatial transcript map across the tissue. Cell boundaries were

defined, and transcripts were allocated to cells through the process of performing cell segmentation. This involved using DAPI images derived from Xenium Nuclei Buffer-stained samples and employing a neural network. The pipeline generated data files that were exported for downstream analyses (**Methods 2.13.2.3**).

Post-Xenium run, the slide was processed for H&E staining as per protocol (10x Genomics, Xenium, Post-Xenium Analyzer H&E Staining, CG000613 | A) to enable direct alignment of gene expression data with histological data from the same tissue section. Slides were imaged on the Zeiss AxioScan Z1 at 10X and 20X magnifications.

2.13 Sequencing & computational analysis

2.13.1 Sequencing

Library quality, size and concentration were assessed using Bioanalyzer 4200 TapeStation (Agilent) and Qubit® 4.0 Fluorometer (ThermoFisher). The final library pools, single cell and spatial, were pooled at a 4nM concentration and sequenced on either an Illumina NextSeq 500/550 (High output v2.5, 150 cycles) or an outsourced NovaSeq PE150. Library sequencing depth and run parameters for were followed as per manufacturer's instructions (10x Genomics, Single cell 5' kit, v1.1, CG000208 Rev F).

2.13.2 Computational processing and analyses

The computational analysis of scRNA-seq and ST data was principally undertaken by Dr. Agne Antanaviciute [AA] with my active participation.

2.13.2.1 scRNA-seq analysis

For scRNA-seq analysis, raw sequencing data were demultiplexed using Illumina bcl2fastq software before undergoing initial quality checks using fastQC. Raw reads were then processed, aligned to the human hg38 reference genome (190) and transcript unique molecular identifier

(UMI) counts were summarised using the Cellranger pipeline from 10x Genomics. Raw UMI count matrices are imported into R for further analysis. Cell calling and index swapping are carried out using DropletUtils R package. Further quality control measures included filtering for empty droplets, cells with high mitochondrial content and low total UMIs.

Downstream analysis was carried out using Seurat R package (191) – data were normalised and highly variable genes were identified. Harmony algorithm (192) was used to batch-correct the pooled data. Harmonised principal components were selected for clustering using Scree plots, with the top components used as input to Louvain clustering algorithm. Clustering solutions were visualised using Uniform Manifold Approximation and Projection (UMAP) visualisation. Cluster cell type annotations were ascribed to epithelial, stromal and immune ($CD45^+$) subsets based on marker gene expression, as previously defined. Differential gene expression (DEG) between conditions was carried out using negative binomial testing, blocking for batch covariates where appropriate. Proportion analyses were carried out with base R, using ggplot2 and ggpubr packages for visualisation or using R package sccomp (193). Additionally, we conducted differential abundance analysis based on graphs using the R package miloR (194).

CITE-Seq antibodies & Hashed sample de-multiplexing: The CITE-Seq and HTO antibody panel data were processed concurrently with scRNA-seq as matched feature barcode libraries. The resulting UMI count matrices were imported into R for analysis. Hashed sample demultiplexing involved filtering for high-quality cells, normalisation, clustering, and assigning sample-of-origin as previously described (195). The demultiplexed cells were visualised using tSNE plots. The correctness of sample demultiplexing was verified by examining the expression patterns of sex-specific genes e.g., *XIST*, ensuring that they appropriately segregated with the sample-of-origin assignments. Doublets and negative cells were filtered out from downstream analysis.

2.13.2.2 Visium ST analysis

Raw sequence data was demultiplexed using bcl2fastq and fastQC software as before (**Methods 2.13.2.1**). Raw reads were aligned and processed using Spaceranger pipeline. Raw UMI count spot matrices, images, spot-image coordinates and scale factors were imported into R. Spot matrix was filtered to keep only spots overlaying tissue sections. Raw UMI counts are then normalised by fitting a negative binomial regression model to account for the variability in total spot RNA content. Additionally, areas of tissue artefacts (e.g., folds, tears, foreign material) were annotated to exclude the ST spots directly covered by these regions. Spatially variable genes were identified using Moran's I metric and the top 2000 were selected for dimensionality reduction and clustering. Clustering was performed using Louvain clustering algorithm and visualised using UMAPs. Cluster distribution overlay the corresponding H&E image to allow visualisation in a spatial context. Cell type predictions were calculated for each spot using factor analysis in Seurat using single cell populations from previously published reference datasets (167, 168, 179) with the addition of the project's generated reference single cell dataset. Corresponding regions between sections were integrated using Harmony algorithm to correct for slide-specific effects.

Pseudo-bulk PCA analysis was carried out using integrated data in DESeq2 R package, transformed using regularised log transformation and principal components were computed using base R. Pathway activity analysis was performed to identify region-specific cellular events. DEG analysis between sections was carried out using negative binomial generalised linear models (glm). Gene Ontology and pathway enrichment analyses (e.g., REACTOME (196)) were carried out to interpret the biological significance of the results.

Fistula Associated Stroma score: In the ST analysis, we devised a composite score to overlay onto the data, capturing the expression profiles of the Fistula Associated Stroma cluster, as characterised by scRNA-seq. The process involved the identification of the genes constituting the

cluster followed by normalisation of gene expression data to mitigate variations. For each ST spot data, we calculated expression scores for the selected genes, combining them into an aggregate score using a weighted sum approach. The resulting aggregate score was then normalised for comparability across the different tissue sections. This normalised score was then overlaid onto the ST data for further visualisation and analysis.

2.13.2.3 Subcellular ST analysis

The following analysis approach was employed for both MERSCOPE and Xenium analysis. Raw imaging data processing was performed using Vizgen MERSCOPE and 10x Genomics Xenium pipelines. Cell segmentation was refined using transcript-density based segmentation algorithm, Baysor algorithm (197), using nuclei-based segmentation (Xenium) or cell-boundary based Cellpose segmentation (MERSCOPE) as a weighted prior. Cell segmentation, transcript coordinates, cell metadata, and morphology images (MERSCOPE only) were imported into R for further analysis. Filtering excluded cells with low molecule detection counts (< 15 transcripts per cell) and low-quality molecules (Q-score <20 for Xenium) to preserve analysis integrity. Cells with high negative control probes were excluded from the analysis. Non-negative control probe data in the remaining cells was normalised using SCTransform, and PCA dimensionality reduction was applied in R. Harmony algorithm corrected batch effects, and reduced dimension components were utilised for clustering and UMAP visualisation as before.

Major cell lineages underwent sub-clustering for higher resolution population analysis. Doublet cells, which arise from cell segmentation failures, were identified based on mixed lineage marker signatures, and were removed from sub-cluster annotations, ensuring clarity in cluster identities. For gene expression analyses not reliant on cell identity (e.g., niche detection), doublet cells were retained to prevent gaps in local tissue area. Using our scRNA-seq data as a reference, gene subset and label transfer workflows were applied. Some cell types e.g., granulocytes, were manually annotated due to technical challenges in profiling these cells. Negative probe signals were carefully examined, and

marker gene expression considered only if more than one molecule per cell was detected. Spatial niches were defined based on local cell type composition.

2.13.3 Data availability

An interactive resource to explore the project's scRNA-seq and Visium spatial data can be accessed here:

https://simmonslab.shinyapps.io/CD_Single_Cell_2022/

https://simmonslab.shinyapps.io/CD_LEVITAS_STROMAL_RUN/ and

https://simmonslab.shinyapps.io/CD_ST_FFPE_ALL_SLIDES/

Access available on request.

2.14 Histopathology characterisation

2.14.1 Histopathological analysis & annotation

All sections utilised for ST analysis had a corresponding digitised H&E image. Two consultant gastrointestinal pathologists, Dr Eve Fryer [EF] and Prof. Michael Vieth [MV], conducted independent histopathological examinations, with limited prior knowledge of the clinical status of the patients. Ensuring patient confidentiality, each slide contained only non-patient identifiable details. EF and MV provided manual annotations of the slides, enabling precise observations of cellular details, identification of pathology, and analysis of tissue architecture. To enhance the robustness of the findings, both pathologists reviewed the same slides. There was minimal inter-observer variability resulting in increased confidence in the assessments. Additionally, a comprehensive review of the formal corresponding histopathology report, encompassing both macro- and microscopic details, was conducted.

2.14.2 H&E staining protocol

H&E staining was essential to assess the morphology of intestinal tissue and confirm the presence of pathology. The tissue sections, embedded in paraffin, underwent deparaffinization to prepare them

for staining. Initially, the slides were immersed in Histoclear, followed by a 4-minute incubation in absolute ethanol. Rehydration of the sections was achieved through a series of ethanol solutions (100%, 96%, 90%, 85%, 70% ethanol) for 4 minutes each, followed by distilled water for 4 minutes, all conducted at room temperature.

For the Hematoxylin staining, the slides were submerged in Hematoxylin solution (Modified Harris, Sigma) for 30-45 seconds. After staining, gentle rinsing under tap water was performed until the solution ran clear. Subsequently, the slides were dipped in Eosin Y solution (Sigma) for 30-45 seconds, with the duration determined by the colour development similar to Hematoxylin, followed by further rinsing with tap water.

Finally, the slides were air-dried before applying a drop of mounting medium (Leica) onto each section. When covering the sections with coverslips, care was taken to avoid air bubbles. The stained sections were then ready for analysis under a microscope and subsequent scanning using the Zeiss AxioScan Z1 at 10X and 20X resolutions.

For fresh frozen sections, the tissue was placed on a slide and fixed with 4% paraformaldehyde. Subsequently, a wash in distilled H2O preceded the H&E staining process, as described above.

2.15 Validation methods

2.15.1 Immunohistochemistry (IHC)

2.15.1.1 EnVision FLEX (Dako) Protocol

Full-thickness tissue samples, fixed in formalin and embedded in paraffin, were sectioned at 5 µm thickness and placed on Superfrost microscope slides (Avantor) before drying. These tissue slides underwent pre-heating at 60°C for 1 hour to minimise tissue detachment. Subsequently, the slides were deparaffinized in 100% Histoclear and rehydrated through a graded ethanol series. Antigen retrieval was optimised by boiling the slides in either a pH 6 or pH 9 buffer (Dako) within a Decloaking Chamber (Biocare Medical) i.e., heat induced epitope retrieval (HIER). Slides were

cooled to room temperature and endogenous peroxidase activity was blocked by H_2O_2 for 30 minutes. Slides were blocked with 2.5% horse serum (Dako) for 20 minutes, to minimise non-specific binding. Slides were then incubated with the primary antibody for either 1 hour at room temperature or at 4^{o}C overnight (please refer to the **Key Resources Table** for the full list of antibodies (**Methods 2.16**). Following a series of washes (Flex Buffer, Dako), slides were incubated with the secondary antibody conjugated with HRP (Dako). The slides were washed in Flex Buffer to remove unbound secondary antibody. Peroxidase activity was detected using DAB substrate (Dako). Slides were counterstained with Haematoxylin for nuclei detection. Finally, the slides were washed with Flex Buffer and distilled water, followed by drying before mounting with Mounting Medium (Leica). Images were acquired on the Zeiss AxioScan Z1 at 20X magnifications.

2.15.1.2 Image analysis and quantification

Digitised images were uploaded to the Visiopharm system (Version 2023.09 x64) for analysis. Utilising Visiopharm's interface, specific regions of interest (ROIs) were manually delineated to encompass distinct tissue areas (e.g., mucosa versus submucosa) and staining patterns for subsequent analysis.

Customised Analysis Protocol Packages (APPs) were developed and trained within Visiopharm to automate the quantification process. These APPs included tailored algorithms specifically designed to identify and quantify the percentage of DAB positivity within the selected regions. The software was equipped with AI machine learning capabilities, allowing it to recognise and differentiate subtle variations in staining intensity throughout the entire tissue thickness, significantly enhancing the accuracy of quantification. To ensure the accuracy and reproducibility of results, the developed protocols and machine learning models underwent rigorous validation against manual quantification methods and corrective measures.

Post-analysis, data visualisation was performed using Visiopharm's output tools, generating quantitative measurements. Statistical analyses and visual representations were generated using

GraphPad Prism software (Version 9.5.1(528)). An unpaired t-test was employed for non-parametric data, with a significance level set at $p < 0.05$, validating the obtained quantitative data. Mean values are presented with error bars representing the standard error of the mean (SEM).

Although not all analyses were conducted in a blinded manner, experiments were replicated to confirm reproducibility.

2.15.2 qPCR

Quantitative PCR (qPCR) was conducted to validate key markers identified from single-cell and ST data. The Fluidigm Flex Six Integrated Fluid Circuits (IFC) workflow facilitated parallel amplification and detection of multiple targets in a high-throughput manner.

RNA extracted from FFPE tissue blocks (**Methods 2.10**) from a representative cohort including fistulating CD patients and controls (CD, non-CD, and healthy individuals) was used for qPCR.

Fluidigm Biomark platform – Flex Six IFC

Reverse transcription utilised the provided Reverse Transcription Master Mix (Standard Bio, PN 100-6300) to generate complementary DNA (cDNA). Small volumes of selected TaqMan gene expression assays were pooled in Tris-EDTA (TE) buffer (invitrogen) for pre-amplification of the generated cDNA. The resulting pre-amplified cDNA was diluted fourfold in TE buffer and used immediately.

The workflow followed the manufacturer's instructions (https://www.standardbio.com). Briefly, the Flex Six IFC (PN 100-6308) was primed with control line fluid. The sample reaction mix consisted of pre-amplified cDNA, TaqMan Fast Advanced Master Mix (Life Technologies, PN 4444557), and 20X GE Sample Reagent (Fluidigm PN 100-6311) in a 4µl reaction volume. TaqMan gene expression assays were diluted in 2X Assay Loading Reagent (PN 100-5359) to a 10x concentration.

Both 3µl of sample reaction mix and TaqMan assay were loaded into the corresponding sample and assay inlets.

The Flex Six IFC was then loaded onto the HX controller for microfluidic mixing and distribution of samples and assays into the reaction chambers. Following this, the IFC was placed in the thermal cycler for qPCR amplifications as per the GE FLEXSix Fast v2 thermal protocol. After thermal cycling, the IFC underwent the post-run script on the HX controller. Data analysis utilised the Fluidigm Real-Time PCR analysis software (Real-Time PCR Analysis User Guide, PN 68000088). Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and Hypoxanthine-guanine phosphoribosyltransferase 1 (HPRT1) served as endogenous controls. The threshold cycle (Ct) values were utilised to calculate gene expression relative to the endogenous controls. Samples and assays were run in duplicate.

Statistical analyses and visual representations were generated using GraphPad Prism software. Non-parametric data were evaluated using an unpaired t-test with a significance level set at $p < 0.05$ to validate the obtained quantitative data. Mean values are presented with error bars depicting the SEM.

2.16 Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER (Catalogue or Part Number)
Antibodies		
APC anti-human CD3 Antibody, OKT3 clone	BioLegend	Cat# 317317
Brilliant Violet 785™ anti-human CD326 (EpCAM) Antibody, 9C4 clone	BioLegend	Cat# 324237
CD45 Antibody, anti-human, FITC, 5B1 clone	Miltenyi Biotec	Cat# 130-113-117
FITC anti-human CD45 Antibody, H130 clone	BioLegend	Cat# 304005
Human TruStain FcXTM (Fc Receptor Blocking Solution)	BioLegend	Cat# 422302
PE anti-human CD90 (Thy1) Antibody, 5E10 clone	BioLegend	Cat# 328109
Zombie Violet™ Fixable Viability Kit	BioLegend	Cat# 423113
TotalSeq™-C0045 anti-human CD4 Antibody	BioLegend	Cat# 344651
TotalSeq™-C0046 anti-human CD8 Antibody	BioLegend	Cat# 344753
TotalSeq™-C0047 anti-human CD56 (NCAM) Antibody	BioLegend	Cat# 362559
TotalSeq™-C0049 anti-human CD3 Antibody	BioLegend	Cat# 344849
TotalSeq™-C0063 anti-human CD45RA Antibody	BioLegend	Cat# 304163
TotalSeq™-C0087 anti-human CD45RO Antibody	BioLegend	Cat# 304259
TotalSeq™-C0088 anti-human CD279 (PD-1) Antibody	BioLegend	Cat# 329963
TotalSeq™-C0145 anti-human CD103 (Integrin αE) Antibody	BioLegend	Cat# 350233
TotalSeq™-C0251 anti-human Hashtag 1 Antibody	BioLegend	Cat# 394661
TotalSeq™-C0252 anti-human Hashtag 2 Antibody	BioLegend	Cat# 394663
TotalSeq™-C0253 anti-human Hashtag 3 Antibody	BioLegend	Cat# 394665
TotalSeq™-C0254 anti-human Hashtag 4 Antibody	BioLegend	Cat# 394667
TotalSeq™-C0255 anti-human Hashtag 5 Antibody	BioLegend	Cat# 394669
TotalSeq™-C0256 anti-human Hashtag 6 Antibody	BioLegend	Cat# 394671
TotalSeq™-C0257 anti-human Hashtag 7 Antibody	BioLegend	Cat# 394673
Antibodies for Immunohistochemistry		
Anti-F3	AtlasAb AMAb	91235 CL 3806

Anti-Frizzled 5 antibody	Abcam	ab75234-100ug
Anti-Frizzled 6 antibody [EPR25319-149]	Abcam	ab290728-10ul
Anti-FRZB	Abcam	ab108169-100ul
Anti-GRHL3 antibody	Abcam	ab221058-100ul
Anti-LGR5	Protein tech	21833-1-AP
Anti-PDGFR alpha antibody [EPR22059-270]	Abcam	ab203491-10ul
Anti-Periostin antibody [EPR20806]	Abcam	ab215199-10ul
Anti-Wnt5a antibody	Abcam	ab235966-100ul
TaqMan Gene Expression Assay (FAM) - Human		
Assay Id Hs00297962_m1 – Grhl3	Thermo Fisher	Cat# 4331182
Assay Id Hs01021800_g1 – Hes2	Thermo Fisher	Cat# 4331182
Assay Id Hs01076029_m1 – F3	Thermo Fisher	Cat# 4331182
Assay Id Hs00154192_m1 – Bmp2	Thermo Fisher	Cat# 4331182
Assay Id Hs00998018_m1 - Pdgfra	Thermo Fisher	Cat# 4331182
Assay Id Hs00182031_m1 – Lrp5	Thermo Fisher	Cat# 4331182
Assay Id Hs00275833_s1 – Fzd7	Thermo Fisher	Cat# 4331182
Assay Id Hs01055414_m1 – IL11	Thermo Fisher	*Cat# 4331182*
Assay Id Hs00968305_m1 – Mmp3	Thermo Fisher	Cat# 4331182
Assay Id Hs00171149_m1 – CCL19	Thermo Fisher	Cat# 4331182
Assay Id Hs00959010_m1 – SPP1	Thermo Fisher	Cat# 4331182
Assay Id Hs01675818_s1 – TWIST1	Thermo Fisher	Cat# 4331182
Assay Id Hs00537087_m1 – TBX15	Thermo Fisher	Cat# 4331182
Assay Id Hs00258278_s1- FZD5	Thermo Fisher	Cat# 4331182
Assay Id Hs01095627_m1 – FZD6	Thermo Fisher	Cat# 4331182
Assay Id Hs02758991_g1 - GAPDH	Thermo Fisher	Cat# 4331182
Assay Id Hs02800695_m1 - HPRT1	Thermo Fisher	Cat# 4331182
Assay Id Hs00998537_m1 - WNT5A	Thermo Fisher	Cat# 4331182
Assay Id Hs00969422_m1 - LGR5	Thermo Fisher	Cat# 4331182
Assay Id Hs015373504_m1 - WNT4	Thermo Fisher	Cat# 4331182
Assay Id Hs00243196_m1 - RYK	Thermo Fisher	Cat# 4331182
Assay Id Hs00158980_m1 - EPCAM	Thermo Fisher	Cat# 4331182
Assay Id Hs04260396_g1 - MKI67	Thermo Fisher	Cat# 4331182
Assay Id Hs01078080_m1 - CDX2	Thermo Fisher	Cat# 4331182
Biological Samples		
Human adult ileal resections	Oxford University Hospital NHS Foundation Trust	REC reference: 16/YH/0247 GI illness - Gastrointestinal Illness in Oxford: prospective cohort for outcomes, treatment, predictors and biobanking; REC reference: 09/H1204/30 IBD biobank - Inflammatory Bowel Disease: prospective cohort for outcomes, treatment, predictors and biobanking; REC reference: 18/WM/0237 TIP Title: Targeting Immune Pathways

FFPE Tissue Samples (OCHRe)	Oxford Research Biobank, Oxford Centre for Histopathological Research	REC reference:19/SC/0173
FFPE Tissue Samples (Bayreuth)	Friedrich Alexander University, Bayreuth, Germany	Ethics number: 23-131 bp Title: An integrative single-cell and spatial transcriptomic analysis of fistulating Crohn's disease
Chemicals, Peptides, and Recombinant Proteins		
10% ammonium persulfate	Millipore-Sigma	Cat# 09913 (100g)
1M Tris pH 7.0	Invitrogen	Cat# AM9850G
Advanced DMEM/F-12-10 x 500 mL	Life Technologies	Cat# 12634028
Cell Staining Buffer	BioLegend	Cat# 420201
CryoStor® CS10 cell cryopreservation media	Sigma-Aldrich	Cat#C2874 (100mL)
Dako Bluing Buffer (used for ST)	Dako	Cat#CS702
Deoxyribonuclease (DNase) I from bovine pancreas	Sigma	Cat#D4263
Deparaffinization Solution	Qiagen	Ca# 939018 (50mL)
Dulbecco's Modified Eagle's Medium (DMEM)	Gibco	Cat#41965-039 (500mL)
Dulbecco's Phosphate Buffered Saline (PBS) ([-] CaCl2, [-] MgCl2)	Gibco	Cat#14190-094 (500mL)
EnVision FLEX Antibody Diluent	Dako	Cat# K800621-2
EnVision FLEX DAB+ Chromogen	Dako	Cat# K8000 (K8002, K8023)
EnVision FLEX Hematoxylin QS (used for IHC)	Dako	Cat# K800821-2
EnVision FLEX HRP	Dako	Cat# K8000 (K8002, K8023)
EnVision FLEX High pH	Dako	Cat# K800021-2
EnVision FLEX Low pH	Dako	Cat# K800521-2
Eosin Y Solution (Modified Alcoholic) (used for ST)	Abcam	Cat# ab246824 (500mL)
Eosin Y solution (used for general H&E staining)	Sigma	Cat# HT110216 (500mL)
Fetal Calf Serum / Fetal Bovine Serum (FCS)	Sigma	Cat#F9665 (500mL)
Formalin buffered 10%	VWR International Ltd	Cat#11699404
Gentle Cell Dissociation Reagent	Stemcell Technologies UK	Cat# 100-0485
GlutaMAX Supplement-100 mL	Life Technologies	Cat# 35050-038 (100mL)
Hank's Balanced Salt Solution (HBSS) medium	Lonza	Cat#10-543F (500mL)
Hemotoxylin Solution (used for standard staining)	Sigma	Cat# HHS16 (500mL)
HEPES Buffer Solution (1M)	GIBCO	Cat#15630-056 (100mL)
Histo-Clear II	Scientific Laboratory Supplies	Cat# NAT1334
Hydrogen peroxide solution (30%)	Sigma	Cat# H1009 (100mL)
Isopentane (2-Methylbutane)	Sigma	Cat#277258 (1L)
Lamina Propria dissociation kit, mouse	Miltenyi Biotech	Cat# 130-097-410
LeviSelect Human CD45 Depletion Kit	Levitas Bio	Cat#1004001
Mayer's Hematoxylin (used for ST)	Sigma	Cat#51275 (500mL)
MEM NEAA (100X)	Gibco	Cat#11140-035 (100mL)
Micromount (Mounting Medium)	Leica	Cat# 3801731 (473mL)

N,N,N′,N′-tetramethylethylenediamine (TEMED)	Millipore-Sigma	Cat# 411019 (100mL)
OCT Embedding matrix for frozen sections	CellPath	Cat#KMA-0100-00A
Penicillin-Streptomycin	Sigma	Cat#P0781 (100mL)
PierceTM Dithiothreitol (DTT)	Thermo Fisher	Cat# A39255
Potassium hydroxide solution	Merck	Cat# P4494 (50mL)
Proteinase K, Molecular Biology Grade	NEB	Cat#P8107S
RNase Inhibitor, Murine	NEB	Cat# M0314L
RNaseZap, RNase Decontamination Solution	Thermo Fisher	Cat# AM9782
Sodium Pyruvate	Sigma-Aldrich	Cat# S8636 (100mL)
TE Buffer pH 8.0	Invitrogen	Cat# AM9849 (500mL)
Trypan Blue stain 0.4%	Invitrogen, Thermo Fisher	Cat# T10282
TrypLE Express Enzyme (1X), phenol red	Gibco	Cat#12605-028 (500mL)
Ultrapure 0.5M EDTA, ph8.0	invitrogen	Cat#15575-038 (100mL)
Xylenes	Sigma Aldrich	Cat# 41736 (1L)
Y-27632 (Dihydrochloride)	Stemcell Technologies UK	Cat# 72302
Critical Commercial Assays		
Agilent High Sensitivity DNA Kit (used with Agilent Bioanalyzer 2100 system)	Agilent Technologies	Cat#5067-4626
Agilent RNA 6000 Pico Kit	Agilent Technologies	Cat# 5067-1513
Chromium Next GEM Chip G Single Cell Kit	10x Genomics	Cat# 1000120
Chromium Next GEM Single Cell 5' Library and Gel Bead Kit v1.1	10x Genomics	Cat# 1000165
Chromium Single Cell 5' Feature Barcode Library Kit	10x Genomics	Cat# 1000080
Chromium Single Cell 5' Library Construction Kit	10x Genomics	Cat# 1000020
Chromium Single Cell V(D)J Enrichment Kit, Human T Cell	10x Genomics	Cat# 1000005
Control Line Fluid Kit— 192.24	Standard Biotools	Cat# 100-4058
CRISPRclean Single Cell RNA Boost Kit - v1.0	Jumpcode Genomics	Cat# KIT1018-1
Dual Index Kit TN Set A	10x Genomics	Cat# 1000250
Dual Index Kit TS Set A	10x Genomics	Cat# 1000251
Dual Index Kit TT Set A	10x Genomics	Cat# 1000215
Flex Six™ Gene Expression IFC	Standard Biotools	Cat# 100-6308
Flex Six™ Gene Expression Reagent Kit— 5 IFCs	Standard Biotools	Cat# 100-6310
LeviSelect Human CD45 Depletion Kit	Levitas Bio	Cat# 1004001
Levitation Reagent	LevitasBio	Cat# 1003001
MERSCOPE 500 Gene Imaging Kit	Vizgen	Cat# 10400006
MERSCOPE 500 Gene Panel	Vizgen	Cat# 10400003
MERSCOPE Cell Boundary Stain Kit	Vizgen	Cat# 10400118
MERSCOPE FFPE Sample Prep Kit, 10 samples	Vizgen	Cat# 10400114
MERSCOPE FFPE Slide Box, 10 samples	Vizgen	Cat# 10500102
MERSCOPE Sample Verification Kit	Vizgen	Cat# 10400007
NextSeq 500/550 High Output kit v2.5 (150 Cycles)	Illumina	Cat# 20024907
Preamp and Reverse Transcription Master Mix—1 Tube	Standard Biotools	Cat# 100-6300
QuBit dsDNA HS Assay Kit, 100 assays	invitrogen	Cat#Q32851

RNeasy® FFPE Kit	Qiagen	Cat# 73504
RNeasy® Plus Micro kit	Qiagen	Cat# 74034
Single Index Kit N Set A	10x Genomics	Cat# 1000212
Single Index Kit T Set A	10x Genomics	Cat# 1000213
TaqMan Fast Advanced Master Mix	Life Technologies	Cat# 4444557
Visium Spatial for FFPE Gene Expression Kit, Human Transcriptome	10x Genomics	Cat# 1000336
Visium Spatial Gene Expression Slide & Reagents Kit	10x Genomics	Cat# 1000187
Visium Spatial Tissue Optimization Slide & Reagent Kit, 4 slides	10x Genomics	Cat# 1000193
Xenium Decoding Consumables (1 run, 2 slides)	10x Genomics	Cat# 1000487
Xenium Decoding Reagents (1 run, 2 slides)	10x Genomics	Cat# 1000461
Xenium Human Colon Gene Expression Panel Kit (2 rxns)	10x Genomics	Cat# 1000642
Xenium Slides & Sample Prep Reagents (2 slides, 2 rxns)	10x Genomics	Cat# 1000460

Instruments & Equipment		
Agilent 2100 Bioanalyzer Instrument	Agilent Technologies	
Agilent 4200 TapeStation	Agilent Technologies	
BD FACSAria Fusion	BD Biosciences	
BD FACSAria IIIu	BD Biosciences	
BD LSR II	BD Biosciences	
Biomark HD	Standard Biotools	
Countess II	Thermo Fisher	Cat# AMQAX1000
Leica DM18 inverted microscope	Leica	
Leica RM2135 Microtome	Leica	
LeviCell-1.0	LevitasBio	Cat# 1000001
MERSCOPE Instrument	Vizgen	Cat# 10000001
MERSCOPE Instrument Accessory Kit	Vizgen	Cat# 20100005
MERSCOPE Photobleacher	Vizgen	Cat# 10100003
Qubit 4.0 Fluorometer	Thermo Fisher	Cat# Q33240
Superfrost Plus Microscope Slides	Avantor	Cat# 631-0108
Visium Accessory Kit	10x Genomics	Cat# 1000194
Zeiss AxioScan Z1 slide scanner	Carl Zeiss Microscopy	

Software (and Algorithms)		
Baysor, version 0.6.3	GitHub	https://github.com/kharchenkolab/Baysor
BD FACS Diva Software v9.0	BD Biosciences	https://www.bdbiosciences.com/en-gb/products/software/instrument-software/bd-facsdiva-software
Cellranger version 7.1.0	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest
fastQC version 0.11.9	Babraham Institute	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Fiji v2.0.0	ImageJ	https://imagej.nih.gov/ij/
FlowJo v10.8.1	FlowJo	https://Flowjo.com
GraphPad Prism v9.5.1 (528)	Dotmatics	https://www.graphpad.com/features
Illumina bcl2fastq version 2.20.0.422	Illumina	https://Illumina.com
MERSCOPE Visualizer v2.3	Vizgen	https://portal.vizgen.com/resources/software
QuPath version 0.3.0	QuPath	https://qupath.github.io/

R package clusterProfiler version 3.16.1	R Bioconductor (198)	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
R package DropletUtils version 1.4.2	R Bioconductor (199)	https://www.bioconductor.org/packages/release/bioc/html/DropletUtils.html
R package ggplot2 version 3.3.2	R CRAN	https://cran.r-project.org/web/packages/ggplot2/index.html
R package ggpubr version 0.2.5	R CRAN	https://cran.r-project.org/web/packages/ggpubr/index.html
R package ggraph version 2.0.3	R CRAN	https://cran.r-project.org/web/packages/ggraph/index.html
R package ggrepel version 0.8.2	R CRAN	https://cran.r-project.org/web/packages/ggrepel/index.html
R package Harmony version 1.0	GitHub (192)	https://github.com/immunogenomics/harmony
R package igraph version 1.2.4.2	R CRAN	https://cran.r-project.org/web/packages/igraph/index.html
R package MAST version 1.14.0	R Bioconductor (200)	https://www.bioconductor.org/packages/release/bioc/html/MAST.html
R package miloR version 1.8.1	GitHub	https://github.com/MarioniLab/miloR
R package pheatmap version 1.0.12	R CRAN	https://cran.r-project.org/web/packages/pheatmap/index.html
R package reticulate version 1.16	R CRAN	https://cran.r-project.org/web/packages/reticulate/index.html
R package sccomp version 1.4.0	R Bioconductor	https://www.bioconductor.org/packages/release/bioc/html/sccomp.html
R package Seurat version 5.1.0	GitHub (191)	https://github.com/satijalab/seurat
R package SeuratWrappers version 0.1.0	GitHub	https://github.com/satijalab/seurat-wrappers
R Package SingleCellSignalR version 1.0	R Bioconductor (201)	http://www.bioconductor.org/packages/release/bioc/html/SingleCellSignalR.html
R package zinbwave version 1.10.1	R Bioconductor (202)	http://bioconductor.org/packages/release/bioc/html/zinbwave.html
SCENIC	GitHub (203)	https://github.com/aertslab/SCENIC
Spaceranger version 2.1.0	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest
Standard Bio Tools Real-Time PCR analysis software version 1.0.2	Standard Biotools	https://www.standardbio.com/FluidigmSite_Assets/PrdSrv_Software/Software-Packages/BiomarkX_Analysis_Software/SBI_AnalysisReal-TimePCRx86_20221216.1.zip
Visiopharm Image Analysis version 2023.09.3.15043 x64	Visiopharm	https://visiopharm.com/
Xenium Explorer v1.3.0	10x Genomics	https://www.10xgenomics.com/support/software/xenium-explorer/downloads
ZEN blue version 3.5.093.0004	Carl Zeiss Microscopy	https://www.zeiss.com/microscopy/en/home.html

Chapter 3: Optimising Full Thickness Intestinal Tissue for Experimental Applications

3.1 Introduction

In recent years, significant advancements have been achieved in single-cell and ST technologies. Theses methodologies have emerged as powerful tools for dissecting the complex pathogenesis of IBD at a molecular level.

At the outset of this project, noticeable gaps were evident in the availability of scRNA-seq data related to fistulating CD. Furthermore, there was a lack of scRNA-seq data derived from full thickness (FT) intestinal tissue obtained from adult patients with CD. Additionally, the absence of ST data in IBD further highlighted the need for comprehensive exploration in this domain.

Consequently, the absence of established protocols served as the motivation for the initial aim of this work, with a primary focus on optimising and validating protocols for preparing FT tissue for both scRNA-seq and ST workflows. Given the rarity and complexity of the fistulating sample type, thorough optimisation was essential.

The main objective of the scRNA-seq optimisations was to isolate a high number of viable cells from small quantities of complex tissues to generate high quality data for downstream analyses. Recognising the challenge posed by low-quality RNA obtained from degraded tissue samples fixed in formalin, efforts were directed towards refining ST workflows to ensure the generation of high-quality, spatially resolved data from archived tissue blocks.

Optimisation experiments were conducted *de novo* or built upon existing work generated within the laboratory (167, 168), leveraging published scRNA-seq experiments and cellular isolation protocols utilising intestinal tissue (204-206), and established methodologies for multimodal profiling (195). This chapter aims to comprehensively document the series of optimisation experiments undertaken throughout the course of the project, effectively addressing the identified gaps in the literature.

3.2 Optimisation experiments for FT intestinal tissue dissociation protocols (by compartment)

3.2.1 Lamina propria liberation

The lamina propria is densely packed with various cell types including immune cells such as T-cells, B-cells, macrophages, and dendritic cells, as well as stromal cells, fibroblast, endothelial cells amongst others. Liberating these cells from the tissue matrix allows for the isolation of individual cells for downstream scRNA-seq analysis.

Identification of an optimal dissociation protocol for full thickness intestinal tissue

The optimisation of FT tissue digestion protocols involved a comprehensive comparison of 6 distinct protocols, with variations incorporated into 3 of these protocols, resulting in a total of 10 distinct protocols under evaluation. These variations encompassed adjustments to concentrations of proprietary enzymes and deviations from protocol to the length of incubation durations. The specifics of each protocol are outlined in **Table 3.1**.

To ensure consistency and reliability in the assessment, a standardised healthy FT tissue section was utilised for each protocol, facilitating direct comparison across the various digestion methods. Assessment outcomes focused on two key parameters: the number of cells obtained at the end of the digestion process, measured both as total cell count and cell count per mg of tissue digested, and the cell viability, assessed using the Countess Automated Cell Counter system. A cell viability threshold of greater than 85% was deemed desirable for subsequent analyses.

A comprehensive summary of the results obtained from the series of digestions is depicted in **Figure 3.1**, highlighting the performance of each protocol in terms of cell yield and viability. From this analysis, four superior protocols emerged: Lamina Propria (Modified), Lamina Propria (CM Modified – 2 hours), Liberase TM (Prot. A) and Collagenase D. These protocols were selected for further comparison using flow cytometry.

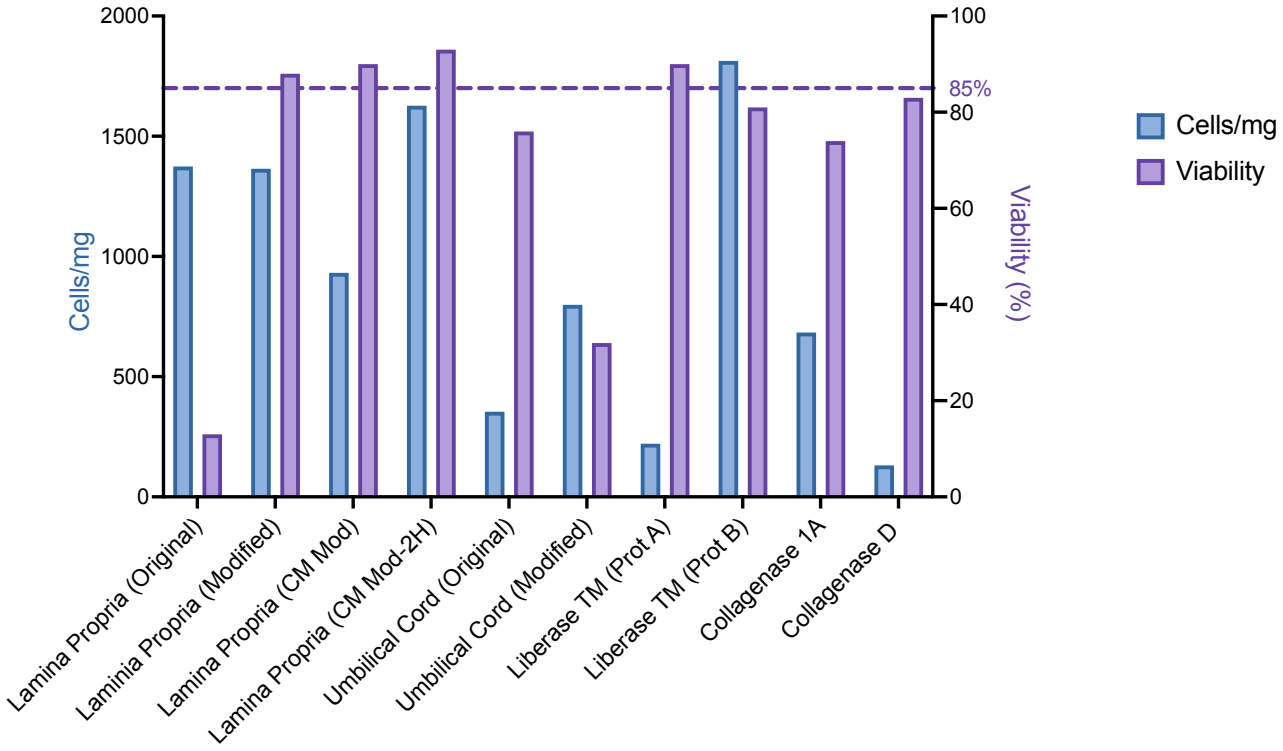


Figure 3.1: Comparison of Digestion Methods by Cell Count and Viability (*previous page*). Comparative analysis of efficiency of digestion protocols as measured by cell yield (cell count/mg) and cell viability. Source of control sample: TIP 471 Healthy, full thickness ileum. [CM: C McGregor, Mod: Modified, Original: Original protocol as per manufacturer's instructions**,** Prot: Protocol]

Table 3.1: Summary of digestion protocols applied to full thickness intestinal tissue.

Protocol	Digestion type / method	Modifications	Protocol source
Lamina Propria (Original)	Enzymatic + mechanical	Nil	Miltenyi Biotec
Lamina Propria (Modified)	Enzymatic + mechanical	Enzyme concentration increase	Miltenyi Biotec
Lamina Propria (CM Mod)	Enzymatic + mechanical	Enzyme concentration increase	Miltenyi Biotec
Lamina Propria (CM Mod-2H)	Enzymatic + mechanical	Increasing enzyme concentration and increasing digestion incubation time to 2h	Miltenyi Biotec
Umbilical Cord (Original)	Enzymatic + mechanical	Nil	Miltenyi Biotec
Umbilical Cord (Modified)	Enzymatic + mechanical	Reducing enzyme concentrations and decreasing the digestion incubation time to 1h	Miltenyi Biotec
Liberase TM (Prot A)	Enzymatic	Nil	(204)
Liberase TM (Prot B)	Enzymatic	Nil	(206)
Collagenase 1A	Enzymatic	Nil	(206)
Collagenase D	Enzymatic	Nil	(204)

Given the intended application of the selected protocol for flow cytometry-based single cell isolation into immune and stromal populations, the flow cytometry panel utilised included specific markers for immune and stromal cell populations (CD3-FITC (OKT3), CD45-FITC (5B1), CD19 (H1B19), CD90-PE (5E10), EpCAM-BV785 (9C4), Live/Dead – Ki67 (**Key Resources Table 2.16**). Initial flow cytometry analysis comparing the Lamina Propria digests revealed a notable absence of the $CD45^+$ population (**Fig. 3.2a**), suggesting potential cleavage of the epitope by the lamina propria digestion enzyme cocktail. Subsequent adjustments, including changing the CD45 antibody clone (clone 5B1 to H130), successfully resolved this issue, enabling accurate identification of both EpCAM and CD45 populations across all four protocols (**Fig. 3.2b**).

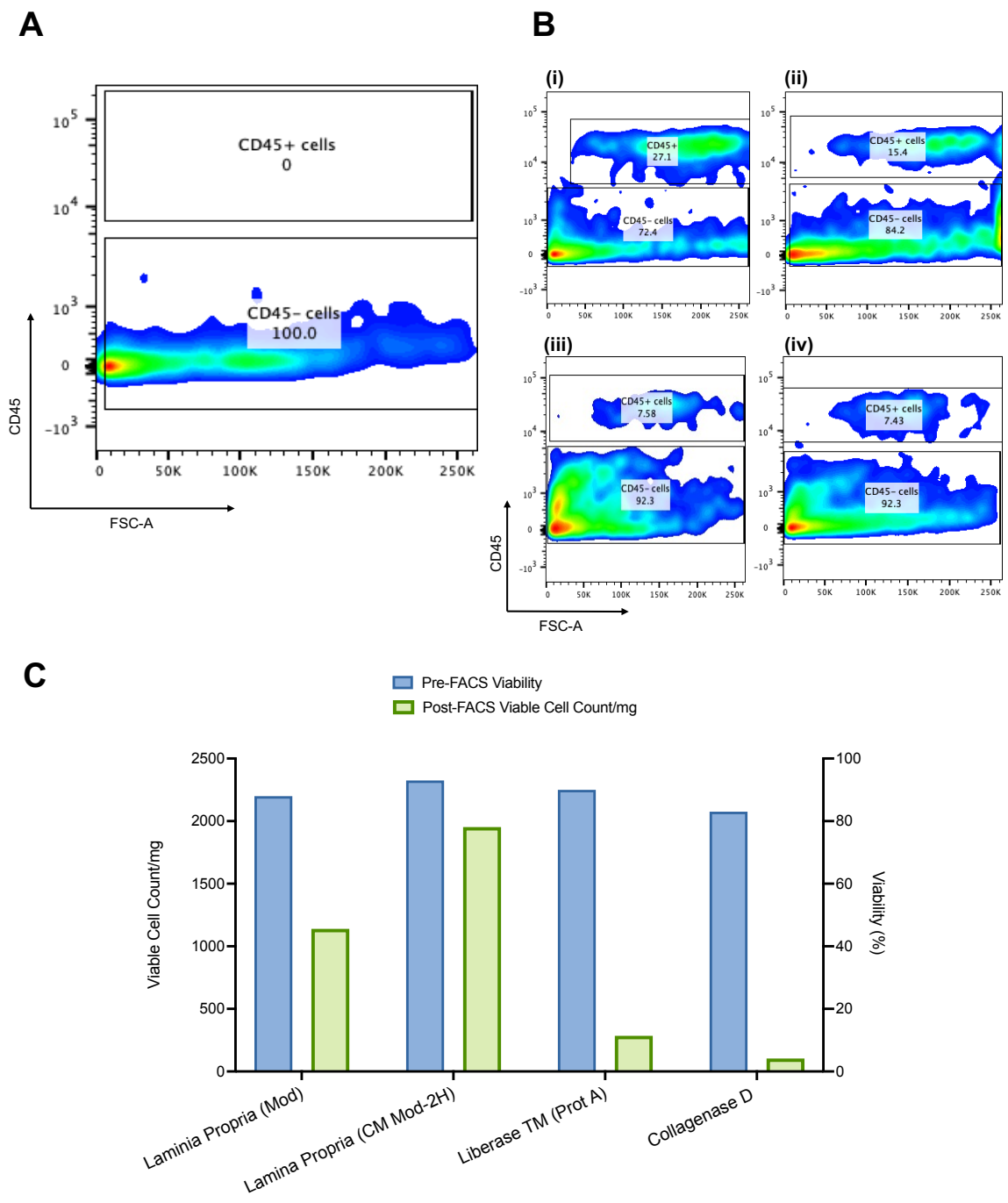


Figure 3.2: Flow-Cytometry Analysis of Antibody Optimisation and Identification of Optimum FT Digestion Protocol. Flow cytometry analysis was conducted to optimise antibody selection and identify the optimal FT tissue digestion protocol (A) The absence of the immune ($CD45^+$) population is shown using the CD45-FITC antibody clone 5B1. (B) The immune ($CD45^+$) population was successfully identified using an alternative CD45-FITC clone (H130), which was compatible with each digestion protocol tested (i) LP (CM-Mod 2H) (ii) LP (Mod) (iii) Collagenase D (iv) Liberase TM (Protocol A). (C) The percentage viability of digested product was assessed pre-FACS (measured by Countess) and post-FACS, with the viable cell count adjusted for the starting tissue mass. Single cells were defined based on FSC-A and SSC, with total viability measured using Zombie-Violet staining.

Following flow cytometry analysis, the Lamina Propria (CM Modified – 2 hours) protocol emerged as the superior choice based on cell count yield and viability post-fluorescence-activated cell sorting (FACS) (**Fig. 3.2c**). To further validate the applicability of this protocol in diseased tissue settings, FT inflamed Crohn's ileal tissue was subjected to the Lamina Propria (CM Modified – 2 hours) protocol, alongside healthy control tissue. The resulting analysis, utilising a revised FACS panel (CD45-FITC (H130), EpCAM-BV421 (9C4) and DAPI-BV421), revealed acceptable cell viability but lower than desirable cell numbers compared to healthy tissue. Notably, the stromal fraction ($CD45^-/EpCAM^-$) exhibited the lowest cell yield compared to the immune ($CD45^+/EpCAM^-$) population (**Fig. 3.3**).

In conclusion, whilst the Lamina Propria (CM Modified – 2 hours) protocol demonstrated promising performance in healthy tissue settings, further optimisation was warranted to ensure optimal cell yield and viability in diseased tissue contexts. Specifically, increasing the amount of starting material was required to enhance the yield of viable cells recovered post-digest and sort, particularly in stromal cell populations.

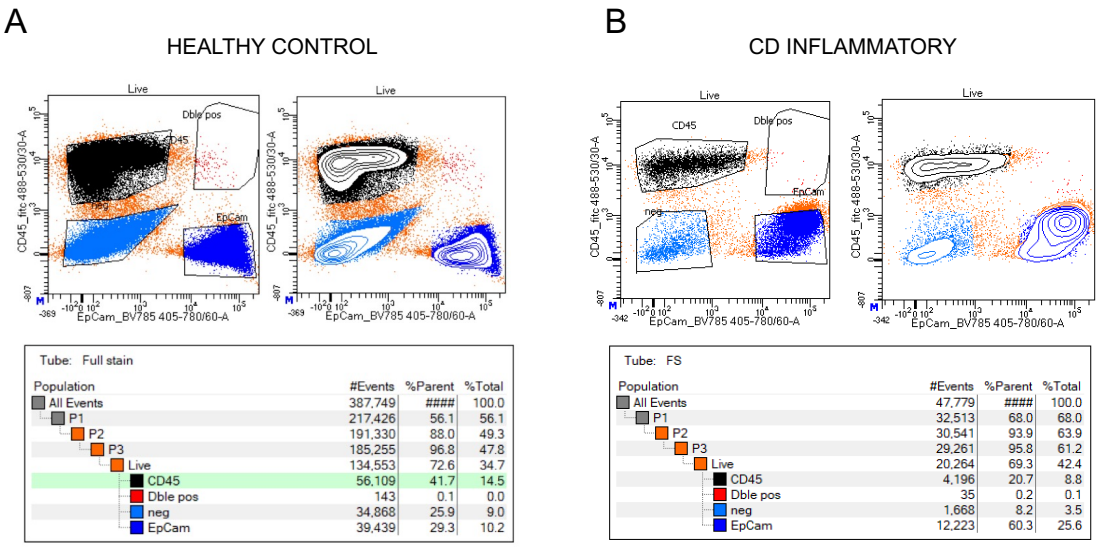


Figure 3.3: FACS Analysis of Digested Full Thickness Ileal Tissue. FACS analysis was performed on digested FT ileal tissue samples obtained from both healthy and inflammatory CD subjects. Summary illustrating gating strategy for known epithelial, immune markers, and stromal ($CD45^-/EpCAM^-$) cells. Representative data from a healthy sample (A) and from an inflammatory CD sample (B) are presented. Notably, the healthy sample yielded more than 6-fold increase in the number of viable cells compared to the CD sample.

3.2.2 Epithelial crypt isolation

The epithelial crypt chelation protocol employed in this project was adapted from the methodology described by Fawkner-Corbett et al. (189) with a modification of a total of two chelation steps of 20 minutes in a waterbath at 37°C (**Methods 2.4**). This protocol was applied to both mucosa separated from FT tissue and directly to FT tissue, where separation of mucosa by blunt dissection was unfeasible.

The efficacy of this epithelial crypt chelation protocol was evaluated through a comparative analysis utilising FACS techniques. Specifically, the protocol's performance in isolating epithelial crypts was assessed by comparing its application to FT healthy tissue versus healthy biopsy tissue. The results of this comparative analysis revealed the protocol's effectiveness in isolating epithelial crypts from FT tissue.

3.3 Optimisation of antibody staining conditions with preferred dissociation protocol

In my investigation comparing digestion protocols by FACS, I observed a consistent decrease in cell viability post sorting, as illustrated in **Figure 3.4a**. Notably, viability levels exhibited a decline throughout the staining process and during FACS, as depicted in **Figure 3.4b(i)**. Moreover, I noted a significant loss of cells during the staining stage, with a reduction factor of 2.7 observed (**Fig. 3.4b(ii)**).

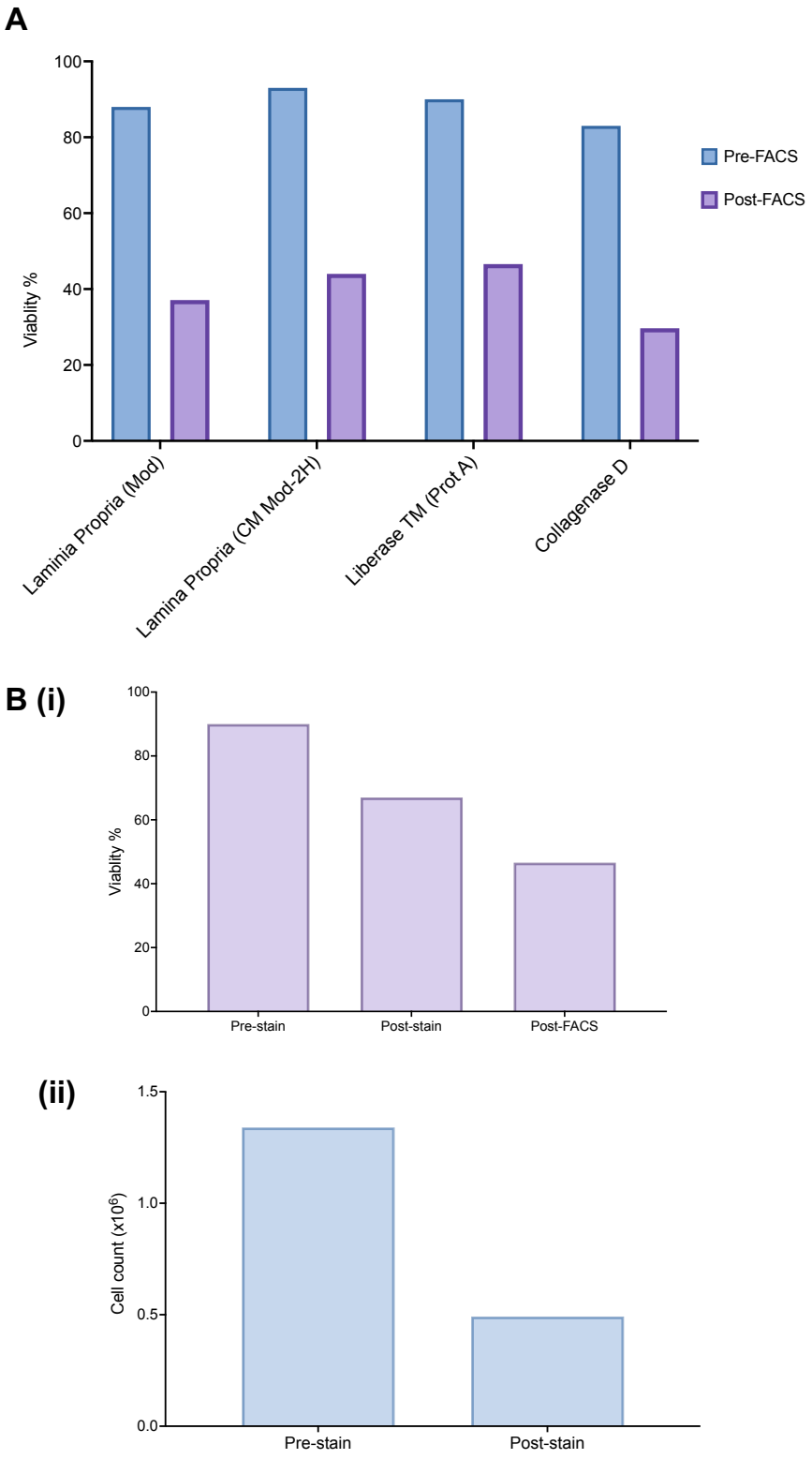


Figure 3.4: Viability and Cell Count of Digested Full Thickness Tissue. (A) Mean percentage viability of sorted, digested FT intestinal tissue before and after FACS across the representative protocols (n=2 per condition), measured using Zombie-Violet staining. (B) Reduction in mean cell viability (i) and total cell count (ii) during the antibody staining process, as measured by Countess (n=4).

Recognising the drop in viability during staining, I hypothesised that the mechanism of action of the Zombie Violet (Live-Dead) staining reagent, requiring a longer duration (10 minutes at room temperature), contributed to cell death. To address this concern, I explored an alternative cell viability stain in the form of DAPI, given its comparatively shorter staining time.

Furthermore, I postulated that staining cells in supplemented DMEM (**Methods 2.4**) might enhance viability compared to the proprietary staining buffer devoid of cell nutrients. To test this hypothesis, I repeated the digestion and staining process under both conditions, employing DAPI cell viability stain. Evaluation using the Countess system demonstrated a marginal superiority of DMEM over staining buffer in terms of viability, as shown in **Figure 3.5a**.

Additionally, I conducted an assessment to determine whether the use of medium as a buffer could impact antibody binding, an aspect that had not been previously explored in our lab. The findings revealed that DMEM did not affect antibody binding, as illustrated in **Figure 5b**.

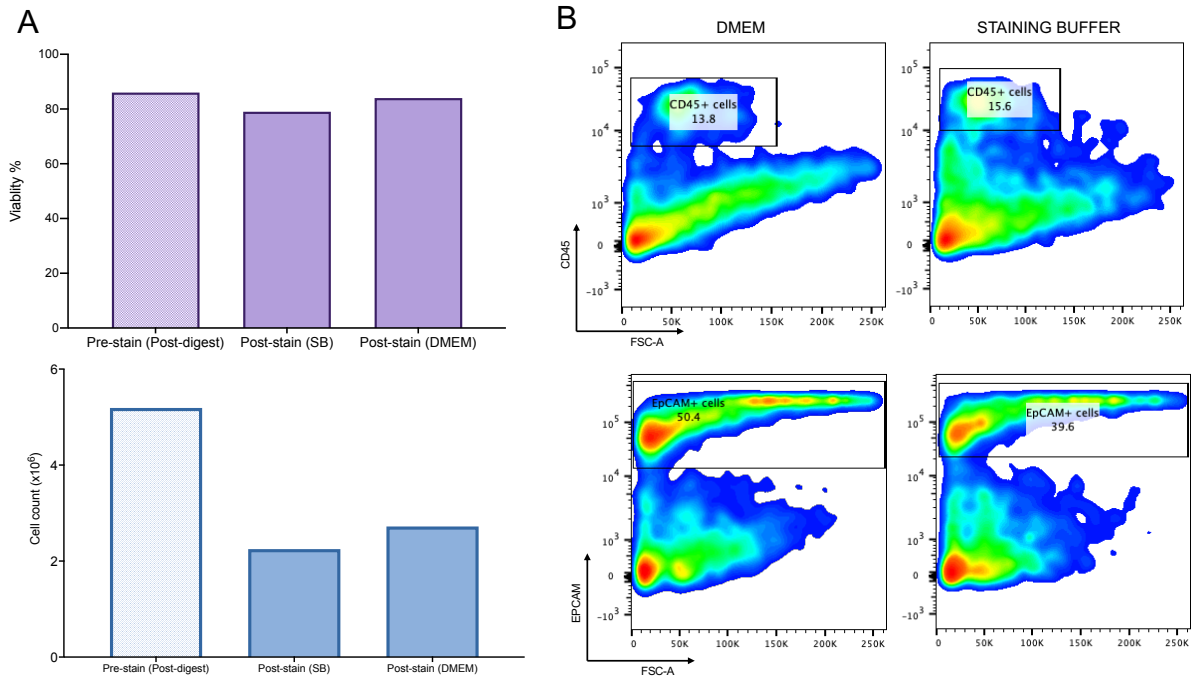


Figure 3.5: Comparison of Staining Buffer and Supplemented Media (*previous page*). (A) Analysis comparing cell viability and cell count of single cells from the same sample, re-suspended for antibody staining in staining buffer versus supplemented DMEM, measured by countess. (B) Flow cytometry analysis of single cells, illustrating the successful identification of immune and epithelial populations with both staining buffer and supplemented DMEM buffered cells.

In parallel, I assessed the compatibility of the Lamina Propria (CM 2-hour) digestion protocol with hashing antibodies (β2M, CD298 epitopes), integral for multiplexing samples in scRNA-seq experiments. Encouragingly, the Lamina Propria (CM 2-hour) digestion protocol demonstrated compatibility with hashing antibodies, as evidenced by successful isolation of β2M and CD298 positive populations via FACS.

In conclusion, based on these findings, I decided to proceed with the use of medium as the buffer for cell staining in flow cytometry and hashing antibody experiments, with DAPI live dead stain adopted as the standard.

3.4 Optimisations for complex FT intestinal tissue in scRNA-seq workflow

3.4.1 Enhancing scRNA-seq recovery rates and data quality for complex tissue samples utilising cell levitation technologies

LeviCell optimisation

Prior to the optimisation efforts detailed below, the stromal fraction exhibited notably poorer viabilities (pre-loading on 10x Chromium) and yielded fewer cells than expected upon sequencing, particularly evident across diseased (involved) samples. Over the course of five runs (five reactions) of scRNA-seq, a total of 10,771 cells were recovered from the stromal fraction, in stark contrast to the over 27,959 cells retrieved from the immune fraction.

To address these challenges, LeviCell, a cell levitation technology, emerged as a promising alternative for more gentle stromal cell isolation. The workflow of LeviCell isolation is detailed elsewhere (**Methods 2.7**). Initially, a pilot experiment [ZC, ASG, PSZ] was conducted to assess LeviCell's efficacy, sequencing six reactions of cells - two from LeviCell isolation and four from non-LeviCell isolation i.e., FACS sorting, allowing for direct comparison. Firstly, LeviCell isolation demonstrated compatibility with antibody hashing, enabling multi-modal profiling such as CITE-Seq or sample multiplexing (**Fig. 3.6a**).

Technically, LeviCell's $CD45^+$ depletion workflow effectively removed unwanted immune cells, as confirmed by the absence of detection of the immune cell marker PTPRC within the clusters (**Fig. 3.6b**). Encouraged by these preliminary findings, subsequent sequencing of the same FT samples, previously subjected to scRNA-seq by FACS isolation, was performed using LeviCell isolation, facilitating direct comparison of the platforms.

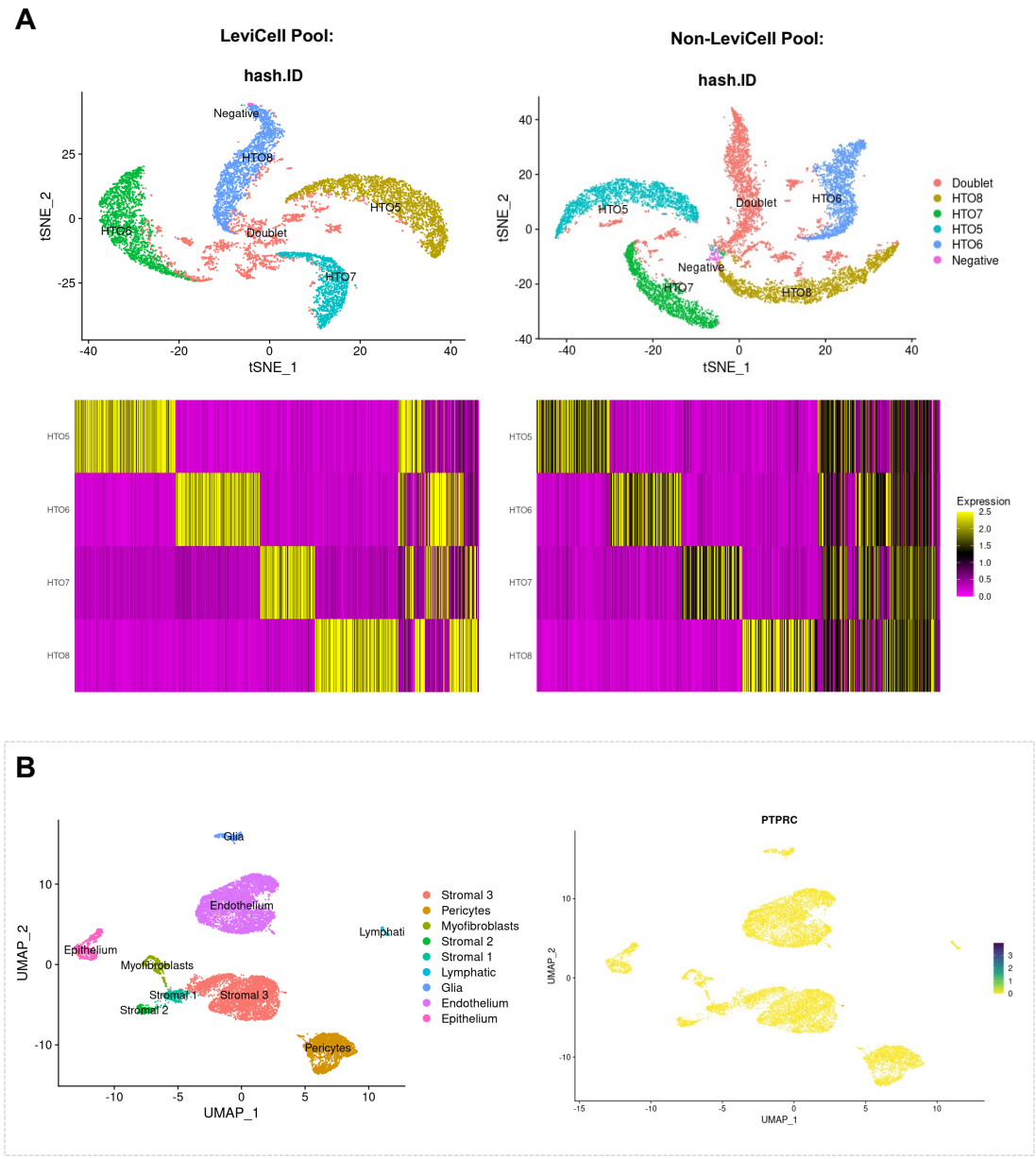


Figure 3.6: LeviCell Isolation Compatibility with Antibody Hashing and Effective CD45⁺ Depletion. (A) tSNE embedding of antibody hashing (HTO) dataset along with corresponding heatmap of scaled (z-scores) normalised HTO values. Four discrete singlet clusters (and doublet clusters) are clearly present with similar profiles in both LeviCell and non-LeviCell pools. (B) UMAP embedding visualising scRNA-seq clusters of cell populations obtained from CD45⁺depleted LeviCell pool, demonstrating successful depletion of immune cells, as measured by the expression of immune cell marker PTPRC.

In total, six reactions of cells were sequenced across three runs using LeviCell isolation. In the first run, approximately 1.2 x 10^6 cells (/ml) were loaded onto the LeviCell, yielding 2.5 x 10^5 viable, CD45- cells (/ml). This yield was consistent across subsequent runs with an average post LeviCell isolation yield of 37.4% (**Table 3.2**), exceeding the 10x Chromium loading requirement of 30,000

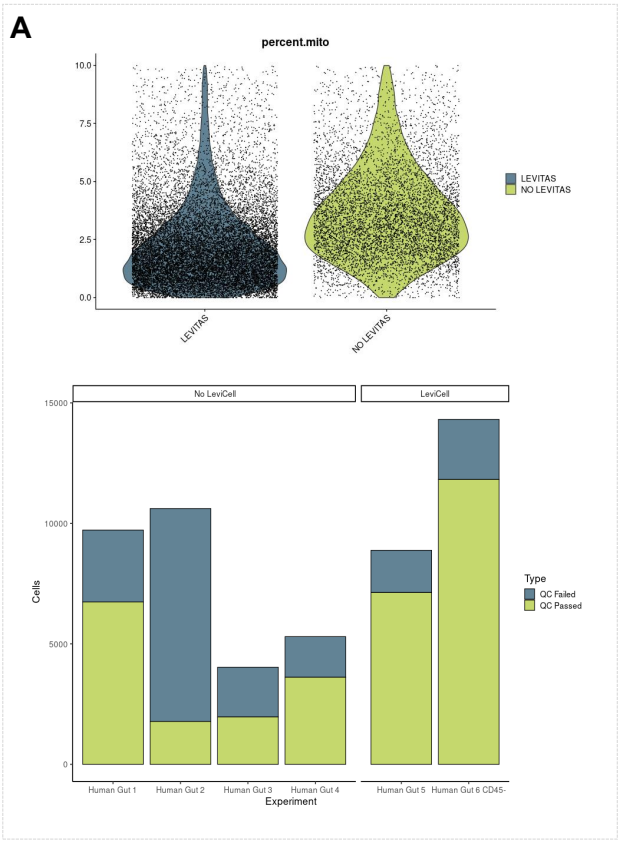
cells. Notably, the live cell enrichment effectively maintained cell viability, with an average post LeviCell and pre-Chromium loading viability of 98.0%, contrasting with 72.4% viability observed with FACS isolation (measured by Countess).

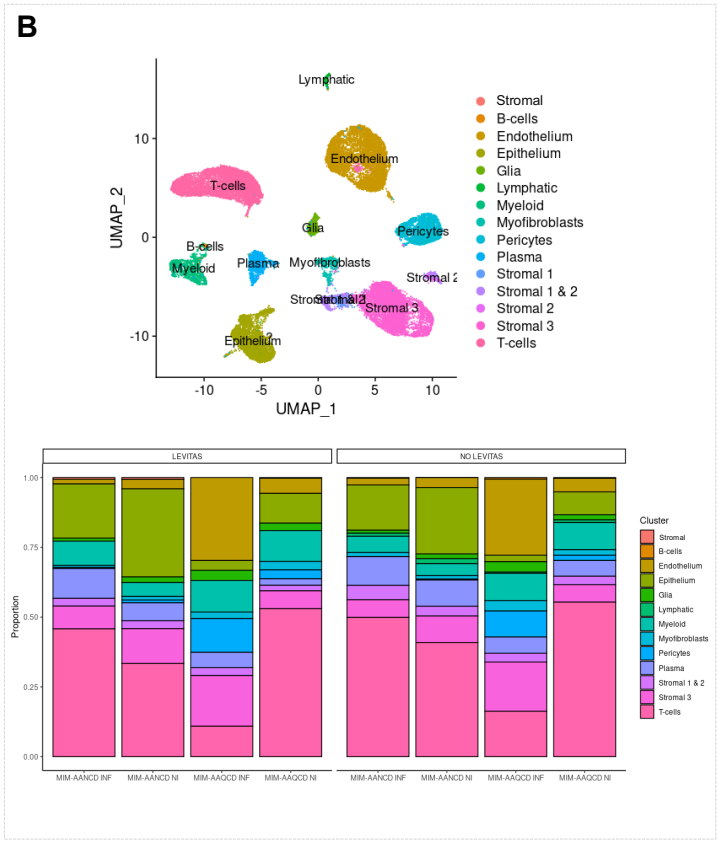
Table 3.2: Overview of viable cell count pre- and post-LeviCell as measured by Countess.

LeviCell Run	Pre-LeviCell		Post-LeviCell		
	No. of cells loaded	**Pooled viability (%)**	**No. of viable cells yielded**	**Percentage yield (%)**	**Pooled viability (%)**
1	332, 200	84.8	161, 777	48.7	95.5
2	250, 000	79.1	119, 455	47.8	98.5
3	350, 000	84.0	54, 917	15.7	99.9
Mean	**310,733**	**82.6**	**112,050**	**37.4**	**98.0**

Furthermore, LeviCell isolation resulted in lower mitochondrial content and fewer cells with low RNA complexity, indicative of better overall cell quality compared to FACS isolation upon analysing sequencing data (**Fig. 3.7a**). Similar to FACS isolation, LeviCell isolation effectively preserved all stromal cell types, including smaller populations such as glial cells and lymphatic cells, while maintaining the balance of cell types within samples (**Fig. 3.7b**). To further ensure balanced representation, I used existing FACS-sorted data from the same samples previously run to identify those with similar immune and stromal proportions for pooling prior to LeviCell isolation.

Encouragingly, LeviCell substantially improved the cell yield from poor-quality samples, demonstrating a five-fold increase in cell recoveries compared to the FACS isolation method (**Fig. 3.7c**). This enhancement enabled better detection of disease-specific populations at the single-cell level (**Chapter 4**). Although some differences were observed in "contaminating" cell populations, with LeviCell isolation capturing more epithelial cells, the FACS method indicated a higher presence of contaminating immune cells. Overall, the superiority of LeviCell isolation was evident, indicating a promising shift in the ability to capture data from poorer quality, more complex samples.


A
percent.mito
LEVITAS
NO LEVITAS
No LeviCell
LeviCell
Cells
Type
QC Failed
QC Passed
Human Gut 1
Human Gut 2
Human Gut 3
Human Gut 4
Human Gut 5
Human Gut 6 CD45-
Experiment


B
Lymphatic
Endothelium
T-cells
Glia
Pericytes
B-cells
Myeloid
Plasma
Myofibroblasts
Stromal 2
Stromal 3
Epithelium
UMAP_1
UMAP_2
Stromal
B-cells
Endothelium
Epithelium
Glia
Lymphatic
Myeloid
Myofibroblasts
Pericytes
Plasma
Stromal 1
Stromal 1 & 2
Stromal 2
Stromal 3
T-cells
LEVITAS
NO LEVITAS
Proportion
Cluster

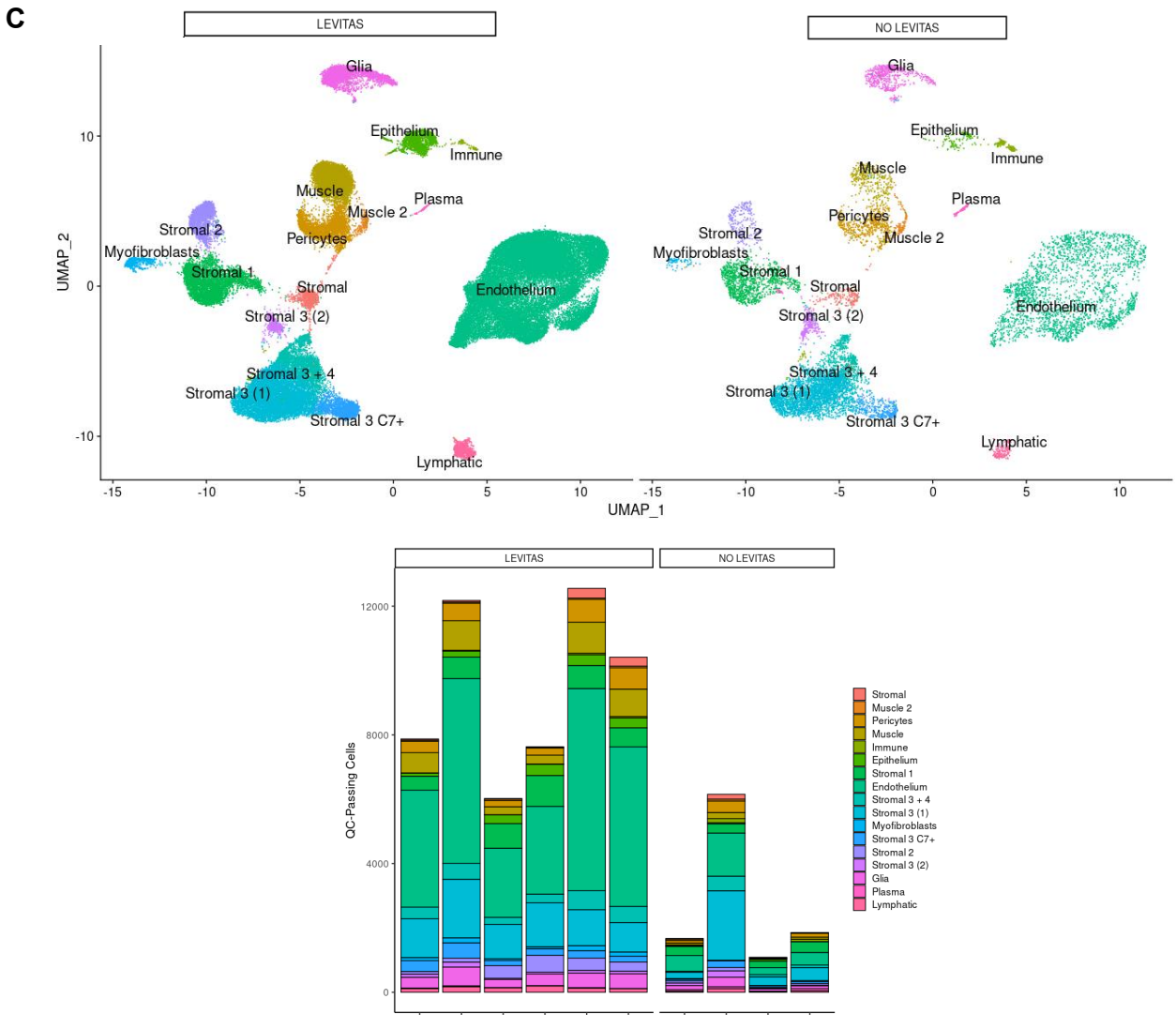

C
LEVITAS
NO LEVITAS
Glia
Epithelium
Immune
Muscle
Muscle 2
Plasma
Pericytes
Stromal 2
Myofibroblasts
Stromal 1
Stromal
Stromal 3 (2)
Endothelium
Stromal 3 + 4
Stromal 3 (1)
Stromal 3 C7+
Lymphatic
UMAP_1
UMAP_2
QC-Passing Cells
Stromal
Muscle 2
Pericytes
Muscle
Immune
Epithelium
Stromal 1
Endothelium
Stromal 3 + 4
Stromal 3 (1)
Myofibroblasts
Stromal 3 C7+
Stromal 2
Stromal 3 (2)
Glia
Plasma
Lymphatic

Figure 3.7: LeviCell Isolation Summary (*previous page*). (A) Violin plots display the distribution of mitochondrial percentage content among cells isolated using LeviCell and non-LeviCell methods. Comparative analysis of QC-passing cells, assessed by mitochondrial content and RNA complexity, in LeviCell and non-LeviCell pools. (B) UMAP embedding visualising scRNA-seq clusters of cell populations obtained from $CD45^{+}$-depleted LeviCell pool, demonstrating preservation of cell types and cell type balance within samples. (C) UMAP embedding depict the number of cells, by cluster, in LeviCell versus non-LeviCell pools. The total number of cells is 56,672 and 10,771 respectively. Stacked bar charts depict the proportional analysis of QC-passing cells by cluster in LeviCell versus Non-LeviCell pool per sample. [LeviCell = Levitas; QC: Quality Control]

3.4.2 Enhancing scRNA-seq data quality and sequencing coverage utilising CRISPR technologies

In our efforts to enhance the quality of single-cell data and sequencing coverage further, we implemented the CRISPRclean Single Cell workflow (Jumpcode Genomics, see **Methods 2.8**), which is compatible with 10x Genomics gene expression libraries. Typically, mitochondrial RNAs and ribosomal RNA account for around 35% and 20% of reads in a library, respectively.

The CRISPRclean workflow aims to deplete highly expressed protein-coding ribosomal and mitochondrial genes, unaligned reads, and non-variable genes, thereby redistributing 50% of sequencing reads to biologically relevant transcripts. Maximising sequencing coverage is vital for accurately detecting low-abundance transcripts and enhancing the sensitivity of scRNA-seq, both of which are crucial for identifying rare cell types. Additionally, higher sequencing coverage improves the accuracy of gene expression quantification by providing a more comprehensive representation of the transcriptome of each cell, enabling finer resolution of cell states and transcriptional dynamics within heterogeneous cell populations.

For the first time within our group, we applied this method to 5-prime 10x Genomics gene expression libraries derived from FT intestinal tissue samples (including CD involved, uninvolved, and healthy samples).

Upon sequencing a CRISPR-depleted epithelial library compared to a non-depleted library, we observed a higher gene detection rate of non-target genes (**Fig. 3.8**). Furthermore, CRISPR depletion led to a 5 to 10-fold reduction in mitochondrial content (**Fig. 3.9a**), which is particularly significant for the epithelial fraction, where mitochondrial content is typically higher (median approximately 25%) (**Fig. 3.9b**). However, variability exists across cell types, with intestinal T-cells typically exhibiting lower mitochondrial content.

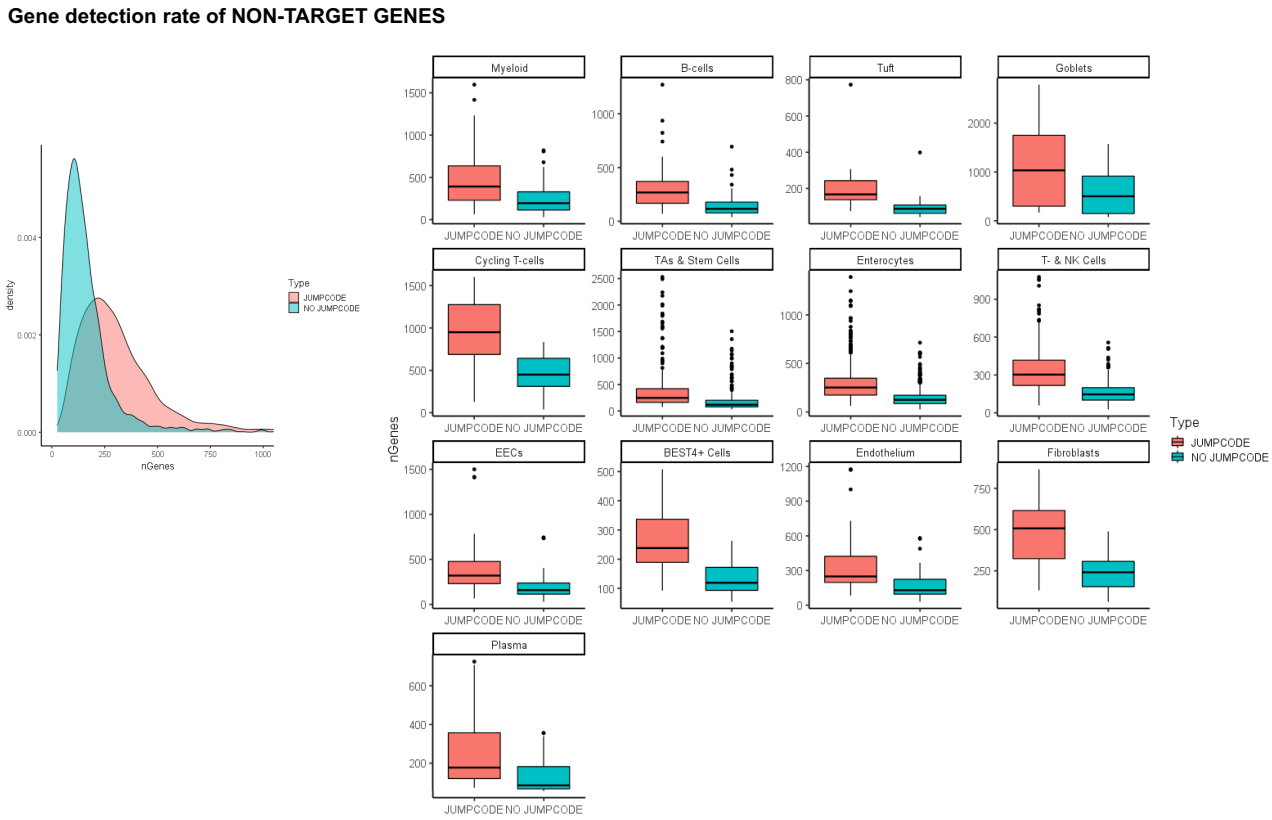


Figure 3.8: Gene Detection Rate of Non-Target Genes in Jumpcode Depleted Pools. Gene detection rate of non-target genes in Jumpcode depleted and non-Jumpcode depleted libraries by cell type (bar plots, 25th, 50th and 75th percentiles shown, colour denoting Jumpcode status).

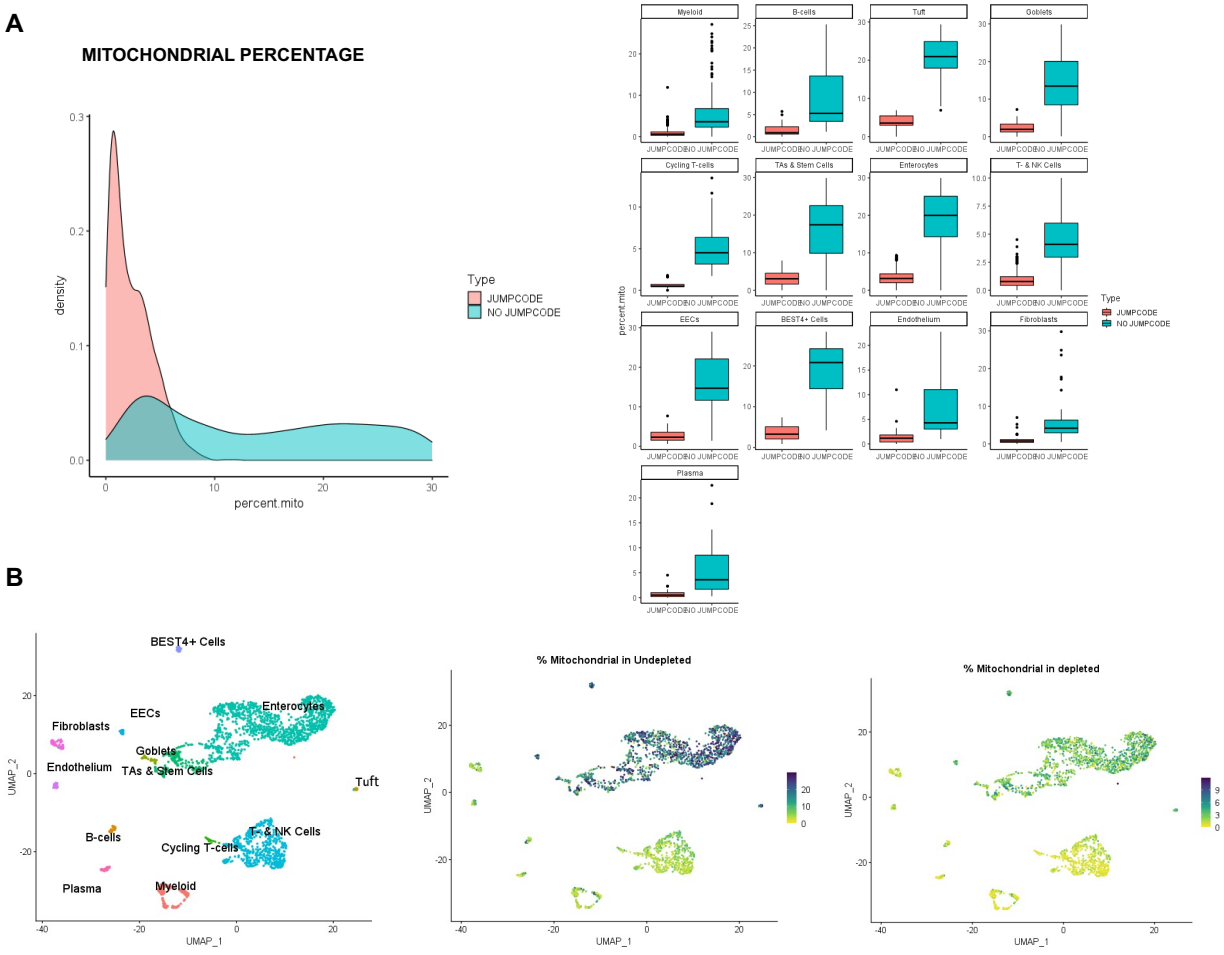


Figure 3.9: Percentage Mitochondrial Content in Jumpcode Depleted Pools. (A) Mitochondrial percentage in Jumpcode depleted and non-Jumpcode depleted pools by cell type (bar plots, 25th, 50th and 75th percentiles shown, colour denoting Jumpcode status) and by UMAP-embedding visualisation (points coloured by percentage expression within cluster) (B).

While ribosomal protein gene percentage decreased in the depleted group, the pattern of variation in decrease also exists for ribosomal protein genes, with more rRNA content observed in T-cells compared to epithelium (**Fig. 3.10a**). The impact of CRISPR depletion on non-variable genes was less evident however (**Fig. 3.10b**).

Interestingly, some of the non-variable gene targets did not show complete depletion, as we were able to identify them in both depleted and non-depleted libraries in similar proportions (**Fig. 3.10b**). However, this was advantageous, as some of these markers are essential for cell annotation in single-cell analyses (e.g., PTPRC for immune lineage, TYROBP for natural IELs and other lymphocyte subtypes, MT1E for certain epithelial populations, KRT19 as an epithelial marker, and FCER1G as a marker for myeloid and NK cells).

We achieved significantly higher sequencing saturation with CRISPR depleted libraries compared to non-depleted libraries, further demonstrating the efficacy of the CRISPRclean workflow. Given its success, we optimised our single-cell workflow to include CRISPRclean depletion in the preparation of all gene expression libraries, thereby improving the downstream quality of our data and reducing our sequencing requirements.

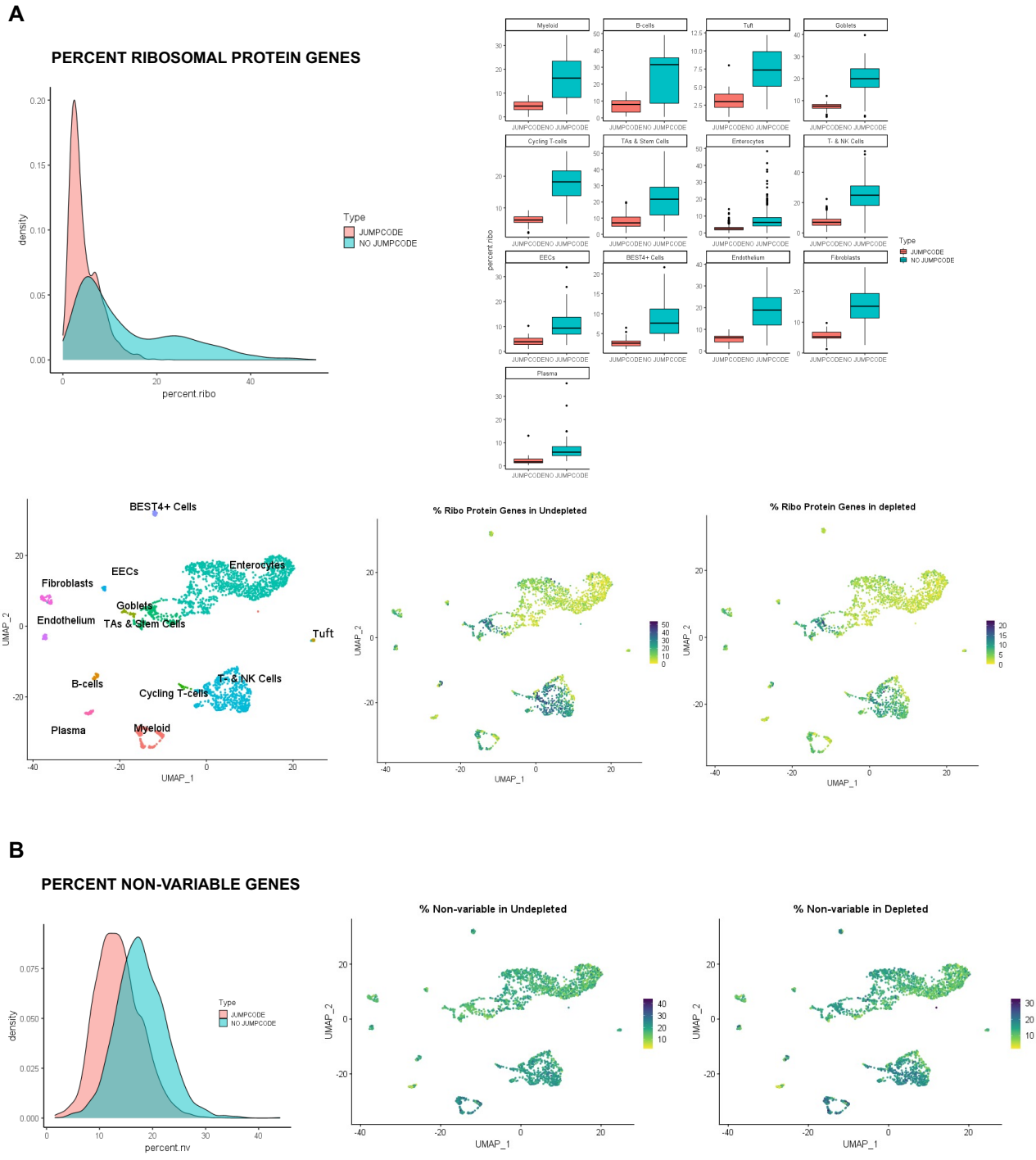


Figure 3.10: Expression of Ribosomal Protein Genes and Non-Variable Genes in Jumpcode-Depleted Pools. (A) Percentage ribosomal protein gene content in Jumpcode depleted and non-Jumpcode depleted pools by cell type (bar plots, 25th, 50th and 75th percentiles shown, colour denoting Jumpcode status) and expression by UMAP-embedding visualisation (points coloured by percentage expression within cluster). (B) Percentage non-variable gene content in Jumpcode depleted and non-Jumpcode depleted pools (colour denoting Jumpcode status) and expression by UMAP-embedding visualisation (points coloured by percentage expression within cluster).

3.5 Optimisations for spatial transcriptomics workflow

3.5.1 Fresh Frozen ST workflow

At the onset of my project, there were no published protocols in the literature for performing fresh frozen ST on FT adult intestinal tissue. To address this gap, a pilot experiment was initiated to assess RNA recovery across the representative intestinal layers observed in FT tissue and to identify any necessary optimisations for tissue permeabilisation.

The pilot experiment utilised FT ileal tissue samples from various conditions, including CD fistulating involved, CD stricturing involved, CD inflammatory involved and healthy, non-IBD samples. Before conducting the pilot experiment, several optimisation experiments were carried out to refine the methodology and ensure the reliability of the results.

3.5.1.1 Configuration of cryopreserved samples for cryosectioning

Preservation of FT sections for spatial analysis in isopentane (2-methylbutane) necessitates precise orientation and cryostat mounting to ensure accurate sectioning. To evaluate sample quality and suitability for spatial analysis, H&E staining was performed to assess the relevant tissue morphology, pathology and absence of significant artefacts.

Identification of the correct sectioning plane was crucial for capturing all representative layers of the tissue. For example, Sample GI 7051 (**Fig. 3.11a**) was oriented in both the transverse and sagittal planes, with the former capturing both the mucosa and submucosa, while the latter focused on the deeper muscle layer. In this example, although the sagittal configuration provided the correct orientation, deeper sectioning would also encompass the muscle layer in addition to the submucosa, thereby enabling a more comprehensive FT representation, emphasising the necessity for contemporaneous H&E reference.

Subsequently, the sagittal configuration emerged as the preferred option for continued sectioning (**Fig 3.11b**). Furthermore, the placement of the tissue sample in the sagittal plane at the time of embedding in OCT compound, the alternative cryopreservation approach, was crucial for optimal results, as depicted in **Figure 3.11c**.

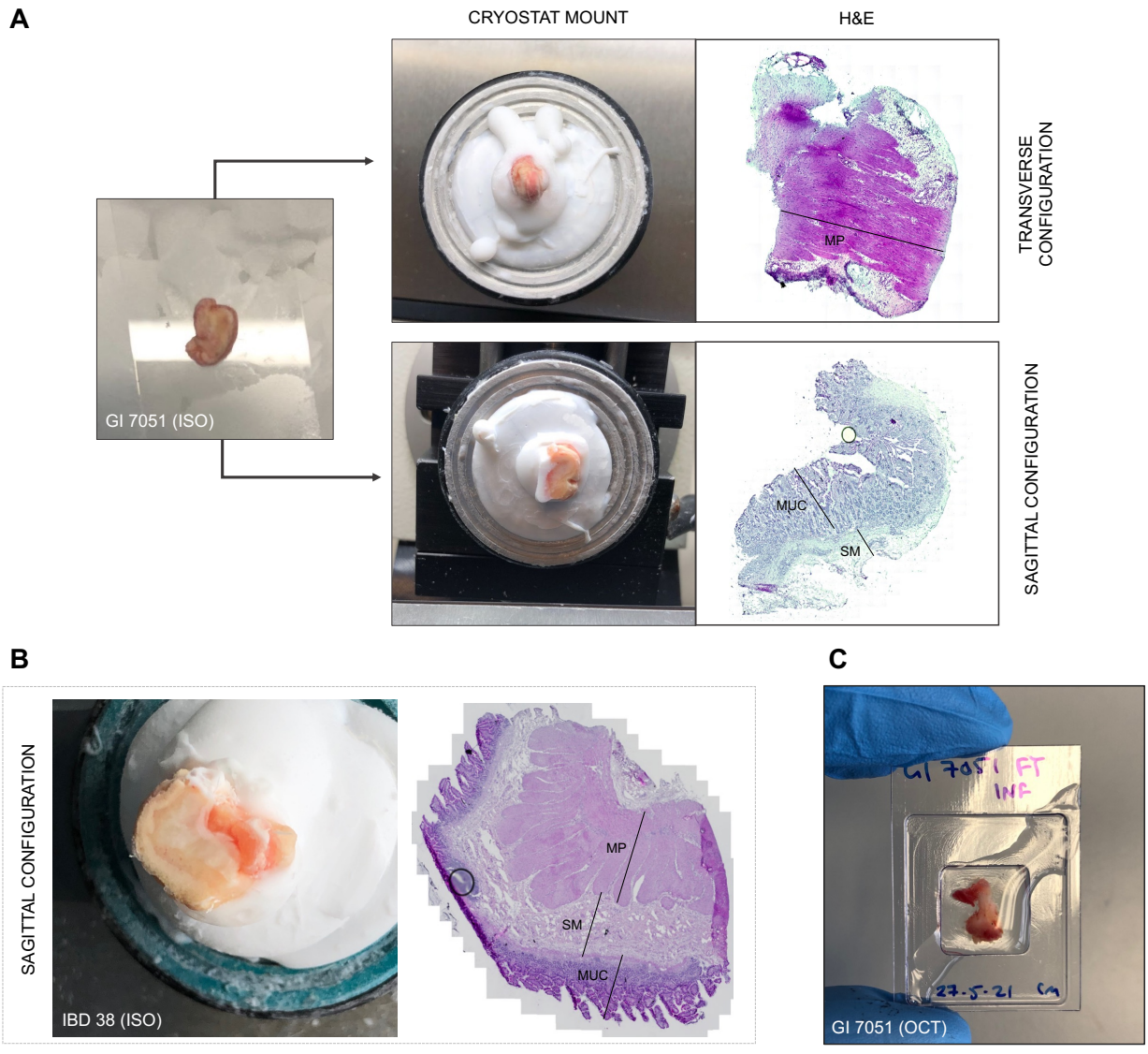


Figure 3.11: Orientation and Configuration of Fresh Frozen Samples for Cryosectioning *(previous page)*. Full thickness tissue samples frozen in 2-methylbutane (isopentane, ISO). (A) Comparison of transverse and sagittal configuration of fresh frozen tissue and corresponding H&E. Transverse configuration capturing muscularis propria only; sagittal configuration capturing representative layers (mucosa and submucosa) (B) Preferred sagittal configuration demonstrating all representative layers as confirmed in corresponding H&E. (C) Example of correct orientation and placement of OCT-embedded sample within cassette. [H&E, haematoxylin & eosin; MUC, mucosa; MP, muscularis propria; SM, submucosa]

3.5.1.2 Assessment of RNA integrity and cryopreservation methods

The RIN serves as a crucial criterion for successful FF ST analysis using the 10x Genomics Visium platform, requiring a minimum threshold of ≥ 7 for each sample. In my project, RIN assessment was conducted for all FF samples, with a comparative analysis performed across various conditions (involved vs. uninvolved), tissue types and cryopreservation methods (refer to **Methods 2.9** for details).

Results revealed variability in RNA quality based on tissue type and cryopreservation method, with submucosa (SM) samples exhibiting the lowest RINs among ISO-preserved tissues. Upon evaluating the disparity in RINs between cryopreservation types, the findings indicated mostly comparable results, albeit slightly higher RINs observed for both FT and SM tissues stored in OCT compared to those preserved in isopentane (ISO) (**Fig. 3.12**).

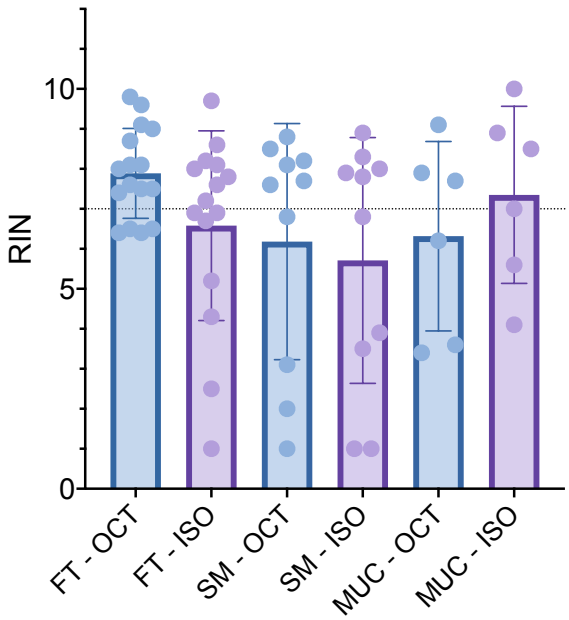


Figure 3.12: Comparative Analysis of RNA Integrity Number (RIN) by Cryopreservation Method and Tissue Type *(previous page)*. Bar chart depicting the comparative analysis of RIN across different cryopreservation methods (ISO vs OCT) and tissue types. RIN values were measured using the Pico assay on Agilent Bioanalyzer, with the acceptable threshold for RIN set at 7. Chart presents the mean values along with the standard deviation.

Ultimately, samples selected for the pilot experiment were chosen based on their representative tissue morphology as assessed by H&E staining and a RIN threshold of ≥ 7. It is noteworthy that consistent cryopreservation methods were employed for all selected samples to mitigate potential biases, with samples for the pilot experiment stored in isopentane.

3.5.1.3 Tissue permeabilisation optimisation experiments

In addressing the absence of commercially available data to inform my project, I conducted a series of experiments aimed at optimising tissue permeabilisation conditions for FT intestinal tissue, encompassing both involved and healthy ileum. For details regarding the tissue optimisation (TO) experiment, please refer to **Methods 2.9**.

The initial run (Run 1) yielded promising results, with the brightest signals observed at both 12 minutes and 36 minutes (**Fig. 3.13**). However, due to the substantial difference between these timepoints, I conducted a subsequent run with a wider range of durations to pinpoint the optimal timing.

In Run 2, I observed an increase in signal intensity with prolonged permeabilisation times. However, this was accompanied by a decrease in specificity and sensitivity, indicative of a more diffused signal, suggestive of tissue overpermeabilisation (**Fig. 3.13**). Therefore, the original 36-minute time point reflected overpermeabilisation rather than optimal signal. After thorough analysis, the 12-minute timepoint emerged as optimal.

It is noteworthy that variations in signal intensity were observed across the FT tissue, irrespective of time points, with the mucosa consistently exhibiting a brighter signal compared to the submucosa. These differences likely reflect variations in tissue quality and thickness.

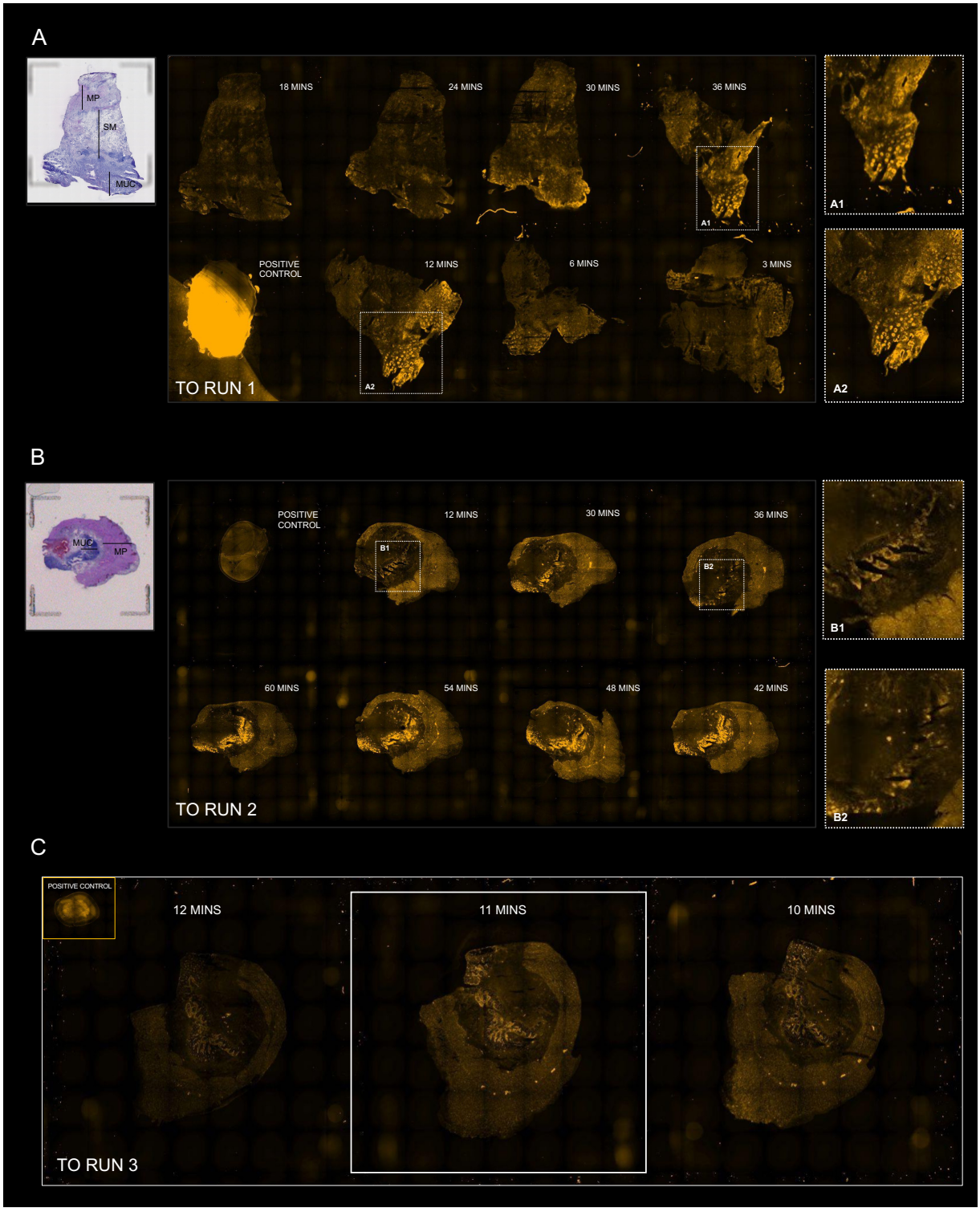


Figure 3.13: Tissue Optimisation for Fresh Frozen Spatial Transcriptomics *(previous page).* Tissue optimisation for full thickness ileal samples preserved in 2-methylbutane. Permeabilisation time series from 3mins to 36mins in Run 1 (A), 12mins to 60mins in Run 2 (B) and 6mins to 14mins in Run 3 (C) with corresponding fluorescence imaging. Reference RNA isolated from PBMC used as positive control. Optimum permeabilisation timepoint identified at 11mins (C). [A: Patient GI 7051 B: Patient GI 7833]

3.5.1.4 Pilot experiment-derived optimisations

The pilot experiment included four conditions as described above (see **3.5.1.**). Despite encountering a partial loss of the CD inflammatory section, due to incomplete tissue adherence pre-staining, we achieved robust sequencing saturation (316,996,247 reads) and consistent gene recovery across sections (excluding the inflammatory section) (**Appx. Fig. A.1**). There was no contamination, with most reads mapping to the genome as one would expect and the majority of reads fell under spots of tissue.

Gene recovery overall was satisfactory, with RNA recovery variations correlating with histopathological features. Regions with high immune cell content exhibited poor recovery, as anticipated. Significant diffusion was observed predominantly on the epithelial side of the tissue bidirectionally, with the halo effect being particularly pronounced in the involved sections (**Fig. 3.14a**).

Following the pilot run, I explored whether slightly reducing the permeabilisation time might ameliorate signal diffusion. A repeated TO run (Run 3), identified a new optimal timepoint at 11 minutes (**Fig. 3.13c**).

This revised timepoint was implemented in the subsequent run. Although there was an improvement in signal diffusion in the epithelium from the crypt base, a significant diffusion of crypt top factors persisted. This consideration, among others, led to the decision not to proceed with fresh-frozen ST. Data from the pilot run is presented within the Appendix for reference.

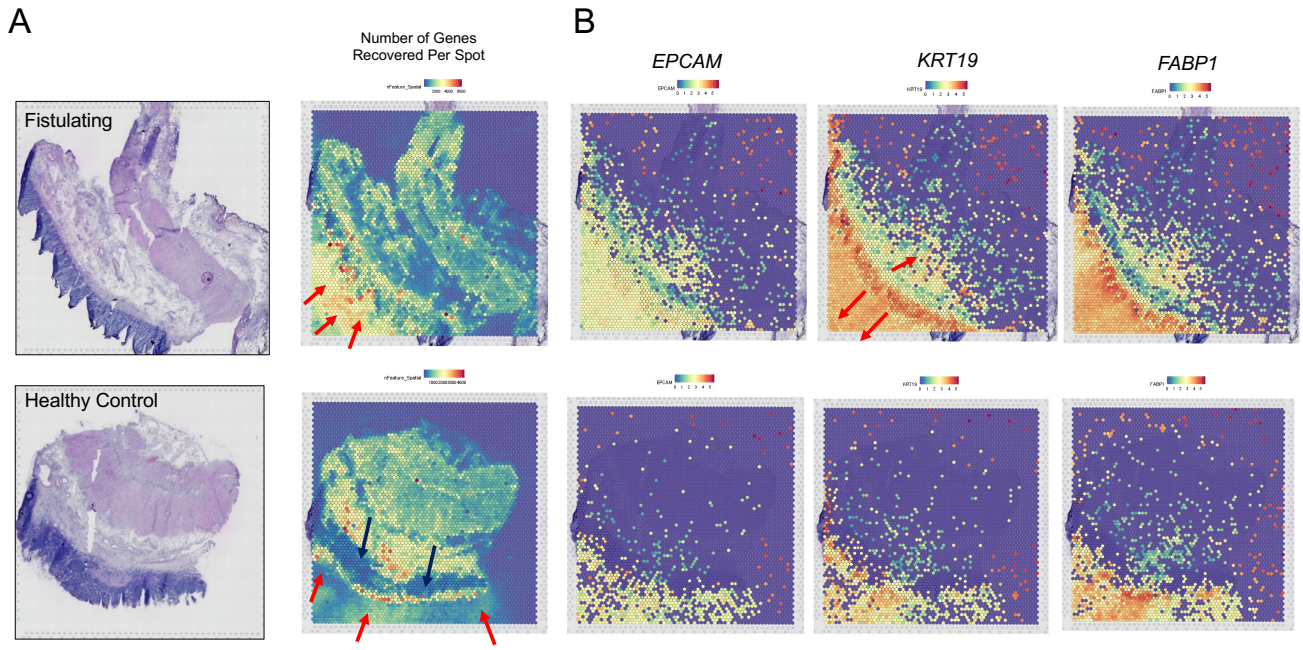


Figure 3.14: Spatial Bleed-Through of Epithelial Signal in Full Thickness Fistulating Crohn's and Healthy Control Spatial Transcriptomes. This figure illustrates the phenomenon of spatial bleed-through of epithelial signal in spatial transcriptomic analysis. (A) Gene recovery in a representative ST slide is depicted for both fistulating CD (top) and healthy control (bottom). Significant signal diffusion, indicated by a halo effect, is observed predominantly on the epithelial sides of the tissue sections (highlighted by red arrows), with minimal diffusion on the non-epithelial sides. Areas of signal drop out are notes in regions of immune cell infiltrate (black arrows). (B) Areas of diffusion mostly enriched for epithelial genes as denoted by expression patterns of *EPCAM*, *KRT19* and *FABP*. [Fistulating CD (Patient TIP 535) and Healthy control (Patient GI 6968)]

3.5.2 FFPE ST workflow

The exploration of the FFPE ST workflow was prompted by the potential to access archived diagnostic tissue blocks stored in formalin from our local hospital and collaborators. This approach significantly expanded our sample cohort by enabling retrospective sampling in addition to prospective collections. For the first time in our group, we trialled the FFPE ST workflow.

During mock runs, I encountered issues with tissue detachment and subsequently performed several optimisations to the 10x Genomics Visium tissue preparation workflow to increase tissue adherence (see **Methods** for details). Unlike the FF workflow, the FFPE workflow utilises a universal permeabilisation condition, eliminating the need for tissue optimisation experiments. Consequently,

we noted a significant improvement in the diffusion of the signal with the FFPE workflow compared to the FF workflow (**Fig. 3.15**).

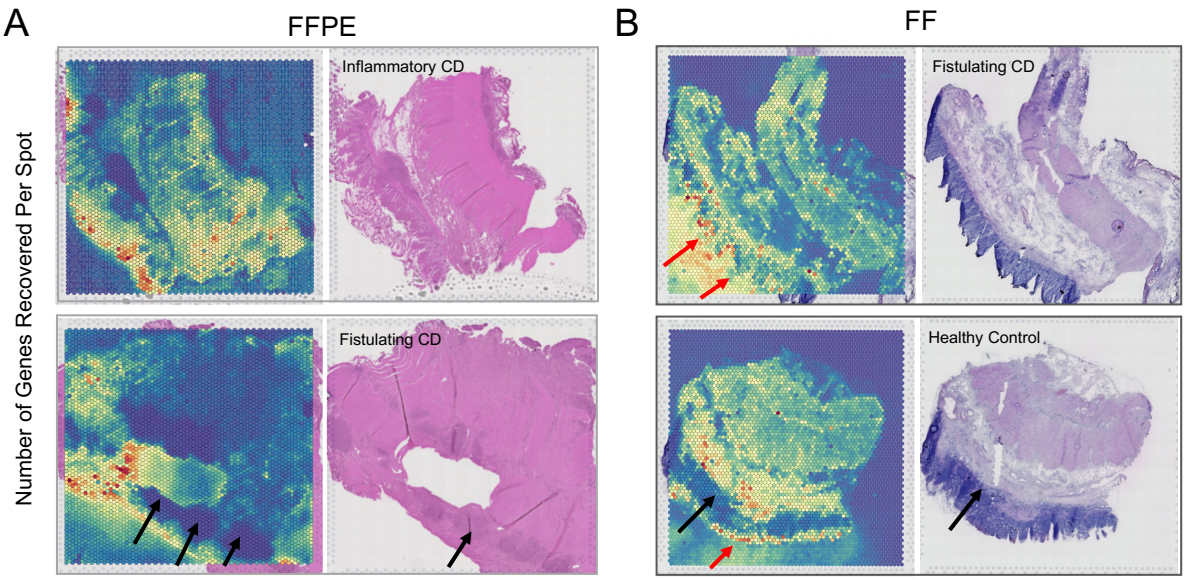


Figure 3.15: Improved Signal Diffusion in FFPE ST. Gene recovery in representative ST slides from FFPE tissue (A) and FF tissue (B) demonstrate an improvement in signal diffusion (red arrows) in FFPE compared to FF samples. However, there is still persistence of signal drop-out in immune rich regions (black arrows) in both FFPE and FF transcriptomes. [A – Inflammatory CD (Patient GI 6966), Fistulating CD (Patient TIP 559); B – Fistulating CD (Patient TIP 535), Healthy control (Patient GI 6968)]

RNA quality assessment for FFPE tissue is measured by the DV200, with a value >50% deemed suitable to proceed with spatial analysis. RNA extracted from FFPE samples tends to be of lower quality due to the chemical fixation process, which can cause fragmentation and cross-linking of RNA molecules. Interestingly, we observed a poor correlation between gene recovery and DV200 values, suggesting that DV200 was not a reliable predictor of data quality (**Fig. 3.16**). Instead, we ensured samples had good tissue morphology and no significant tissue artefacts.

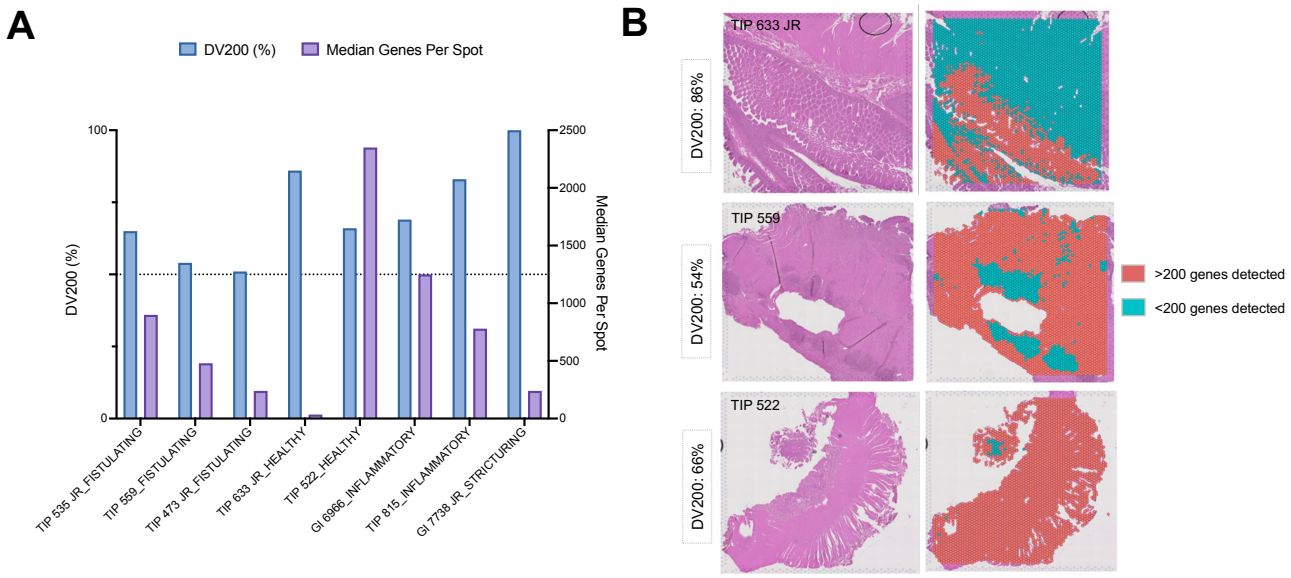


Figure 3.16: Gene Detection and DV200 Correlation. Gene detection and DV200 correlation in FFPE spatial transcriptomes. (A) Bar chart depicting DV200 value and median gene per spot count per sample. (B) Representative spatial transcriptomes of healthy and diseased full thickness CD samples visualising gene detection (cut off 200 genes per spot) with associated DV200.

Building upon the success of FFPE Visium, we explored subcellular FFPE ST using the Vizgen MERSCOPE platform. We were the first to pilot FT intestinal tissue on the platform by sampling three FT adult, Crohn's intestinal sections for the verification (optimisation) run. Pleasingly, we observed a polyT and DAPI signal, albeit with variations across samples types and within the sample (**Fig. 3.17a**). These initial findings indicated the potential of the technology following some adaptations.

To adapt the established protocol for larger FT intestinal tissue, we made several modifications, including extending the duration of tissue digestion and increasing the concentration of proteinase K. We also extended tissue clearance and photobleaching, to quench autofluorescence signals, prior to imaging. Subsequently, we observed some improvement in transcript recovery in the same section from Run 1 compared to Run 2 (**Fig. 3.17b**).

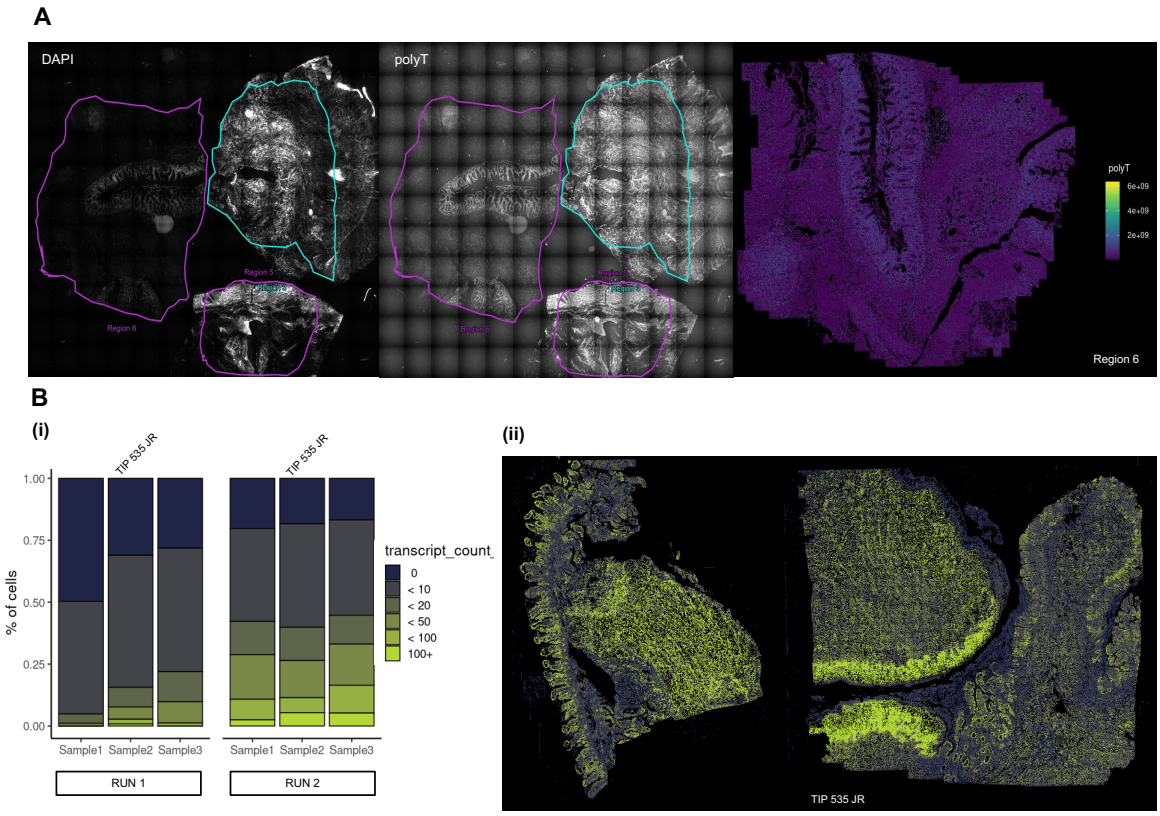


Figure 3.17: MERFISH RNA Transcript Recovery Analysis. (A) Visualisation of DAPI and polyT signals in full thickness fistulating Crohn's samples. Signal intensity is represented by brightness. (B) Stacked bar chart illustrating the transcript count per sample for both Run 1 and Run 2. Note that Sample 2 (Patient TIP 535 JR) is the same in both runs (i). The MERFISH spatial transcriptome of two full thickness intestinal samples visually displays the transcript count (ii) (colour key displayed). [A – Region 3: Patient JR 9174, Region 5: Patient BAY 113385, Region 6: Patient JR 17451; B – L: TIP 852, R: TIP 535 JR]

However, in subsequent runs, we encountered varied results despite running samples with DV200 values >50% (deemed 'good' quality by the MERSCOPE protocol). Issues highlighted included low or no gene recoveries in specific regions of tissue, inconsistent correlation between polyT region signal and transcript recovery, and occasional loss of focus or failed cell segmentation (**Fig. 3.18a-b**). Despite these challenges, where signal was recovered, it exhibited biologically relevant expression patterns, demonstrating high sensitivity and specificity **(Fig. 3.18c**), particularly around the fistula tract.

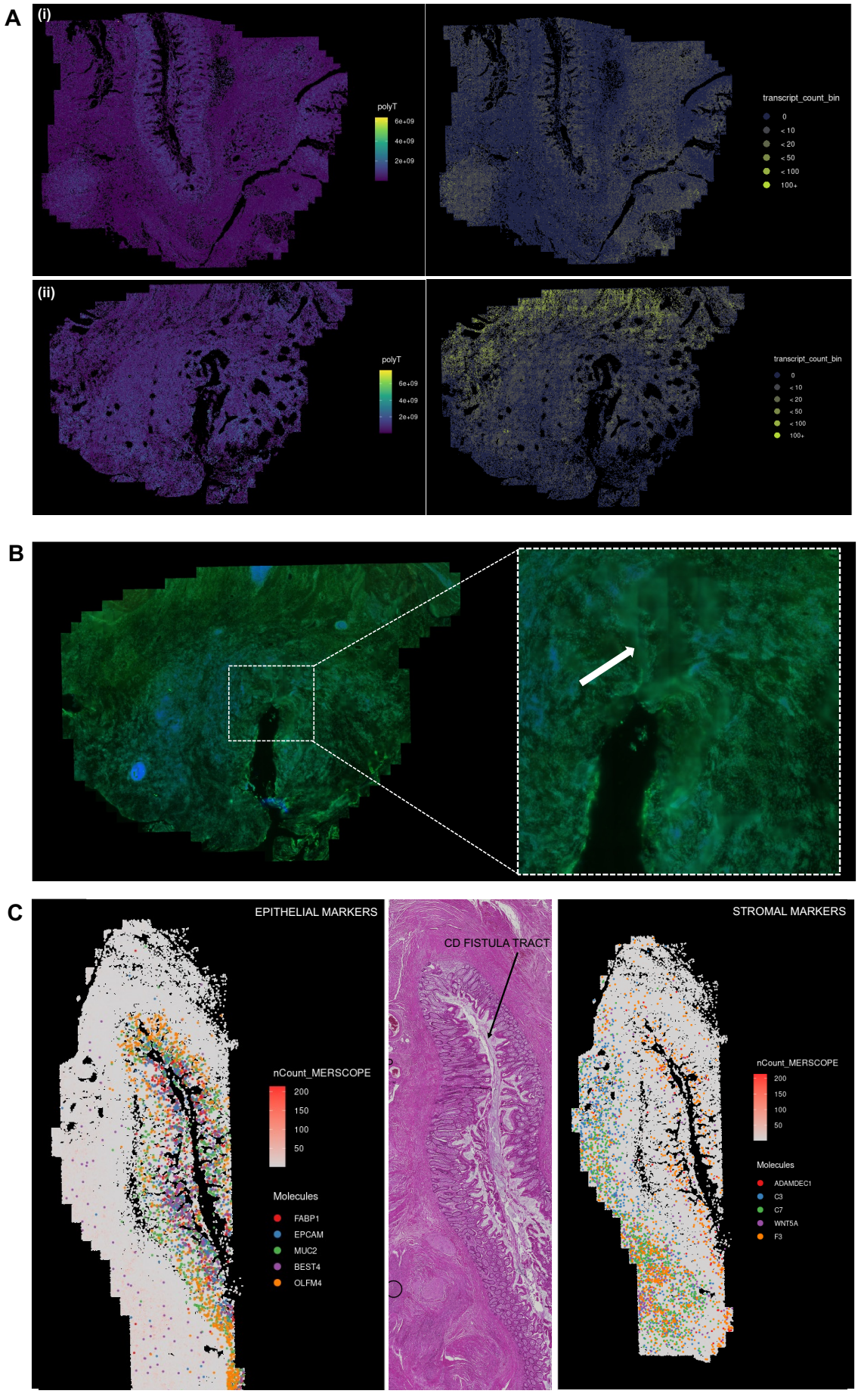

A
(i)
polyT
6e+09
4e+09
2e+09
transcript_count_bin
0
< 10
< 20
< 50
< 100
100+
(ii)
polyT
6e+09
4e+09
2e+09
transcript_count_bin
0
< 10
< 20
< 50
< 100
100+
B
C
EPITHELIAL MARKERS
nCount_MERSCOPE
200
150
100
50
Molecules
FABP1
EPCAM
MUC2
BEST4
OLFM4
CD FISTULA TRACT
STROMAL MARKERS
nCount_MERSCOPE
200
150
100
50
Molecules
ADAMDEC1
C3
C7
WNT5A
F3

Figure 3.18: MERFISH Quality Check (*previous page*). (A) MERFISH visualisation of paired polyT and RNA transcript recovery signal across two representative fistulating Crohn's sections demonstrating correlation (i) and lack of correlation (ii). [i, Patient JR 17451; ii, JR 9174] (B) Example of loss of focus within the field of view (Patient JR 9174). (C) Visual representation of the sensitivity and specificity of the MERFISH 500-plex panel to accurately localise epithelial markers (*FABP1, EPCAM, MUC2, BEST4, OLMF4*) and stromal markers (*ADAMDEC1, C3, C7, WNT5A, F3*) within a fistulating Crohn's section (Patient JR 17451).

Despite further optimisations, we continued to observe inconsistent results, leading us to conclude that inherent limitations exist within the chemistry of the probe hybridisation technology and tissue type. This conclusion further supported by unsuccessful attempts made by Vizgen's Research & Development team to optimise the technology for our samples. Subsequently, a successful pilot with the 10x Genomics High Resolution ST platform, Xenium, using the same tissue samples that had previously failed with MERSCOPE, provided additional evidence supporting the presence of constraints within MERSCOPE's probe hybridisation chemistry.

For the purposes of my thesis, I will include subcellular ST results from Xenium only (**Chapter 5**).

3.6 Interim discussion

This chapter serves as a comprehensive documentation of a series of optimisation experiments undertaken to prepare FT intestinal tissue for scRNA-seq and spatial transcriptomic analysis in the context of CD. Specifically, the protocols presented enabled the successful isolation of viable immune, stromal and epithelial cells from FT tissue. Among the protocols tested, the Lamina Propria (CM modified – 2 hours) protocol, emerged as the most promising for yielding viable immune and stromal cell populations, as confirmed by subsequent flow cytometry analysis. Additionally, optimisations in epithelial crypt isolation and antibody labelling were key in enabling cell viability and preserving epitopes, essential for downstream analyses.

Further optimisation was achieved with the implementation of LeviCell, which significantly enhanced the recovery of stromal cells, particularly in diseased tissue settings and improved the cell yield and quality of scRNA-seq data from challenging samples without major impacts on sample cell

type composition. Notably, the optimised LeviCell protocol was found to be compatible with antibody labelling, enabling multi-modal profiling such as CITE-Seq or sample multiplexing. This multiplexing capability simplifies and shortens the lab workflow by allowing the simultaneous sequencing of multiple samples, thereby avoiding batch correction issues associated with individual sample runs.

Efforts to enhance scRNA-seq data quality and sequencing coverage through CRISPR technologies yielded promising results, reducing mitochondrial content and significantly increasing sequencing saturation, thereby enhancing gene detection rates. The CRISPRclean workflow also allowed the redistribution of sequencing coverage, reducing sequencing costs - a significant consideration given the high costs associated with scRNA-sequencing.

In ST, our pilot experiments with FF and FFPE tissue workflows highlighted the importance of tissue preparation and permeabilisation optimisation for accurate spatial analysis. While challenges persisted with certain methodologies, notably FF and MERSCOPE, robust methodologies for FFPE ST were successfully replicated in 10x Genomics Visium and Xenium workflows.

These findings serve as the foundation for the data presented in the rest of this thesis, contributing valuable insights into the molecular landscape of fistulating CD.

Chapter 4: Unbiased Single-Cell Transcriptomic Atlas of CD-Associated Fistula Tissue

4.1 Introduction

In this part of the project, my aim was to generate a multi-modal single-cell reference atlas to comprehensively compare intestinal homeostasis and dysregulation in fistulating CD. Fistulating CD pathology is widely thought to be characterised by an initial epithelial defect arising from chronic intestinal inflammation, dysregulated wound healing processes and transmural inflammation extending throughout the entire thickness of the intestinal wall. These pathological processes culminate in significant alterations to the normal tissue architecture resulting in the formation of an abnormal passage i.e., a fistula (7, 207).

Given the complex nature of fistulating disease, it was important to transcriptomically profile the epithelial, stromal, and immune compartments of FT tissue sequentially. Such profiling allowed for the delineation of the epithelial and mesenchymal remodelling patterns, along with immune landscape changes unique to fistulating tissue.

In the preceding chapter, a series of methodologies were charted to optimise the single-cell workflow specifically for diseased, FT tissue analysis. These optimisations were crucial in enabling the subsequent identification of novel, fistula-specific stromal populations with enhanced precision and resolution.

The results provide novel insights into the heterogeneity and functional characteristics of epithelial, stromal, and immune cell populations within both involved and uninvolved fistulating CD tissue, in comparison to controls. This comprehensive characterisation provides unprecedented depth into the cellular dynamics underlying fistulating CD pathology.

The primary objective of this chapter is to delineate the cellular landscape of fistulating CD tissue through the application of scRNA-seq analysis. This cellular atlas also serves as a foundational reference for the interpretation of subsequent spatial analyses, further elucidating the spatial organisation and interactions within the diseased intestinal microenvironment.

4.2 Clinical cohort

Following obtaining informed consent, tissue samples were collected from patients diagnosed with fistulating, inflammatory and stricturing CD who underwent elective or emergent surgical ileal resection, as well as patients without IBD who served as healthy controls. These samples were obtained from areas of involvement, including the fistula tract in cases of fistulating tissue, as well as from distant uninvolved (non-involved) areas. Pathology status was initially assessed through macroscopic evaluation by the operating surgeon and subsequently confirmed through histopathological examination.

In total, the study profiled 32 tissue samples sourced from 22 individual donors. This cohort was carefully balanced, comprising both involved and paired uninvolved tissue samples. The breakdown

of samples based by condition, cellular compartment, involvement status, and cell count, is delineated in **Figure 4.1**.

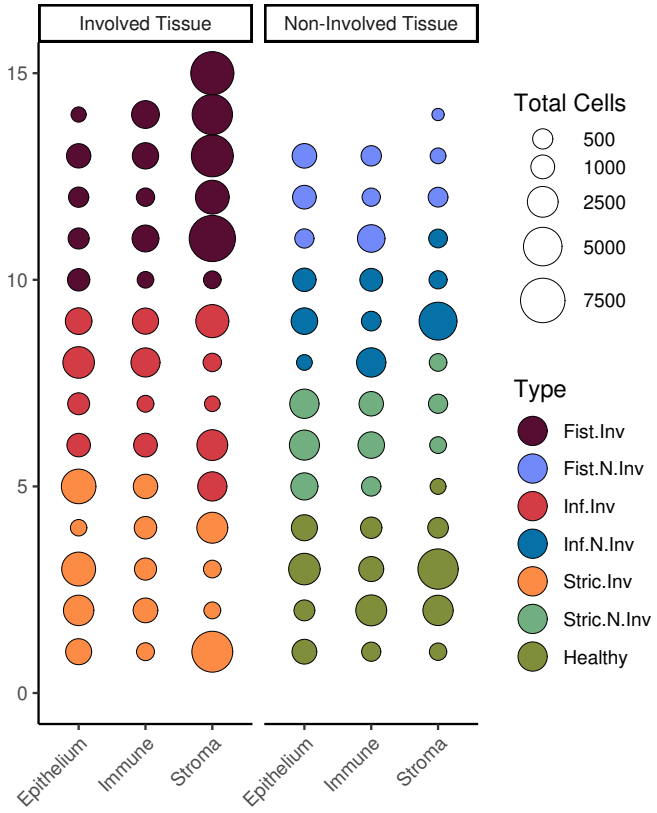


Figure 4.1: Single-Cell RNA-Sequencing Sample Overview. scRNA-seq experiment sample overview dot-plot depicting sample distribution across condition type, involvement status, cellular compartment and number of high-quality post-QC cells recovered per sample. [Fist, Fistulating; Inf; Inflammatory; Inv, Involved; N.Inv, Non-involved; Stric, Stricturing; QC, Quality Control]

4.3 Experimental overview

4.3.1 Cell isolation, sequencing and initial analyses

Tissue samples were promptly processed upon collection and cryopreserved to enable simultaneous processing for scRNA-seq, thus minimising batch effects. Optimised protocols, detailed in **Chapter 3**, were employed to isolate viable cells from FT tissue, enhance cell yield and quality, and optimise sequencing for scRNA-seq experiments. Lamina propria digestion and epithelial crypt chelation

protocols were applied to obtain single-cell suspensions of stromal, immune and epithelial cells from full thickness tissue.

Cells isolated via lamina propria digestion underwent FACS or LeviCell isolation to obtain stromal and immune populations (see **Fig. 4.2; Methods 2.6, 2.7**). Each sequencing run included involved, paired uninvolved and/or healthy samples from at least three donors (total of 5 samples) (**Appx. Table A.1**). Patient samples were pooled by reaction type (stromal, immune and epithelial) and subjected to 5' droplet-based technology (Chromium, 10x Genomics, **Methods 2.8**) to generate CRISPR-depleted gene expression, cell surface protein, and T-cell receptor (TCR) libraries. Final libraries were quantified, pooled, and sequenced on Illumina platforms to specific sequencing depths: 25,000 read pairs per cell for gene expression libraries and 5,000 read pairs per cell for TCR (V(D)J) enriched and cell surface protein libraries.

Raw sequencing data underwent initial processing with bcl2fastq software and quality checks using FastQC. The data were aligned to the human hg38 reference genome and transcript UMI counts were summarised through the Cell Ranger pipeline. Subsequent analysis utilised the Seurat R package, with batch correction performed using the Harmony algorithm. Cluster marker detection, differential gene expression analysis, proportional analyses, and visualisation were conducted using various R packages.

The scRNA-seq data outlined in this chapter is categorised according to cellular compartment.

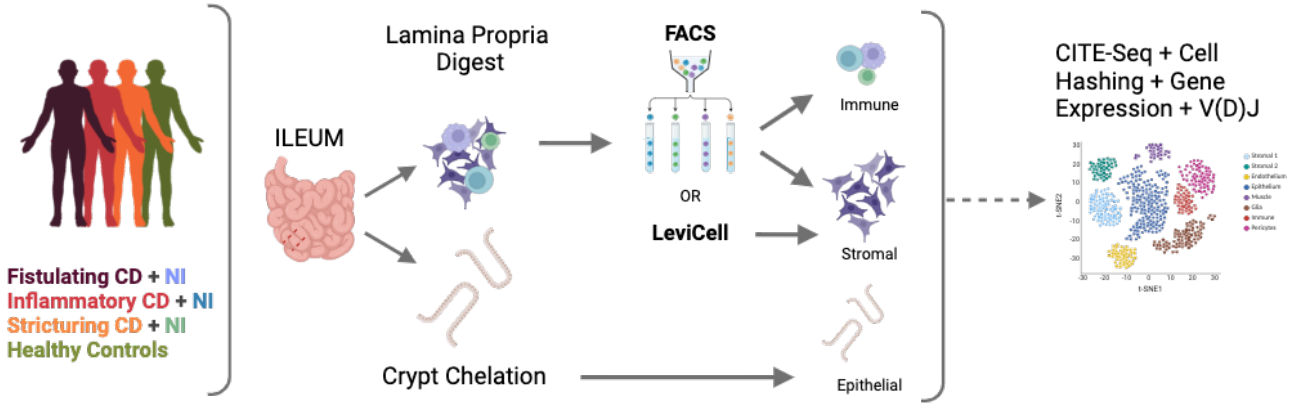


Figure 4.2: Single-Cell RNA-Sequencing Experimental Workflow. Schematic providing an overview of the experimental design for multiplexed scRNA-seq analysis of fistulating, inflammatory and stricturing CD samples along with healthy controls. Surgically resected ileal tissue from each condition underwent enzymatic lamina propria digestion. The resulting cell suspension was then processed either through FACS to isolate stromal (CD45$^-$/EpCAM$^-$) and immune (CD45$^+$/EpCAM$^-$) populations or through LeviCell CD45 depletion to isolate single-cell stromal population. Additionally, full-thickness tissue underwent epithelial crypt chelation to isolate epithelial cells, followed by the generation of cell surface protein, gene expression and TCR libraries for sequencing using the 10x Genomics Chromium. [CD, Crohn's disease; FACS, Fluorescence-activated cell sorting; TCR, T-cell receptor; NI, Non-involved]

4.4 Epithelial scRNA-seq data

4.4.1 Cell quality control process & cell recoveries

In total, 37,133 epithelial (EpCAM$^+$) cells were successfully recovered from the samples. After quality control (QC) and batch correction procedures, unsupervised clustering identified the major epithelial cell types within the small intestine (**Fig. 4.3a**), annotated using cluster marker gene annotations from established datasets (167-169). Notably, these clusters exhibited consistent integration across multiple runs, with minimal need for further batch correction (**Fig. 4.3b(i)**).

Table 4.1: Cell Recoveries by Condition and Run, post QC

		Stromal	Epithelial	Immune
Condition	Fistulating INV	30996	3098	5331
	Fistulating UI	633	2516	2561
	Healthy	9022	5878	4803
	Inflammatory INV	8256	5926	4713
	Inflammatory UI	5407	2589	3425
	Stricturing INV	8957	11039	4096
	Stricturing UI	841	6087	3030
	TOTAL	**64,112**	**37,133**	**27,959**

		Stromal	Epithelial	Immune
Run	Run 1	6153	2667	9003
	Run 2	1673	7941	2340
	Run 3	1088	7426	7132
	Run 4	1857	6619	4324
	Run 5	0	12480	5160
	LeviCell Run 1	6858	-	-
	LeviCell Run 2	9862	-	-
	LeviCell Run 3	6022	-	-
	LeviCell Run 4	7630	-	-
	LeviCell Run 5	12554	-	-
	LeviCell Run 6	10415	-	-
	TOTAL	**64,112**	**37,133**	**27,959**

INV, involved; UI, uninvolved; QC, quality control

Only 4,209 (2,667 post-QC) singlet epithelial cells were recovered from Run 1, the least among all the runs (**Table 4.1**). To address this limitation and increase the number of cells recovered from the epithelial fraction, a decision was made to load the pooled epithelial reaction twice in subsequent runs. This adjustment resulted in a significant increase in epithelial cell yield, with a mean total cell recovery per run of 7,426.

4.4.2 Cell cluster annotations and characterisation

Upon UMAP embedding visualisation, eighteen clusters were identified across 37,133 cells in the epithelial dataset (**Fig. 4.3a**). These clusters encompassed all of the key secretory and absorptive populations (**Fig. 4.3c**). For instance, EECs expressing *CHGA* and *KCNB2*, and goblet cells expressing genes involved in the maintenance of the intestinal barrier (*MUC2*, *SPINK4*, and *FCGBP)*.

Enterocytes, the predominant absorptive cell type, were distinguished by *ALDOB*, *CDHR5*, and *CYP3A5* expression. Stem cells and their daughter cells, transit amplifying (TA) cells, exhibited shared gene expression of *GPX2*, *REG1A*, *OLFM4*, *KRT18*, and *SOX4*, with higher expression of *GPX2*, *REG1A*, and *OLFM4* in TA cells, reflecting their more proliferative and regenerative state within the intestinal crypt.

Additionally, rarer populations such as BEST4, characterised by *OTOP2*, *CA7*, *GUCA2B* expression (167), and Tuft cells, characterised by *FYB1*, *LRMP*, and *PLCG2*, were readily identified.

Identification of an EpCAM⁺ IL11⁺ population

A previously undescribed population within the epithelial dataset was identified, characterised by *IL11* expression, and appeared to be specific to involved tissue in fistulating and stricturing CD (**Fig. 4.3a, b(ii), c**). The top markers of this novel cluster (**Fig. 4.3c**) suggest activation of cellular processes related to epithelial turnover (*KRT18*), tissue repair (*CDHR5*, *IL11*), mucosal inflammation, immune responses (*CXCL1/5*), and metabolism (*ALDOB*).

Specifically, the cells of this population exhibited significant expression of pro-fibrotic cytokine IL11, known to promote tissue repair, inflammation and modulate the intestinal immune response (208, 209). Interestingly, IL11 has been linked to a number of inflammatory diseases including CD, in which IL11 is correlated with CD risk allele, *NOD2* (210). Chemokines CXCL1 and 5, also expressed by these cells, play a role in neutrophilic recruitment, inflammation, and immune responses. Additionally, the expression of *CDHR5*, a member of the cadherin superfamily which gives rise to a cell adhesion molecule (211), suggests a role in modulating cell-to-cell interactions and tissue morphogenesis in response to injury. The upregulation of metabolic pathway activity through *ALDOB* and *ANPEP* may be required to support increased cellular proliferation in diseased states.

This observation suggested a cluster specific to CD, potentially involved in tissue repair, immune recruitment and altered metabolism. However, the population size was notably small (40 cells),

mainly arising from a single run (**Fig. 4.3b(i)**). Despite identification of this cluster across more than one sample, additional sampling is warranted to validate these findings for confirmation of its significance in intestinal epithelia.

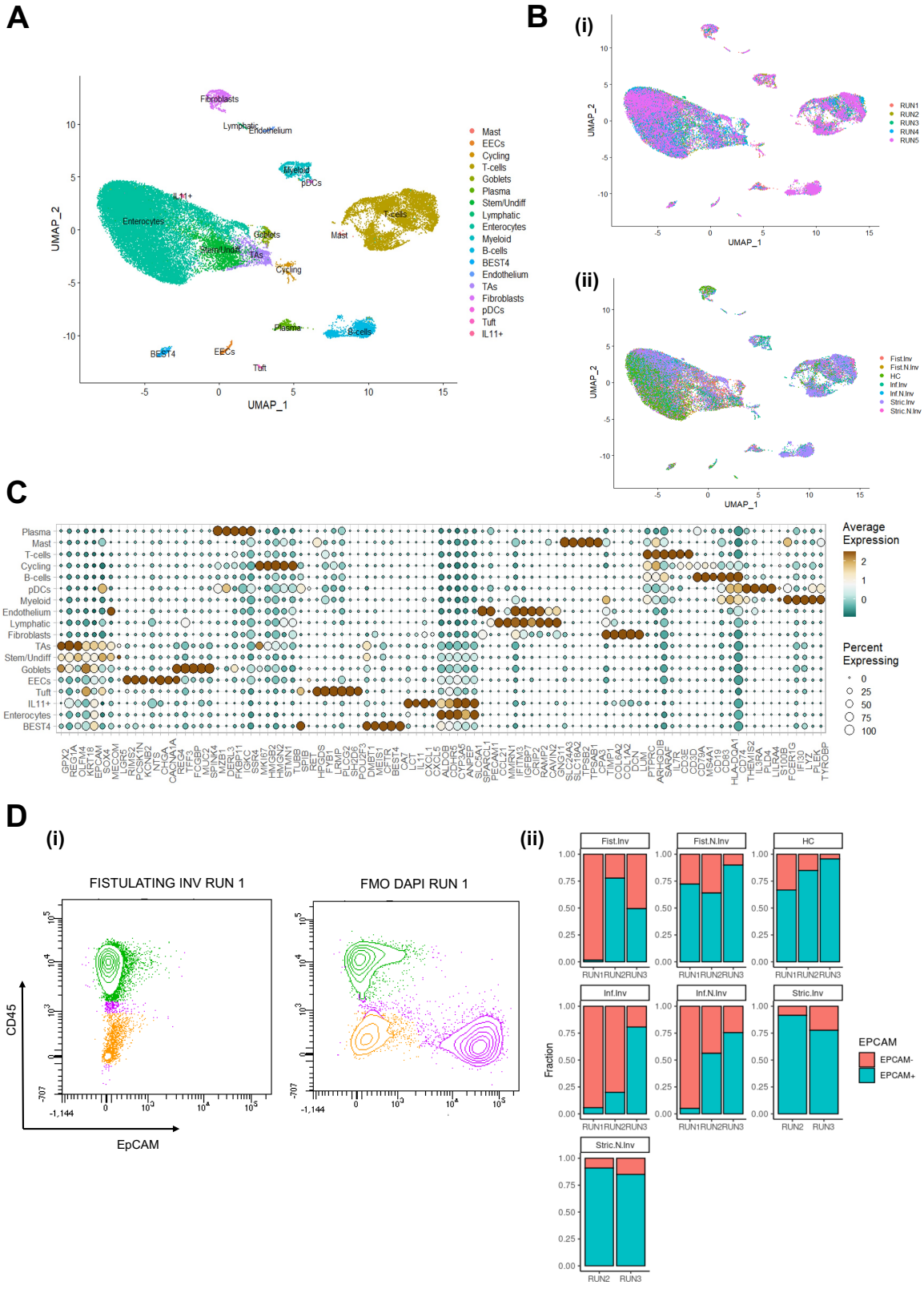


Figure 4.3: Merged Epithelial scRNA-Seq Data (*previous page*). (A) UMAP embedding visualising scRNA-seq clusters of cell populations obtained from epithelial crypt dissociation. (B) UMAP embedding visualising scRNA-seq clusters of epithelial cells by single cell run (i) and by condition (ii). (C) Dot-plot heatmap showing top epithelial marker gene expression by clusters as identified in (A), coloured by average expression. The relative size of the dot represents the percentage of cells per cluster that expresses each marker. (D) FACS plot of fistulating involved sample (Run 1) demonstrating the absence of an EpCAM+ population compared to the populations identified in FMO-DAPI (i). Proportion of EpCAM+ cells per condition and per run (Run 1 – 3 only) (ii). [FACS, Fluorescence-activated cell sorting; FMO, Fluorescence minus one]

4.4.3 Epithelial cell proportions in fistulating involved CD vs controls

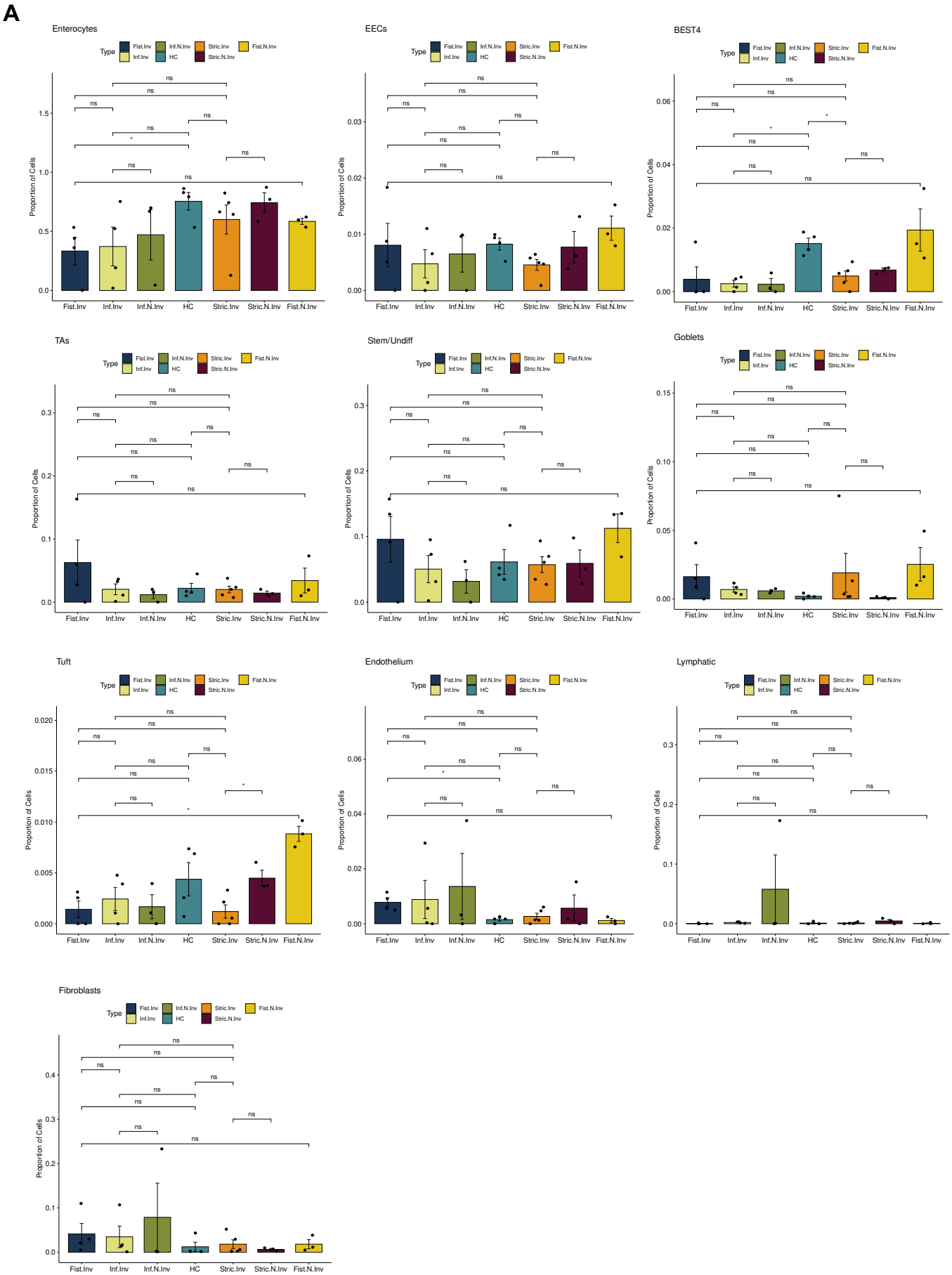
In the initial run, a notably low yield of EpCAM+ cells from fistulating involved tissue raised questions regarding potential biological variance or sample dependence (**Fig. 4.3d**). This observation was further confirmed by FACS analysis, where the absence of EpCAM+ cells in the fistulating involved sample of Run 1 was evident, contrasting with successful identification of both EpCAM and CD45 positive populations in DAPI gating controls (**Fig. 4.3d(i)**). However, subsequent runs demonstrated successful recovery of EpCAM+ cells from fistulating involved tissue, suggesting sample dependency in the case of Run 1 (**Fig. 4.3d(ii)**). Interestingly, a significant proportion of the EpCAM+ cells in Run 1 were found to be intraepithelial T-cells. Consequently, TCR libraries were generated from the epithelial fraction in addition to the immune fraction in subsequent runs to further explore this observation (**see 4.6**).

Variations between epithelial cell populations by condition and per run are illustrated in **Figure 4.4** (and **Appx. Fig. A.2**). While significance was not reached, trends were observed showing that differentiated epithelial cells such as enterocytes, tuft cells, and BEST4 were among the least abundant in fistulating involved tissue, whereas undifferentiated epithelial cells; transit amplifying cells, and stem cells were among the most expanded in fistulating involved tissue compared to controls (**Fig. 4.4a**).

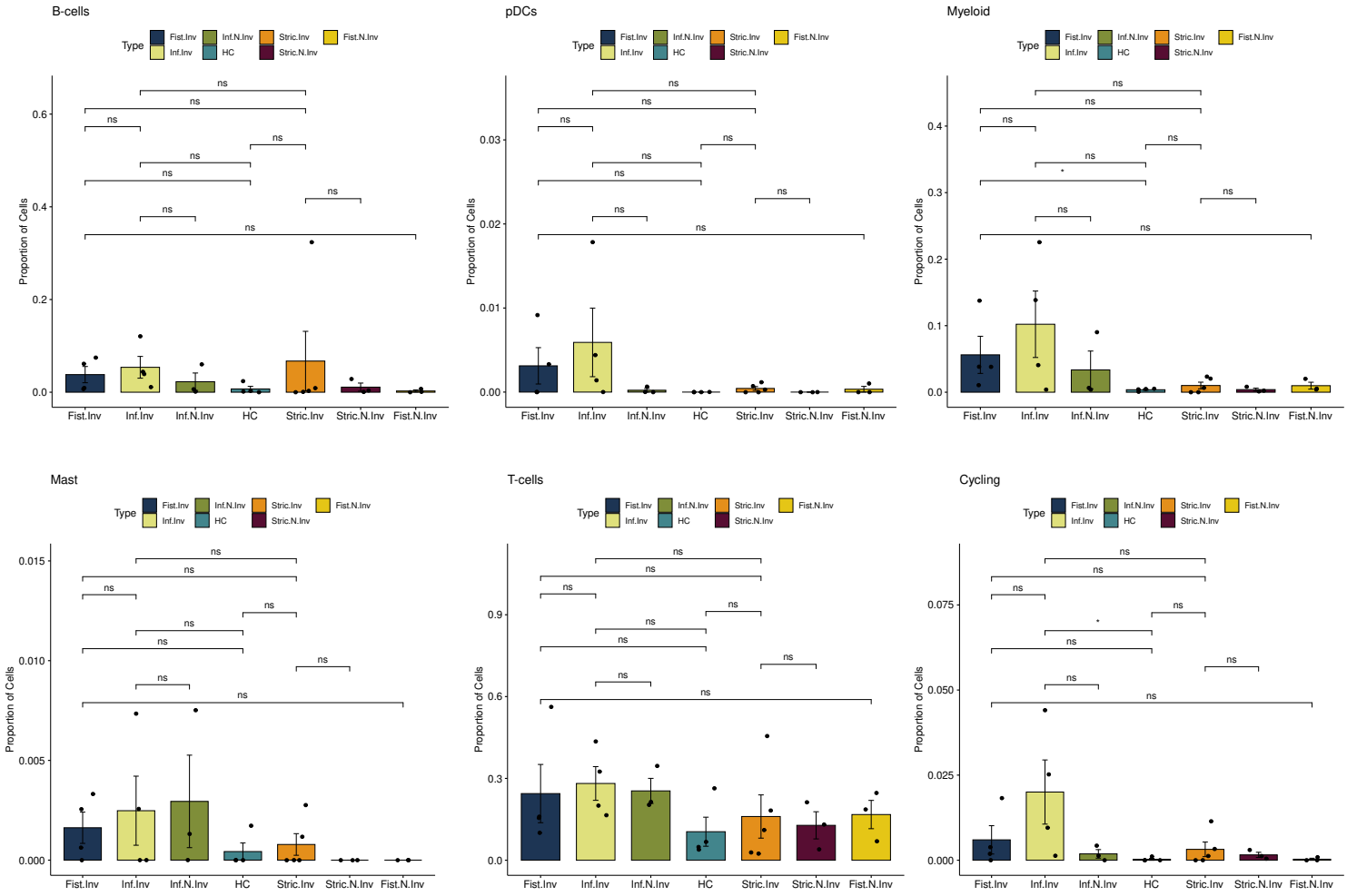
Variations in immune cell composition within the epithelial fraction were also observed, with an expansion of myeloid, plasma and B-cells in fistulating involved and inflammatory CD involved

samples (**Fig. 4.4b-c**). Additionally, an expansion of cycling T-cells was identified in the involved samples (fistulating, inflammatory, and stricturing) over healthy and uninvolved tissue.

Notably, tuft cells and enterocytes were found to be least abundant in fistulating involved tissue, while goblet cells appeared to be the most abundant secretory population in fistulating involved tissue.


A
Enterocytes
EECs
BEST4
TAs
Stem/Undiff
Goblets
Tuft
Endothelium
Lymphatic
Fibroblasts
Type
Fist.Inv
Inf.Inv
Inf.N.Inv
HC
Stric.Inv
Stric.N.Inv
Fist.N.Inv
Proportion of Cells
ns
*

B


B-cells
pDCs
Myeloid
Mast
T-cells
Cycling
Type
Fist.Inv
Inf.Inv
Inf.N.Inv
HC
Stric.Inv
Stric.N.Inv
Fist.N.Inv
Proportion of Cells
ns
*
Fist.Inv
Inf.Inv
Inf.N.Inv
HC
Stric.Inv
Stric.N.Inv
Fist.N.Inv

C

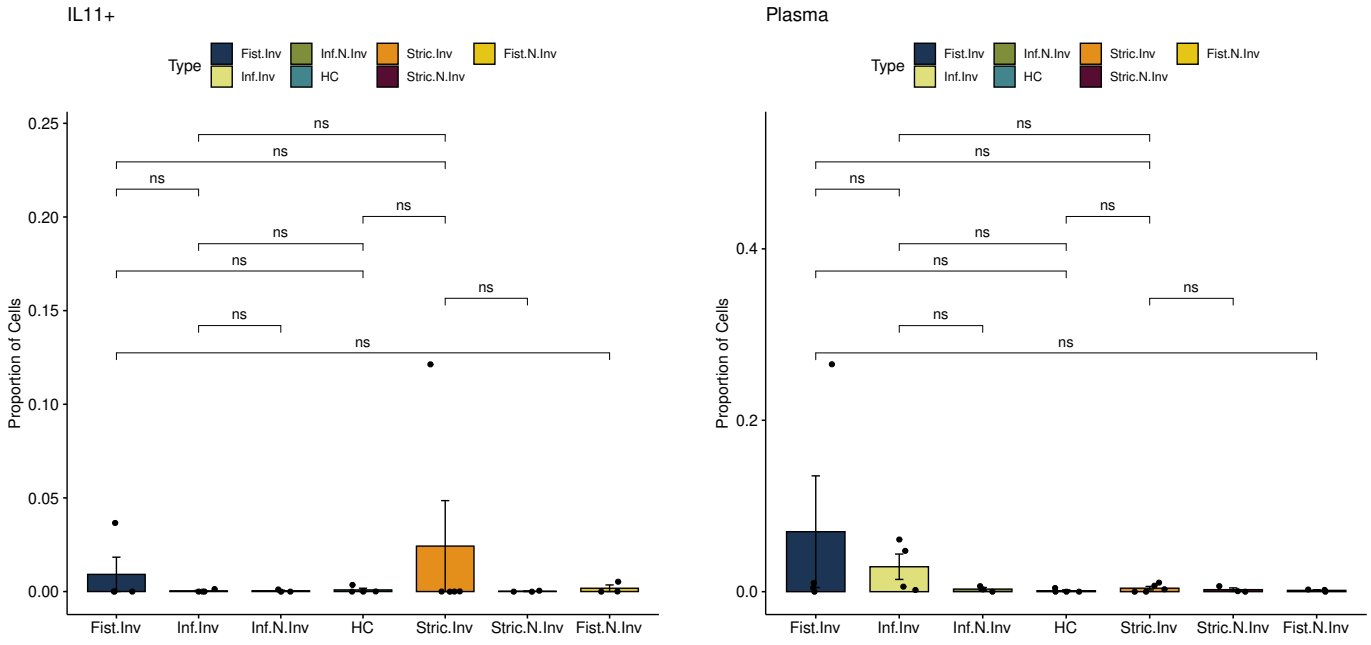

IL11+
Plasma
Type
Fist.Inv
Inf.Inv
Inf.N.Inv
HC
Stric.Inv
Stric.N.Inv
Fist.N.Inv
Proportion of Cells
ns
Fist.Inv
Inf.Inv
Inf.N.Inv
HC
Stric.Inv
Stric.N.Inv
Fist.N.Inv

Figure 4.4: Cellular Composition Differences in Epithelial Isolation Reactions (*previous pages).* Epithelial scRNA-Seq data showing: (A) Cellular composition differences in epithelial crypt isolation reactions, per cluster, including differentiated and undifferentiated epithelial cells. (B) Differences in immune cell clusters within epithelial crypt isolation reactions. (C) Composition differences in $EpCAM^+$ $IL11^+$ cluster and plasma cell clusters. Statistical significance is set at $p<0.05$ using Wilcox test [EECs, enteroendocrine cells; Fist, Fistulating; HC, Healthy control; Inf, Inflammatory; Inv, involved; N inv, Non-involved; Stri, Stricturing; TAs, transit amplifying cells]

4.4.4 Differential gene expression and gene ontology enrichment analysis

Differential gene expression analysis was conducted on total epithelial cells from fistulating involved tissue compared to healthy, non-IBD tissue, revealing distinct patterns (**Fig. 4.5**). Gene ontology (GO) biological process enrichment analysis of fistulating involved tissue indicated a downregulation of metabolic processes compared to healthy tissue (**Fig. 4.5b(i)**). Notably, pathways associated with the regulation of cell-to-cell adhesion, immune cellular interactions, and nutrient transport were significantly upregulated in fistulating tissue, as revealed by GO enrichment analysis.

A simultaneous induction of genes involved in cell survival (*PHLDA1, BAG3*), immune responses (*ITLN1*), epithelial differentiation (*SPDEF*) (212), barrier function (*ITLN1*), intercellular communication (*CEACAM5, GJB3*) and nutrient transport (*SLC1A5, LRP8*) (213) was observed in fistulating involved tissue (**Fig. 4.6**). Specifically, elevated expression of *PHLDA1* and *BAG3* in the intestinal epithelium may influence cell survival and apoptosis pathways (214, 215), while increased expression of *CEACAM5* and *GJB3* may modulate cell-to-cell interactions via adhesion molecules and gap junctions, respectively. In addition to its function as a cell adhesion molecule, CAECAM5, upregulated in UC, is implicated in inhibiting cellular differentiation and programmed cell death in colonocytes. Furthermore, it serves as a recognised tumour marker in colorectal cancer (216). GWAS have identified *ITLN1* polymorphisms as risk alleles for CD (217).

The top differentially expressed genes identified in healthy tissue suggest adaptive responses to various physiological conditions, including oxidative stress (*GPX4*) (218), nutrient transport (*SLC52A1*) and xenobiotic exposure (*SULT1A1*) (219) as one might expect (**Fig. 4.5a**).

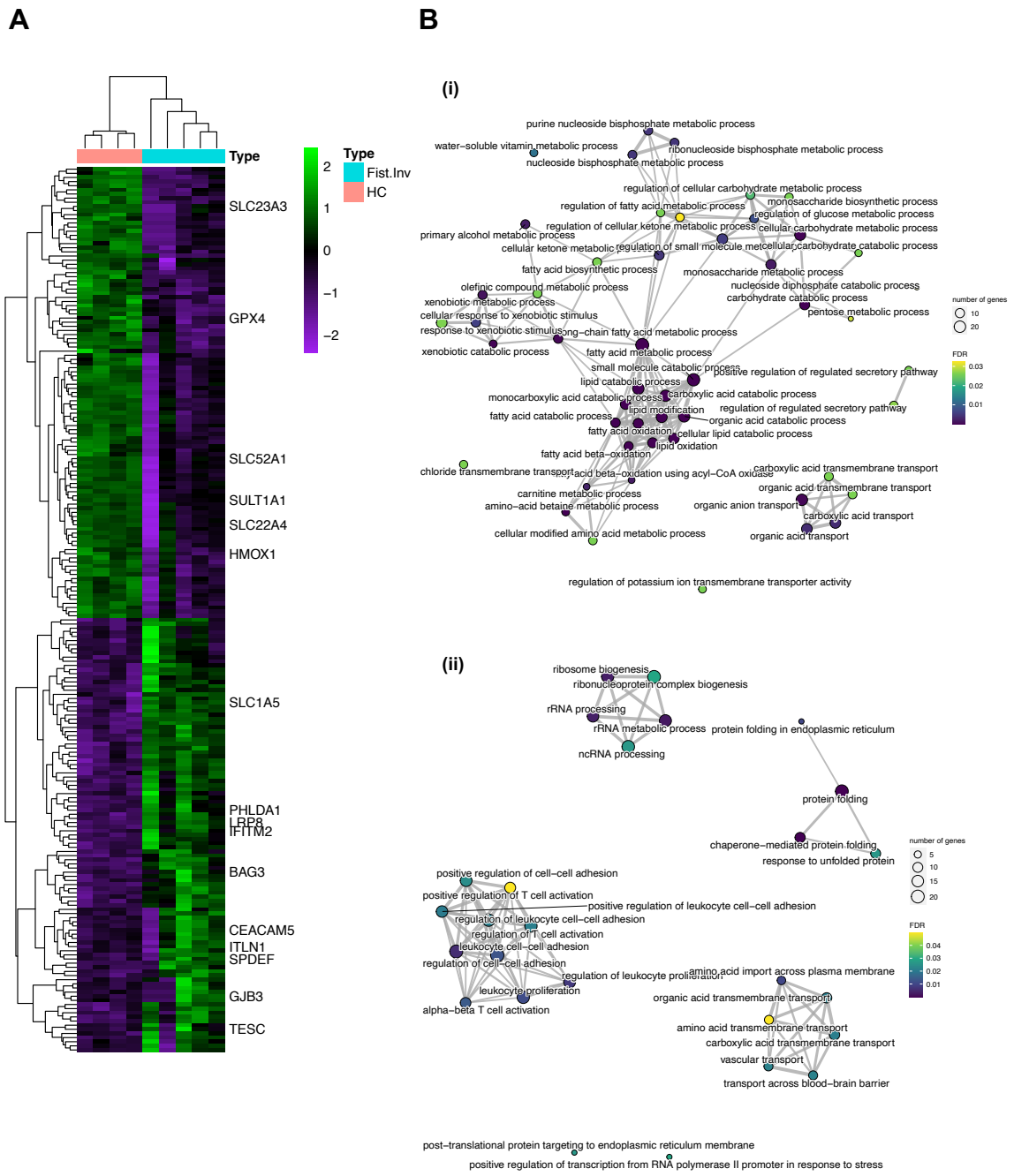
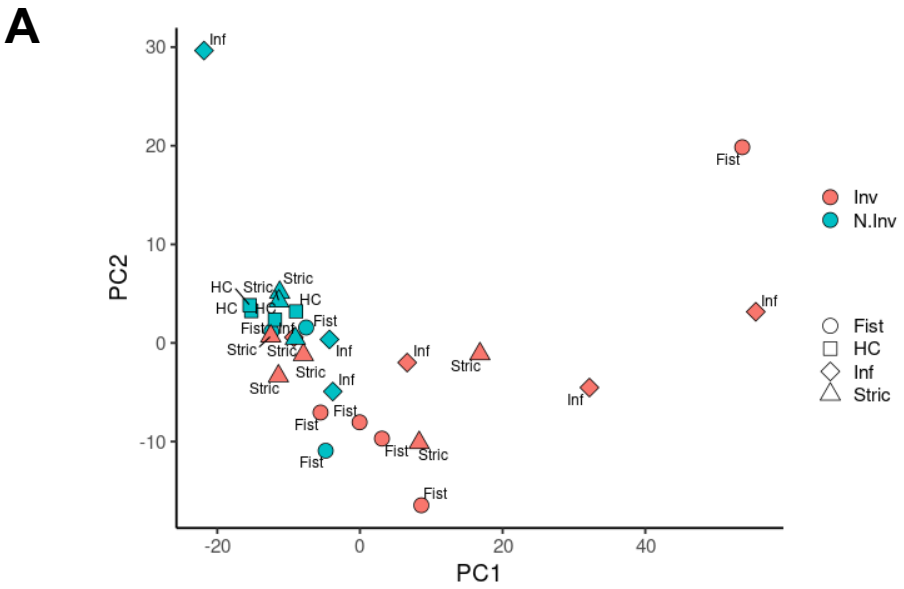


Figure 4.5: Differentially expressed genes enriched in fistulating involved tissue compared to healthy control tissue. (A) Heatmap visualising relative expression of selected differentially expressed genes, identified using the Wald test and corrected for false discovery rate by the Benjamini-Hochberg method, in fistulating involved tissue (blue, n=5) and healthy control tissue (pink, n=4). (B) Network plot showing overrepresented gene ontology terms for the significantly downregulated gene set (i) and significantly upregulated gene set (ii) in fistulating involved cells compared to healthy controls. Statistical significance was assessed using a hypergeometric test and visualised with emapplot.


A
PC2
PC1
Inv
N.Inv
Fist
HC
Inf
Stric


B
Fistulating Involved vs Health
-log10 FDR
log2 Fold Change
DOWN
N.S
UP
HRCT1
ASPA
NEU4
PITPNC1
TSHZ2
LRP8
SLC7A11
AL162725.2
TPM4
FMO1
PLEKHM3
CDV3
PHLDA1
STRIP2
DHDH
STK26
NAAA
AEN
MUC1

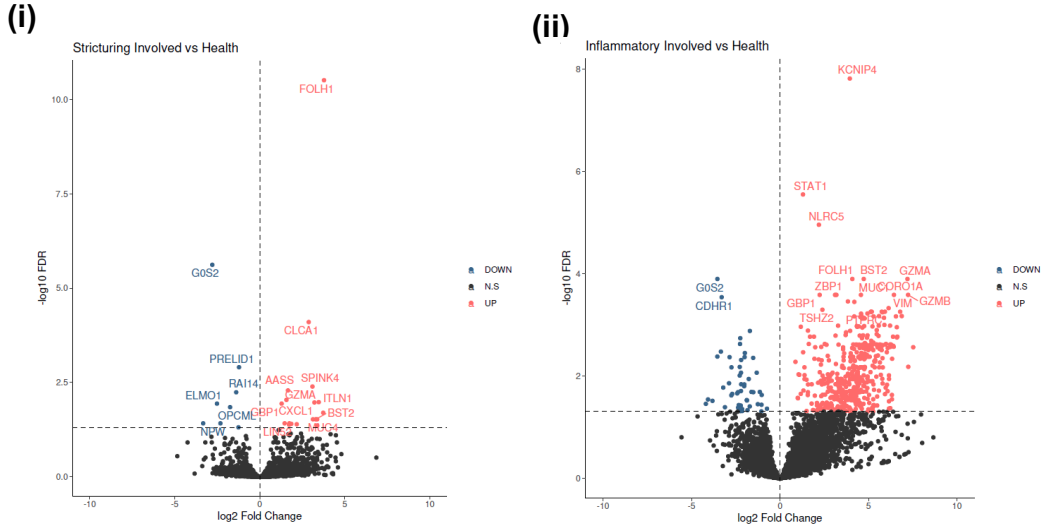

C
(i)
Stricturing Involved vs Health
(ii)
Inflammatory Involved vs Health
-log10 FDR
log2 Fold Change
DOWN
N.S
UP
FOLH1
G0S2
CLCA1
PRELID1
RAI14
AASS
SPINK4
ELMO1
GZMA
ITLN1
CXCL1
BST2
OPCML
KCNIP4
STAT1
NLRC5
CDHR1
ZBP1
MUC1
CORO1A
GBP1
VIM
GZMB
TSHZ2

Figure 4.6: Differentially Expressed Genes in Complex CD Compared to Healthy Control Tissue (*previous page*). (A) Principal component analysis (PCA) based on condition type and involvement status ($EpCAM^+$ cells only). (B) Volcano plot displaying genes that exhibit differential expression between fistulating involved tissue (n=5; positive fold change expressed logarithmically) and healthy control (n=4; negative fold change expressed logarithmically). The black horizontal line represents a 1% false discovery rate, with the most significant genes highlighted. (C) Volcano plots illustrating genes differentially expressed between stricturing involved tissue (n=5) (i) and inflammatory involved (n=4) (ii), compared to healthy controls (n=4).

Enterocyte populations from fistulating involved tissue compared to fistulating uninvolved tissue exhibited an induction of genes associated with maintaining the structural integrity of the intestinal epithelium. These genes are involved in cell adhesion, migration, signalling, and immune regulation (e.g., *ITGA6, CD55, LAMB3*), suggesting dysregulation in these cellular processes (**Fig. 4.7a**). Conversely, downregulated genes such as *REG1A* and *REG1B* might indicate impaired capacity for epithelial cell proliferation and tissue regeneration.

Comparison between fistulating involved and stricturing involved tissues revealed significant enrichment differences in undifferentiated epithelial populations, particularly stem cells and TA cells (**Fig. 4.7b-c**). In fistulating involved tissue, progenitor cells exhibited significant upregulation of genes involved in cytoskeletal organisation, tissue remodelling (*FHL2, EPB41L2*), and immune responses (e.g., *CD55*), suggesting active attempts at tissue repair and remodelling primarily within undifferentiated cells. High expression levels of both *CD55* and *FHL2* have been correlated with poor prognostic indicators in colon cancer (220, 221).

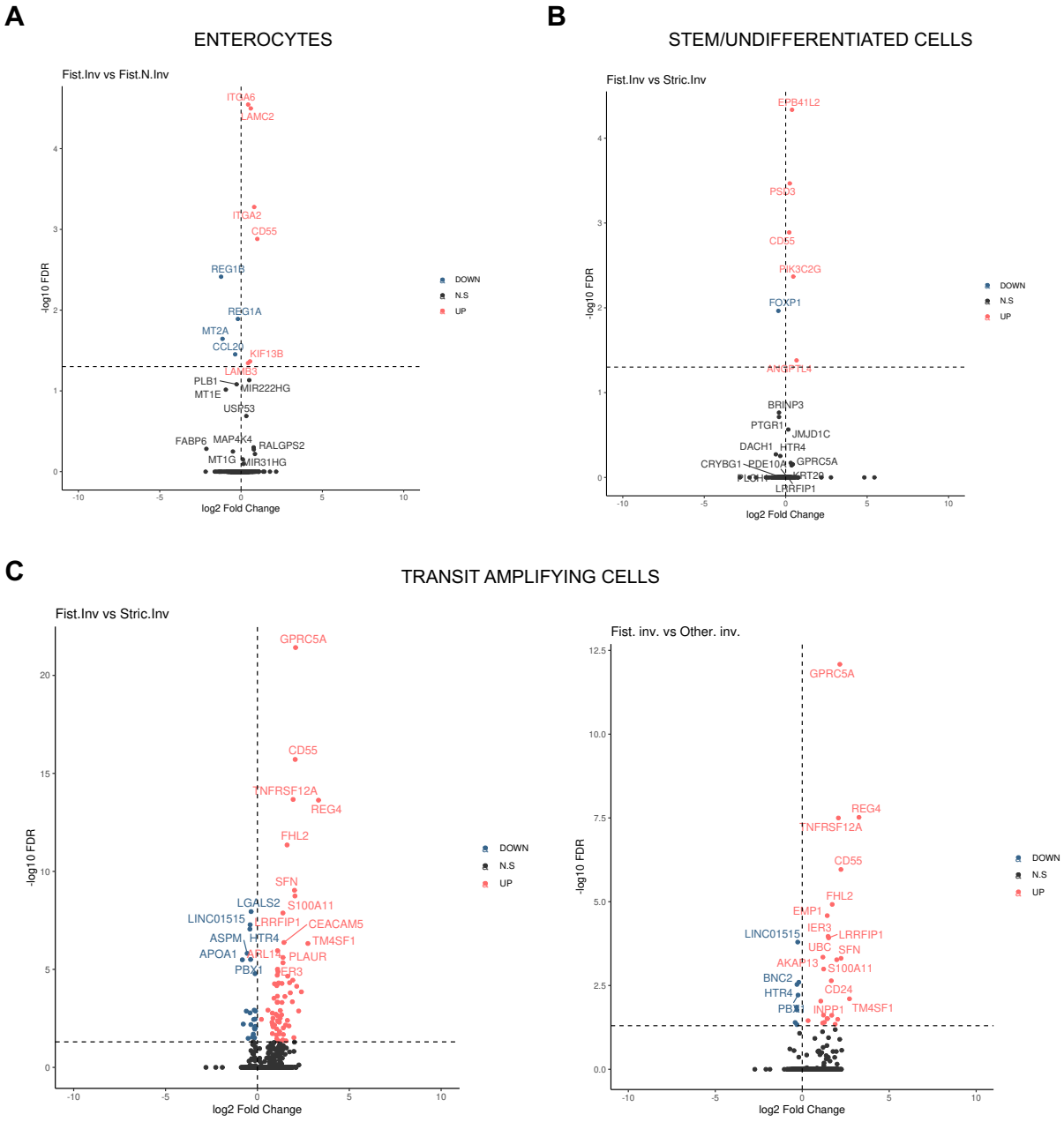


Figure 4.7: Differentially Expressed Genes in Selected Epithelial Cell Populations in Complex CD Tissue. (A) Volcano plot illustrating genes that exhibit differential expression by enterocytes from fistulating involved tissue (n=5; positive fold change expressed logarithmically and fistulating non-involved tissue (n=3; negative fold change expressed logarithmically). The black horizontal line represents a 1% false discovery rate. (B) Volcano plot showing genes that are differentially expressed by stem cells from fistulating involved tissue (n=5; positive fold change expressed logarithmically) compared to stricturing involved tissues (n=5). (C) Volcano plot visualising genes that are differentially expressed by TA cells from fistulating involved tissue (n=5; positive fold change (expressed logarithmically) and either stricturing involved tissues (n=5) or other involved tissue (n=9).

4.5 Stromal scRNA-seq data

4.5.1 Cell recoveries, QC process, cluster characterisation

A total of 64,112 high-quality singlets (CD45$^-$/EpCAM$^-$) were recovered from 34 samples, a significant increase attributed to optimised protocols, including the use of cell levitation technology discussed in **Chapter 3.4.1**. Prior to these refinements, the stromal fraction yielded far fewer cells (see **Table 4.1**). Additionally, the decision to load the stromal reaction twice significantly improved cell recovery.

Subsequent computational analysis revealed 18 distinct clusters using UMAP embedding visualisation (**Fig. 4.8**), with the majority delineated by characteristic gene expression signatures (168). We successfully captured the full complement of cell populations in deeper intestinal layers, including distinct subsets within the muscle tissue. All anticipated cell types were well-represented across experimental runs, facilitating robust data integration (**Appx. Fig. A.2**). The enhanced cell yields not only unveiled novel stromal subsets but also clarified smaller subpopulations identified in my pilot analyses.

4.5.2 Stromal signalling and remodelling in fistulating Crohn's tissue: Compositional analysis and gene expression profiling

The stromal compartment in fistulating CD tissue undergoes substantial compositional restructuring, indicative of a complex interplay of cellular dynamics within the diseased microenvironment. Notably, compared to healthy controls, we observed a significant expansion of a novel fibroblast subset (IL11$^+$EpCam$^-$) characterised by the expression of *IL11* as described below (see **4.5.3**). While statistical significance was not uniformly reached, trends revealed a notable increase in various cell types in fistulating CD tissue, including muscle cells, pericytes, myofibroblasts, stromal 4 cells, and endothelial cells, concomitant with a reduction in stromal 2 cells, stromal 1 cells, and lymphatic cells.

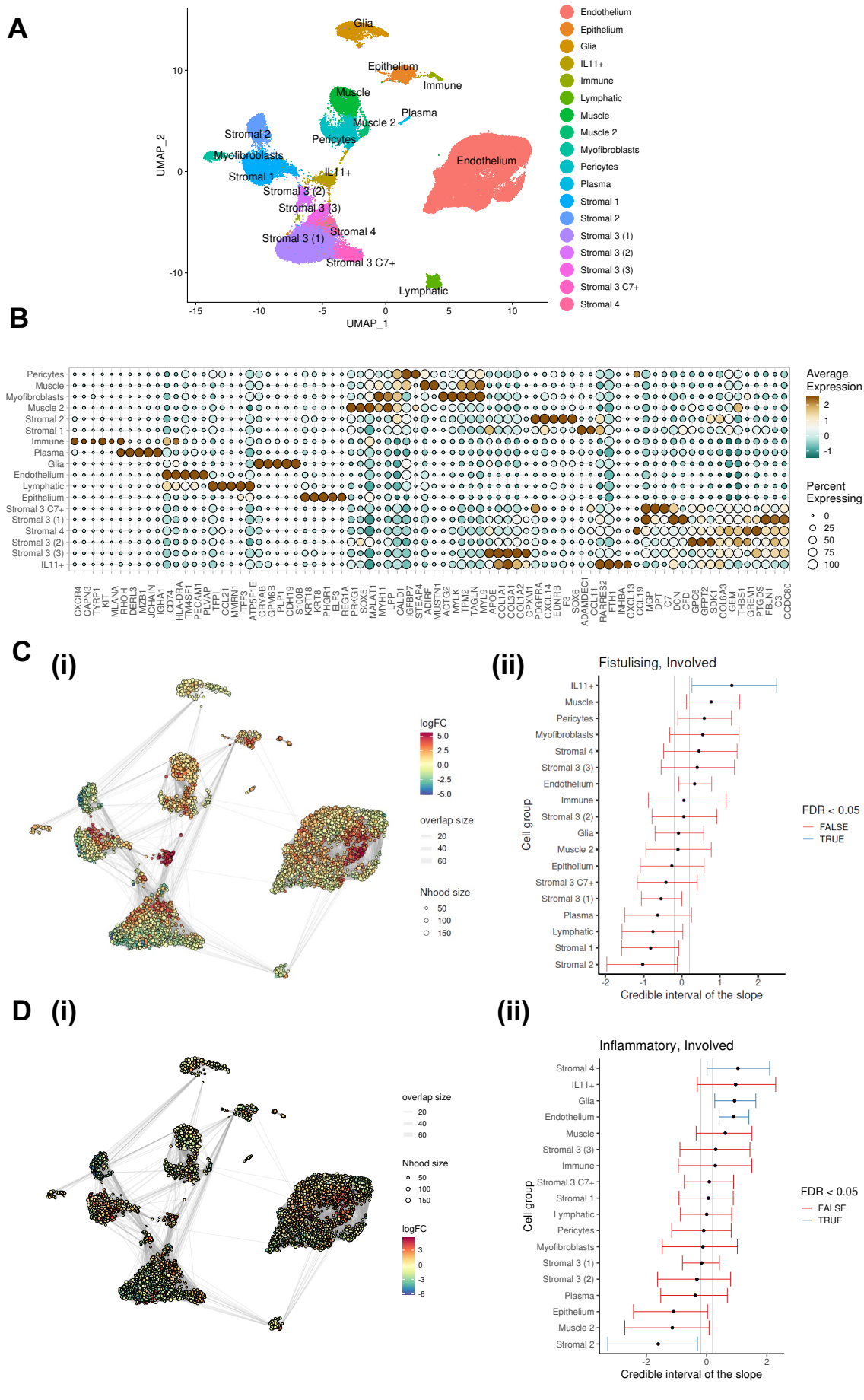

A
Glia
Epithelium
Immune
Muscle
Plasma
Muscle 2
Pericytes
Stromal 2
Myofibroblasts
Endothelium
Stromal 1
IL11+
Stromal 3 (2)
Stromal 3 (3)
Stromal 4
Stromal 3 (1)
Stromal 3 C7+
Lymphatic
UMAP_1
UMAP_2
Endothelium
Epithelium
Glia
IL11+
Immune
Lymphatic
Muscle
Muscle 2
Myofibroblasts
Pericytes
Plasma
Stromal 1
Stromal 2
Stromal 3 (1)
Stromal 3 (2)
Stromal 3 (3)
Stromal 3 C7+
Stromal 4
B
Average Expression
Percent Expressing
C
(i)
logFC
overlap size
Nhood size
(ii)
Fistulising, Involved
Cell group
Credible interval of the slope
FDR < 0.05
FALSE
TRUE
D
(i)
logFC
overlap size
Nhood size
(ii)
Inflammatory, Involved
Cell group
Credible interval of the slope
FDR < 0.05
FALSE
TRUE

Figure 4.8: Merged Stromal scRNA-seq Data (*previous page)*. (A) UMAP embedding visualising scRNA-seq clusters of cell populations obtained from stromal cell isolation. (B) Dot-plot heatmap displaying top stromal marker gene expression across the clusters identified in (A), with colour representing average expression levels and dot size indicating the percentage of cells in each cluster expressing the respective markers. (C) Differential abundance analysis comparing fistulating vs. healthy samples, visualised using a Milo network (i) showing log fold change by local neighbourhood and corresponding forest plot (ii). (D) Differential abundance analysis comparing inflammatory CD vs. healthy samples, visualised using a Milo network (i) and corresponding forest plot (ii). The false discovery rate (FDR) and significance threshold are set at $p < 0.05$. [FC, fold change; Nhood, neighbourhood]

Focused analysis on fibroblast subsets allowed for a deeper exploration of molecular alterations underlying fistulating CD pathology. Differential expression analysis of genes within fistulating fibroblasts, compared to their healthy counterparts, unveiled a distinct pattern of upregulated genes. Notably, genes implicated in tissue remodelling (*TWIST1, ABL2, IL11, ARL4A*), immune responses (*CXCL3, IL11*) and repair mechanisms (*CXCL3, TWIST1*) exhibited heightened expression levels within the intestinal stroma of fistulating CD patients (222-224).

Specifically, *TWIST1*, a transcription factor crucial for development, plays a pivotal role in mesenchymal cell differentiation and tissue remodelling processes such as EMT (222). Its aberrant expression is associated with tumourigenesis and has been implicated in modulating immune cell infiltration, cytokine production, and inflammatory signalling pathways in the context of IBD (225).

Similarly, the cytokine IL11 exerts diverse functions in the intestine, including maintaining epithelial integrity, promoting tissue repair, and modulating mucosal immune responses. Recognised as a pro-fibrotic marker, IL11 is predominantly expressed by activated fibroblasts (210), exerting significant effects on fibroblasts themselves by promoting the production of ECM proteins (226). Dysregulation of IL11 signalling has been implicated in CD pathogenesis, with single-cell data confirming the presence of $IL11^{+}$ inflammation activated fibroblasts within the diseased microenvironment (169).

In contrast, a noticeable downregulation of genes implicated in maintaining ECM integrity and immune regulation, such as *PCOLCE2*, *MFAP5*, and *TRAC*, was evident in the stromal tissue of fistulating CD. This decrease in expression levels suggests potential disruptions in ECM dynamics, stromal architecture, and an altered immune microenvironment (227-228). Specifically, MFAP5,

which has been observed to be downregulated in the stromal tissue of colorectal carcinoma, raises questions regarding its role in the context of the tumour microenvironment or wound healing pathways. However, the precise mechanisms remain unaddressed within the literature (227). Considering its reduced expression in fistulating tissue, it is plausible that aberrations in wound healing processes may contribute to the observed changes in fistulae.

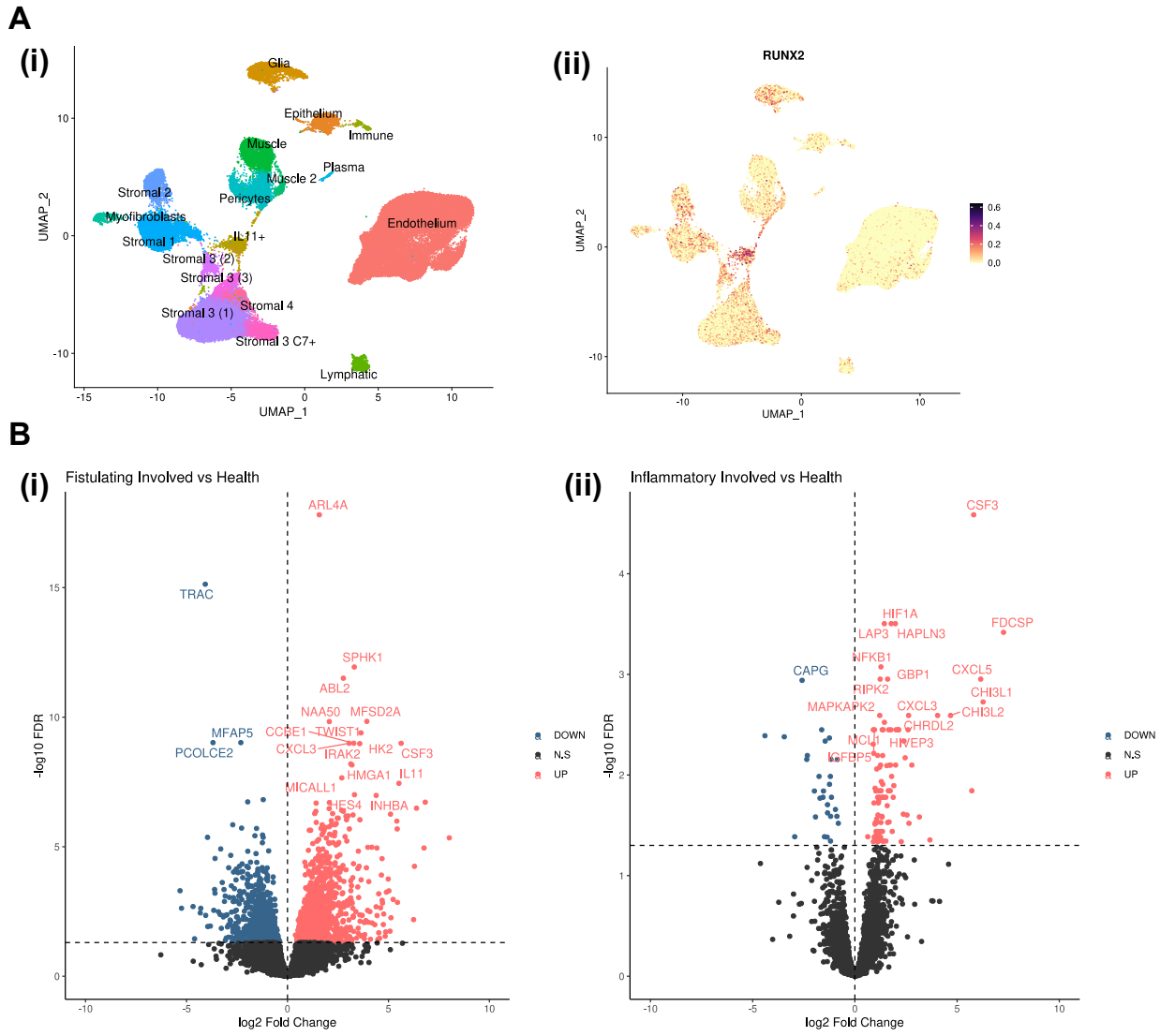


Figure 4.9: Differentially Expressed Genes in Stromal Cell Populations in Fistulating Crohn's Disease Tissue (*previous page*). (A) UMAP embedding visualising scRNA-seq clusters of stromal cell populations i), with a UMAP embedding overlay showing the expression of *RUNX2* across these clusters (ii). (B) Volcano plot illustrating differentially expressed genes in stromal cells from fistulating CD-involved tissue compared to healthy tissue. Positive fold changes (logarithmically expressed) indicate higher expression in fistulating CD, while negative fold changes (logarithmically expressed) represent higher expression in healthy tissue. The black horizontal line indicates the significance threshold, based on a 1% false discovery rate (i). Volcano plot showing differentially expressed genes in stromal cells from inflammatory CD-involved tissue compared to healthy tissue (ii).

4.5.3 Characterisation of a novel fistula associated stromal subset

Cluster analysis revealed a stromal counterpart to the previously identified IL11^{+}EpCam^{+} cluster in the epithelial dataset. This stromal (CD45^{-}/EpCam^{-}) cluster is principally characterised by the expression of *IL11* amongst others (**Fig. 4.10b**). After the optimisation of the single-cell workflow, this novel stromal cluster exhibited clearer delineation and comprised a larger cell population compared to its epithelial counterpart. Notably, approximately 10% of the total stromal compartment was constituted by this cluster.

Remarkably, this cluster exhibited a distinct association with fistula samples, and in smaller proportions in other involved CD samples, while being absent in healthy controls reflecting relative CD specificity. Notably, statistical analysis revealed a significant enrichment of this cluster exclusively in fistulating CD samples (**Fig. 4.10a**).

Molecular characterisation of this cluster revealed a repertoire of genes indicative of its functional role. Notably, top markers implicated roles in ECM remodelling and fibrotic responses (*COL1A1, COL3A1, COL1A2*), metabolism (*FTH1*), and immune activation (*CXCL13*). Additionally, heightened expression levels of transcriptional regulators associated with stromal remodelling and EMT, such as *TWIST1*, *RUNX2*, and *SNAI2* (229), were observed, suggesting involvement in wound healing processes gone awry.

GO enrichment analysis further elucidated pathways upregulated in the IL11^{+} stromal cluster, predominantly related to ECM organisation, regulation of EMT and tissue and leucocyte migration (**Fig. 4.10c**). Distinctive features of this cluster included significantly upregulated expression of pro-fibrotic signalling effector molecules such as IL11, MMPs (e.g., MMP3, 9) and CXCL13.

Notably, examination of the expression pattern of this cluster suggested potential lineage relationships with three previously identified fibroblast subsets (stromal 2, 3, and 4) characterised in our research group (168). Stromal 2 (S2) was distinguished by its location proximal to epithelial cell crypt niches and is characterised by the expression of transcription factor *SOX6*, morphogens (*WNT5A, WNT5B, BMP2, BMP5*) and the secreted Wnt antagonist *FRZB*. Together with other preferentially induced genes crucial to tissue repair, such as *POSTN*, *S2* has been shown to contribute significantly to epithelial cell renewal (230) (168). Stromal 3 (S3), identified as submucosal structural cells, exhibits enrichment for processes associated with ECM organisation, contributing notably to vascular support. Characterised by elevated expression levels of key ECM proteins including C7, OGN, and ASPN, S3 plays a pivotal role in maintaining structural integrity within the submucosal environment. Finally, stromal 4 (S4) displayed expansion in UC and was enriched for pro-inflammatory fibroblastic reticular cell (FRC)-associated genes, lymphocyte trafficking cytokines (CCL19 and CCL21), and T-cell co-stimulatory TNF-superfamily ligands (TNFSF14/LIGHT), suggesting a role in immune response mobilisation in intestinal inflammation (**Fig. 4.10b, d**).

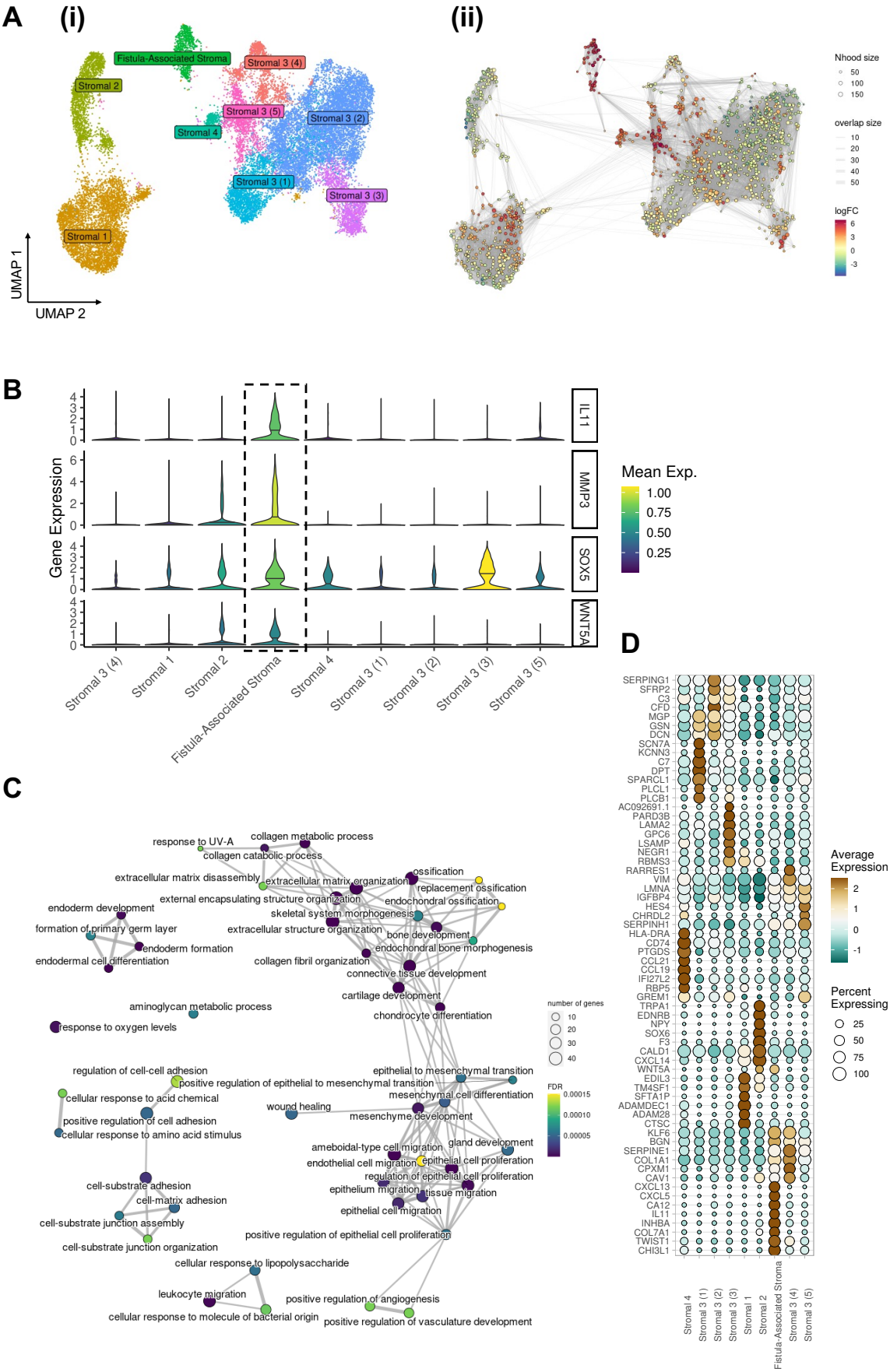

A
(i)
Fistula-Associated Stroma
Stromal 3 (4)
Stromal 2
Stromal 3 (5)
Stromal 4
Stromal 3 (2)
Stromal 3 (1)
Stromal 3 (3)
Stromal 1
UMAP 1
UMAP 2
(ii)
Nhood size
50
100
150
overlap size
10
20
30
40
50
logFC
6
3
0
-3
B
Gene Expression
IL11
MMP3
SOX5
WNT5A
Mean Exp.
1.00
0.75
0.50
0.25
Stromal 3 (4)
Stromal 1
Stromal 2
Fistula-Associated Stroma
Stromal 4
Stromal 3 (1)
Stromal 3 (2)
Stromal 3 (3)
Stromal 3 (5)
C
response to UV-A
collagen metabolic process
collagen catabolic process
extracellular matrix disassembly
extracellular matrix organization
ossification
replacement ossification
external encapsulating structure organization
endochondral ossification
endoderm development
skeletal system morphogenesis
formation of primary germ layer
extracellular structure organization
bone development
endoderm formation
endochondral bone morphogenesis
endodermal cell differentiation
collagen fibril organization
connective tissue development
cartilage development
aminoglycan metabolic process
chondrocyte differentiation
response to oxygen levels
regulation of cell-cell adhesion
epithelial to mesenchymal transition
positive regulation of epithelial to mesenchymal transition
cellular response to acid chemical
mesenchymal cell differentiation
wound healing
positive regulation of cell adhesion
mesenchyme development
cellular response to amino acid stimulus
ameboidal-type cell migration
gland development
endothelial cell migration
epithelial cell proliferation
regulation of epithelial cell proliferation
cell-substrate adhesion
epithelium migration
tissue migration
cell-matrix adhesion
epithelial cell migration
cell-substrate junction assembly
positive regulation of epithelial cell proliferation
cell-substrate junction organization
cellular response to lipopolysaccharide
leukocyte migration
positive regulation of angiogenesis
cellular response to molecule of bacterial origin
positive regulation of vasculature development
number of genes
10
20
30
40
FDR
0.00015
0.00010
0.00005
D
SERPING1
SFRP2
C3
CFD
MGP
GSN
DCN
SCN7A
KCNN3
C7
DPT
SPARCL1
PLCL1
PLCB1
AC092691.1
PARD3B
LAMA2
GPC6
LSAMP
NEGR1
RBMS3
RARRES1
VIM
LMNA
IGFBP4
HES4
CHRDL2
SERPINH1
HLA-DRA
CD74
PTGDS
CCL21
CCL19
IFI27L2
RBP5
GREM1
TRPA1
EDNRB
NPY
SOX6
F3
CALD1
CXCL14
WNT5A
EDIL3
TM4SF1
SFTA1P
ADAMDEC1
ADAM28
CTSC
KLF6
BGN
SERPINE1
COL1A1
CPXM1
CAV1
CXCL13
CXCL5
CA12
IL11
INHBA
COL7A1
TWIST1
CHI3L1
Stromal 4
Stromal 3 (1)
Stromal 3 (2)
Stromal 3 (3)
Stromal 1
Stromal 2
Fistula-Associated Stroma
Stromal 3 (4)
Stromal 3 (5)
Average Expression
2
1
0
-1
Percent Expressing
25
50
75
100

Figure 4.10: Characterisation of a Novel Fistula-Associated Stromal Population (*previous page*)

(A) UMAP embedding visualising scRNA-seq clusters of fibroblast populations derived from isolated stromal cells (i). Differential abundance analysis of fibroblast populations comparing fistulating CD involved samples to healthy controls, visualised using a Milo network showing log fold change by local neighbourhood (ii).

(B) Violin plot characterising average expression of *IL11*, *MMP3*, *SOX5* and *WNT5A* across fibroblast populations. The colour gradient represents the average expression levels, with lighter colours indicating higher gene expression.

(C) Network plot displaying gene ontology analysis of marker genes in the IL11+/Fistula-Associated stromal cluster. Enrichment significance was assessed using a hypergeometric test and visualised with emapplot.

(D) Dot-plot heatmap showing top stromal marker gene expression across fibroblast clusters identified in (A(i)). Colour intensity represents average expression level, while the relative size of each dot indicates the percentage of cells within each cluster expressing the respective markers.

The IL11$^+$ fibroblast cluster shared expression of inducible morphogens with S2, such as *WNT5A* and *BMP2*, indicating potential involvement in epithelial differentiation and regenerative morphogen pathways (**Fig. 4.10b**). Additionally, unique to this cluster was the significant enrichment of Wnt ligand WNT2 and its receptor FZD1, which have been implicated in ISC maintenance, proliferation, and fibroblast activation (231). S2 marker, POSTN, typically expressed in the colon within the intestine (179), was notably absent in the S2 cluster derived from TI samples in my data. However, *POSTN* expression was observed in the IL11$^+$ stromal cluster identified from the same TI samples. Within the intestinal stromal context, POSTN plays a crucial role in regulating ECM remodelling and epithelial-mesenchymal crosstalk, thereby influencing tissue repair and barrier function (232). The aberrant expression of *POSTN* within the IL11$^+$ stromal cluster is further indicative of a pathological process. How this IL11$^+$ stromal cell state arises will be interesting to infer via a transcriptional regulatory network analysis coupled with pseudotemporal ordering.

The expression patterns of transcription factors *PRRX1* (233) and *OSR2* (234), known for their involvement in mesenchymal development and tissue patterning, are shared within the S3 subset and IL11$^+$ cluster. Additionally, the presence of complements (C3) and chemokines (CXCL13, CCL19) reflecting immune regulatory roles observed in the IL11$^+$ cluster is also evident in S4. Further understanding of the relationships of the different stromal states observed in this analysis will be assisted with transcriptional regulatory network analysis and by spatial and intercellular signalling analysis.

For the purposes of my thesis, I will now denote this pathological stromal subset (IL11^{+}) as 'Fistula-Associated Stroma' (FAS).

To further investigate the FAS subset, gene regulatory network (GRN) analysis was conducted to study the regulation of these stromal cells (**Fig. 4.11**). The analysis revealed several upregulated and/or uniquely expressed transcription factors (TF) and TF regulons. Specifically, *TWIST1* and *TWIST2*, previously mentioned (**Chapter 4.5.2**), are known for their roles in regulating patterning and morphology during development. Additional TFs associated with morphogenesis were identified, including *PRRX1* and *PRRX2*, which are involved in limb development, with mutations leading to craniofacial abnormalities (235), and *RUNX2*, a regulator of skeletal morphogenesis (229). Furthermore, GRN analysis indicates that the upregulation of *ELK1* and *STAT1* likely drives aberrant *WNT* signalling in these cells, promoting proliferation. Taken together, these findings indicate that FAS cells play a role in the tissue remodelling processes that lead to fistula formation.

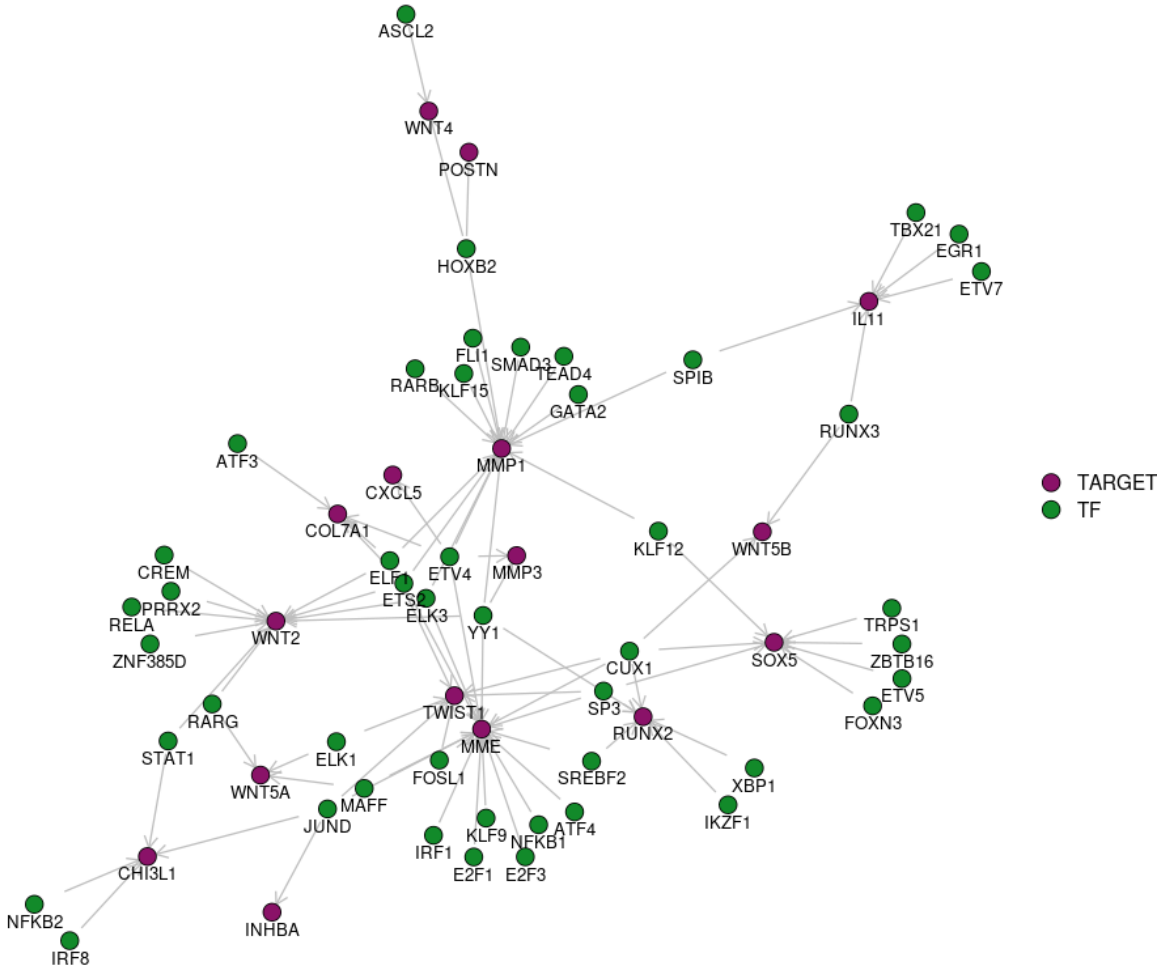


Figure 4.11: Gene regulatory network of Fistula-associated stromal (FAS) cells. Gene regulatory network analysis of Fistula-Associated Stromal (FAS) cells derived from fistulating Crohn's disease tissue. The network visualisation highlights the complex interactions between transcription factors (TFs) and their downstream target genes within the FAS cell population. TFs are represented by green nodes, and target genes are depicted by purple nodes. Arrows indicate the direction of regulatory influence. The network was constructed using SCENIC regulons and visualised with the Fruchterman-Reingold layout algorithm in the igraph package.

4.5.4 Characterisation of expanded CD-specific subpopulation of pericytes

Following optimisation of the stromal workflow, a distinct subpopulation of pericytes emerged prominently. Although not reaching statistical significance, this pericyte population exhibited expansion in involved CD samples, with the highest cell abundance observed in fistulating involved cases. Characterised by a signature reminiscent of S4 cells, these pericytes notably expressed CCL19 and CCL21 (**Appx. Fig. A.3**), suggesting their involvement in modulating immune cell behaviour and function within the intestinal microenvironment.

Pericytes, traditionally recognised as integral components of the intestinal microvasculature, extend their role beyond vascular support to encompass immune regulation, influencing processes such as inflammation and tissue repair (236, 237).

Analysis of preferentially induced genes in fistulating pericytes compared to healthy controls revealed implications in pathways related to inflammatory signalling and immune regulation within the intestinal stroma. Notably, the production of chemokines such as CXCL12 and CCL19 suggests a role in orchestrating immune cell recruitment and activation. Furthermore, these pericytes may contribute to lipid metabolism and transport processes, as indicated by the expression of genes such as *APOE* (238) and *FKBP5* (239). Additionally, their interaction with the ECM is implied through the regulation of ECM composition and remodelling, as evidenced by the expression of genes including *PCOLCE* and *THBS1* (240).

4.6 Immune scRNA-seq data

4.6.1 Cell recoveries, QC process, cluster annotations

The immune dataset yielded a total of 27,959 cells for analysis, from which seven broad clusters were identified through UMAP embedding visualisation. Further sub-clustering of the T-cell population enabled the identification of seventeen distinct clusters representing various T-cell subsets, including naïve (CD4 naïve, CD8 naïve), activated (cycling), and effector T-cell populations (e.g., FGFB2, GZMK^{+}) using characteristic gene expression signatures (170) (**Fig. 4.12**).

Across the five runs, there was generally good replication in terms of the recovered cell populations and the overall composition of the samples. Notably, a significant proportion of the epithelial fraction comprised immune cells, predominantly T-cells identified as IELs. Both the T-cells from the immune

fraction ($CD45^+/EpCam^-$) and the epithelial fraction ($EpCam^+$) were analysed to understand their composition and diversity.

Comparative analysis revealed largely comparable compositions between the immune and epithelial T-cell populations, with minor differences observed (**Fig. 4.12e**). Specifically, the immune fraction exhibited slightly higher proportions of $GZMK^+$ effectors, CD4 naïve, and regulatory T-cell (Treg) populations compared to the epithelial fraction. However, common markers associated with IELs, such as CD160 and KLRC2, were observed across both fractions, suggesting consistent representation of intraepithelial T-cell populations.

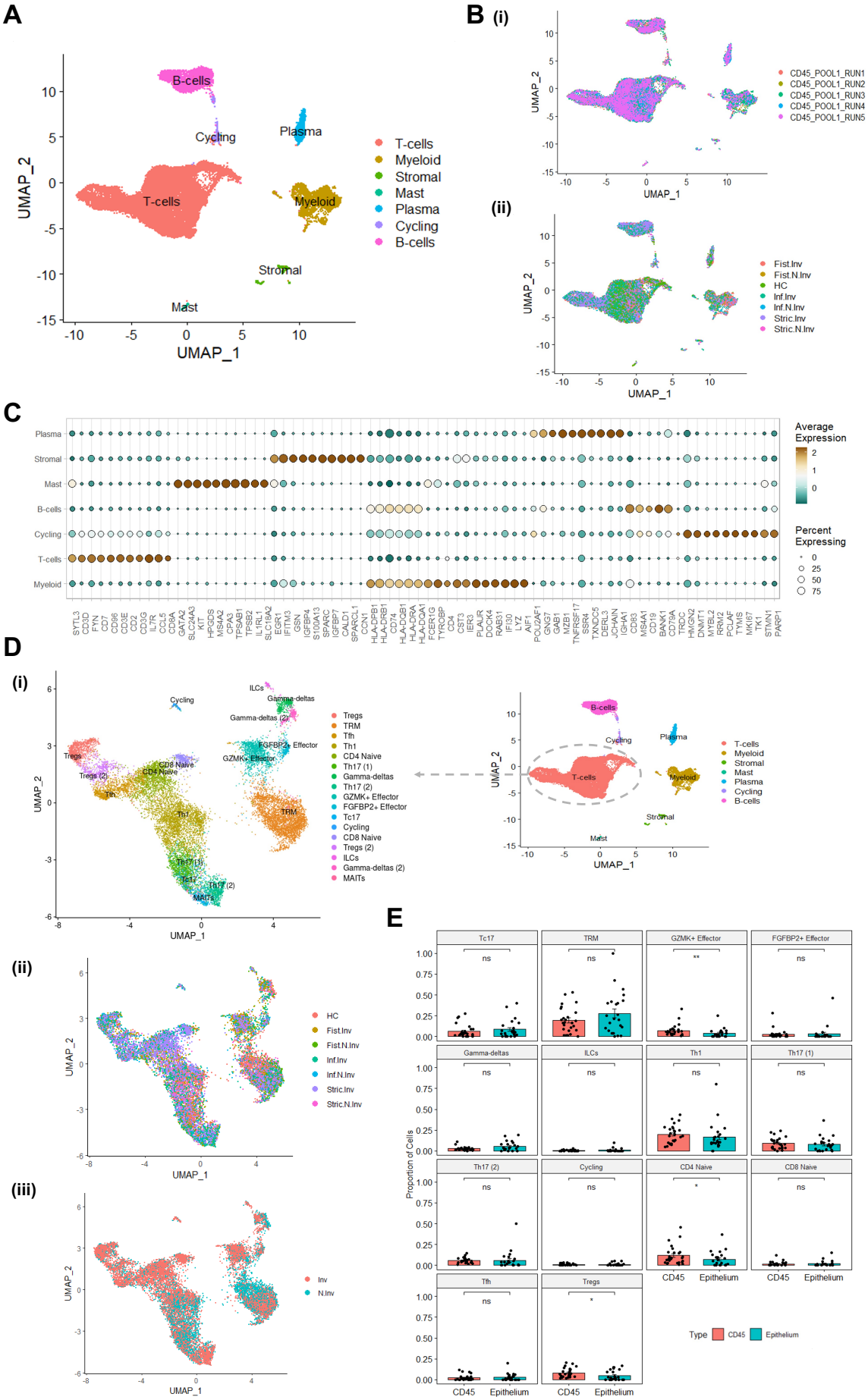

A
B-cells
Cycling
Plasma
T-cells
Myeloid
Stromal
Mast
UMAP_2
UMAP_1
T-cells
Myeloid
Stromal
Plasma
Cycling
B-cells
B (i)
CD45_POOL1_RUN1
CD45_POOL1_RUN2
CD45_POOL1_RUN3
CD45_POOL1_RUN4
CD45_POOL1_RUN5
(ii)
Fist.Inv
Fist.N.Inv
HC
Inf.Inv
Inf.N.Inv
Stric.Inv
Stric.N.Inv
C
Plasma
Stromal
Mast
B-cells
Cycling
T-cells
Myeloid
Average Expression
Percent Expressing
D
(i)
Tregs
TRM
Tfh
Th1
CD4 Naive
Th17 (1)
Gamma-deltas
Th17 (2)
GZMK+ Effector
FGFBP2+ Effector
Tc17
Cycling
CD8 Naive
Tregs (2)
ILCs
Gamma-deltas (2)
MAITs
(ii)
(iii)
Inv
N.Inv
E
Tc17
TRM
GZMK+ Effector
FGFBP2+ Effector
Gamma-deltas
ILCs
Th1
Th17 (1)
Th17 (2)
Cycling
CD4 Naive
CD8 Naive
Tfh
Tregs
Proportion of Cells
CD45
Epithelium
Type

Figure 4.12. Merged Immune scRNA-seq Data (*previous page*). (A) UMAP embedding visualising the clustering of immune cells (CD45$^+$) based on scRNA-seq data. (B) UMAP embeddings further stratify immune cell clusters by single-cell run (i) and by condition (ii). (C) Dot-plot heatmap illustrating the expression of top immune cell marker genes across the identified clusters in (A). The colour intensity reflects the average expression level of each marker within the clusters, while the size of the dots indicates the proportion of cells within each cluster expressing the respective markers. (D) UMAP embedding focusing on the T-cell subsets within the broader immune cell population. This panel visualises T-cell clusters (i), further stratified by condition (ii) and involvement status (iii). (E) Bar plot comparing the proportion of T-cell subsets isolated from the lamina propria (CD45$^+$ cells) versus intraepithelial lymphocytes isolated from the epithelial compartment. [Inv, involved; N.Inv, non-involved]

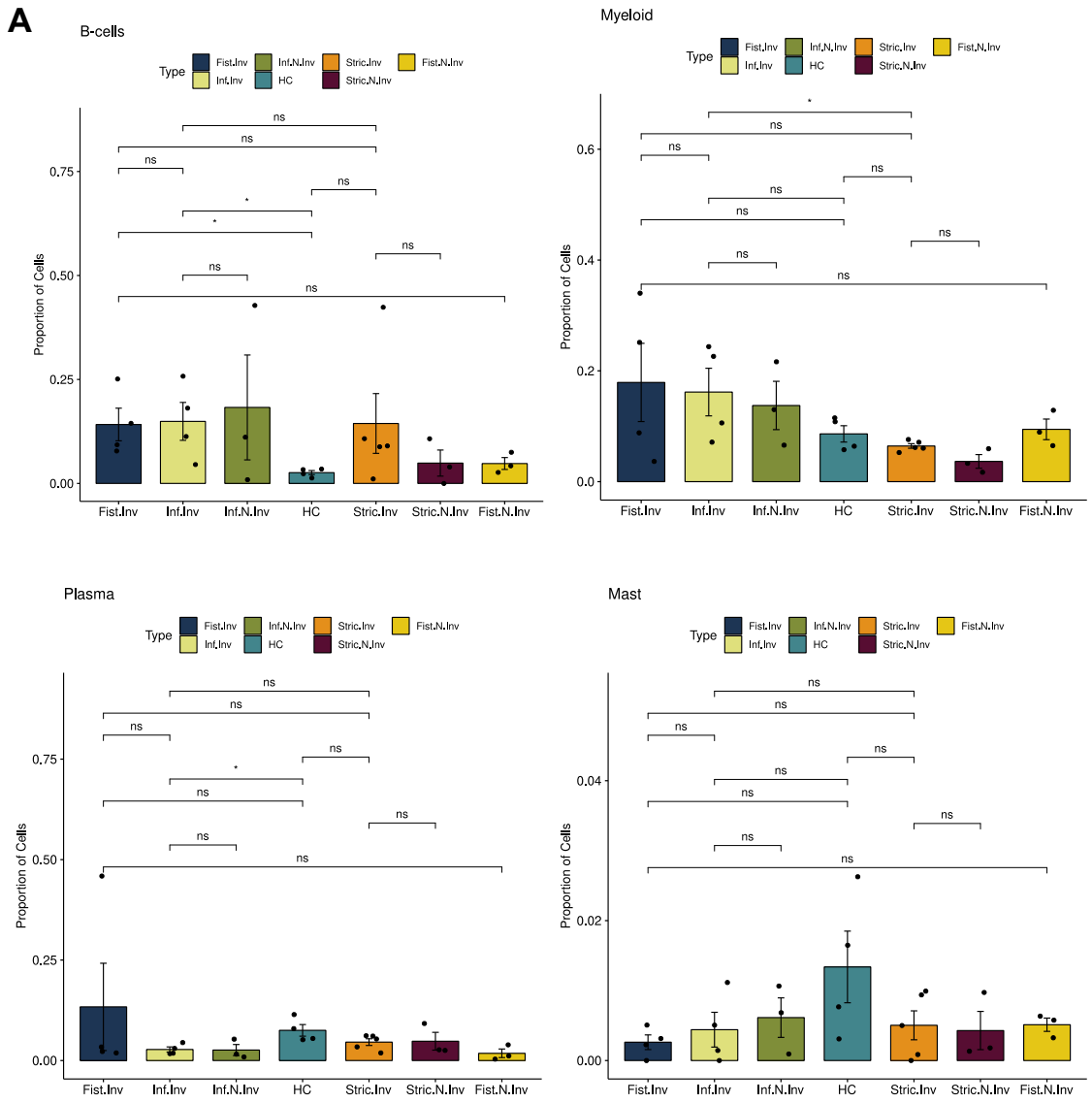
4.6.2 Immune cell composition in fistulating CD

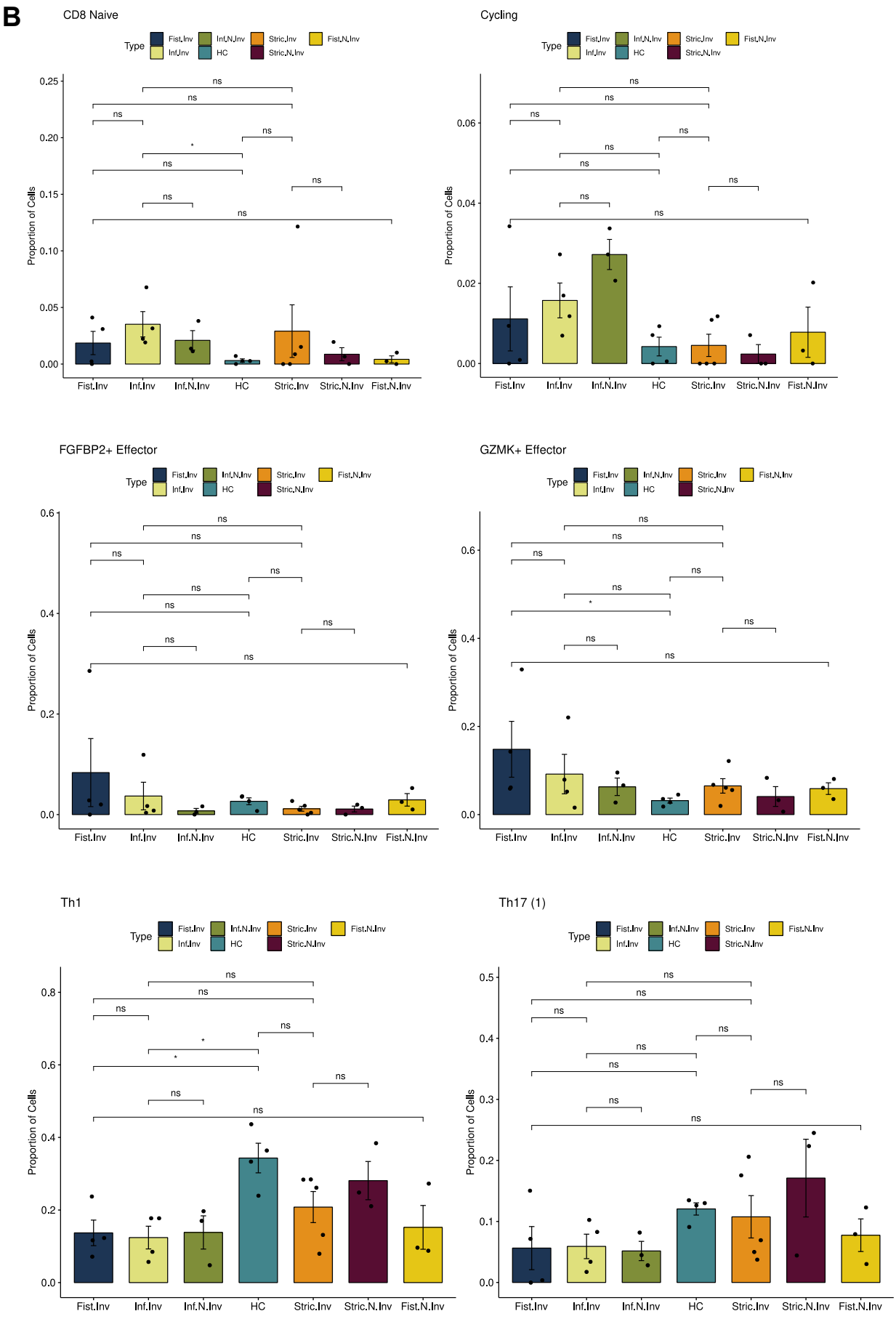
In the immune compartment of fistulating involved CD tissue, significant differences in cellular composition were noted compared to controls. Specifically, a marked increase in B-cells was observed in fistulating involved samples compared to healthy controls, indicating a potential role for B-cell-mediated immune responses in the disease process (**Fig. 4.13a**). Additionally, although not statistically significant, there was an expansion in the myeloid compartment in fistulating involved samples compared to controls (**Fig. 4.13a**), suggesting potential alterations in innate immune cell populations in diseased tissue. Differentially upregulated genes in the myeloid compartment of fistulating involved samples compared to healthy tissue were consistent with a more activated myeloid state (*STAT1*), inflammation (*STAT1, TYMP*) and tissue remodelling (*SDC2, THBS1, CTSB*) (241, 240) (**Fig. 4.14**).

Furthermore, variation was observed within T-cell subsets across different disease conditions. Overall, diseased tissues exhibited a higher abundance of effector, Tc17, Th17(2), Tregs and cycling T-cells compared to healthy controls. Specifically, in fistulating involved tissue, there was a notable expansion of T-follicular cells (Tfh) (**Fig. 4.13b**).

Additionally, a significant enrichment of GZMK$^+$ effector T-cells was observed in fistulating involved tissue compared to healthy controls, along with FGFBP2$^+$ effector T-cells, although the latter did not reach statistical significance. Remarkably, a significant reduction in resident memory T-cells (TRM) and Th1 cells was evident in fistulating involved tissue compared to healthy controls (**Fig. 4.13b**).

Additionally, Th17(2) cells appeared to be more abundant in the epithelium of fistulating involved samples, indicating a potential role in mucosal inflammation and barrier dysfunction. Regulatory T-cells, both Treg and Treg(2) subsets were low in healthy controls, with Tregs prominently present in the epithelium of fistulating samples and Treg(2) more prominent in the stroma of fistulating samples.




B
CD8 Naive
Cycling
FGFBP2+ Effector
GZMK+ Effector
Th1
Th17 (1)
Type
Fist.Inv
Inf.Inv
Inf.N.Inv
HC
Stric.Inv
Stric.N.Inv
Fist.N.Inv
Proportion of Cells
ns
*
Fist.Inv
Inf.Inv
Inf.N.Inv
HC
Stric.Inv
Stric.N.Inv
Fist.N.Inv

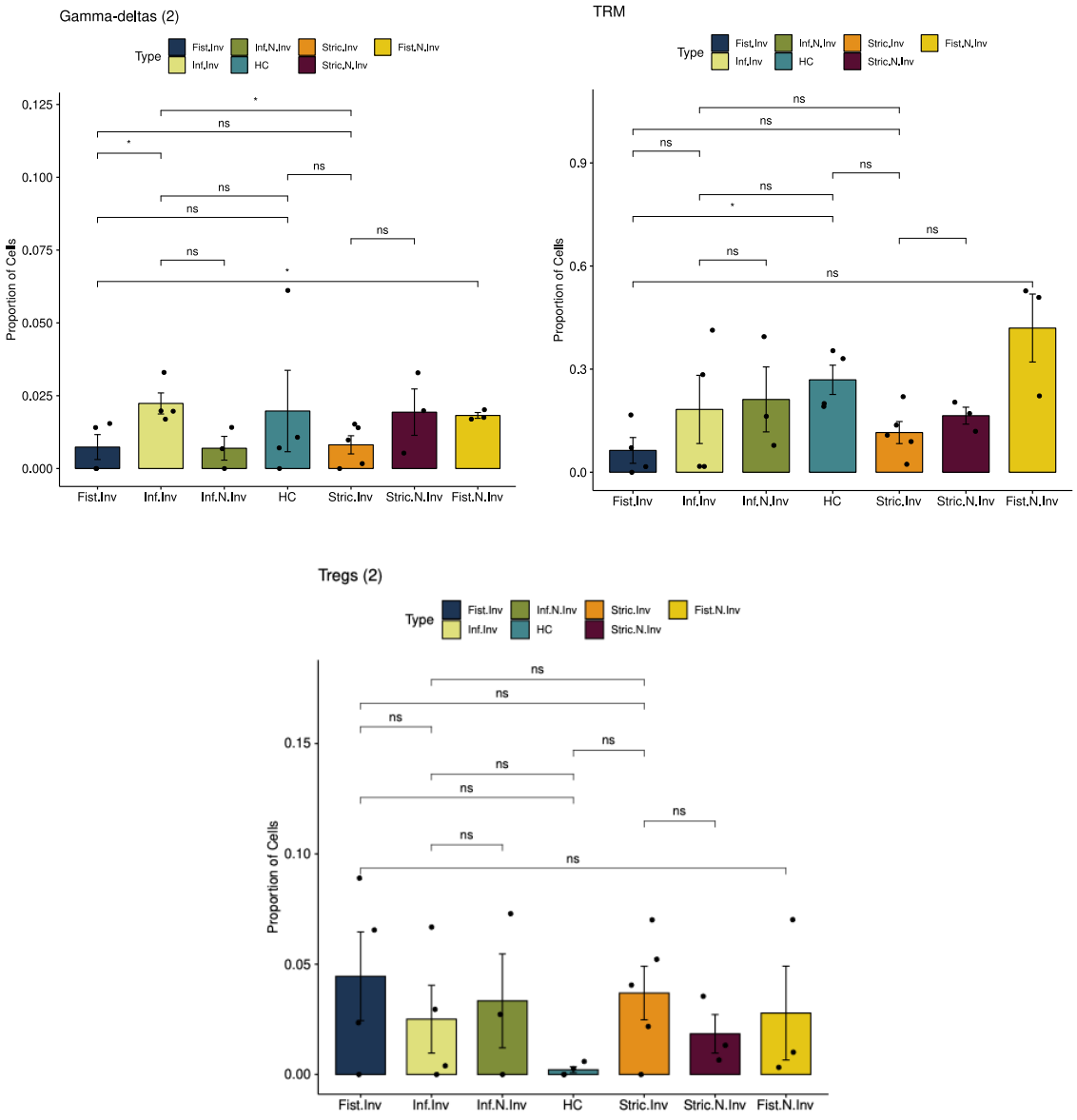


Figure 4.13: Cellular Composition Differences in Immune Isolation Reactions (*cont. previous pages).* Immune scRNA-Seq data showing: (A) Cellular composition differences in immune isolation reactions, per cluster – B-cells, myeloid, mast and plasma cells. (B) Cellular composition differences of T-cells in immune cell isolation reactions. Statistical significance is set at $p<0.05$ using Wilcox test [GZMK, granzyme K; Fist, Fistulating; HC, Healthy control; Inf, Inflammatory; Inv, involved; N inv, Non-involved; Stri, Stricturing; TAs, transit amplifying cells; TRM, tissue resident memory]

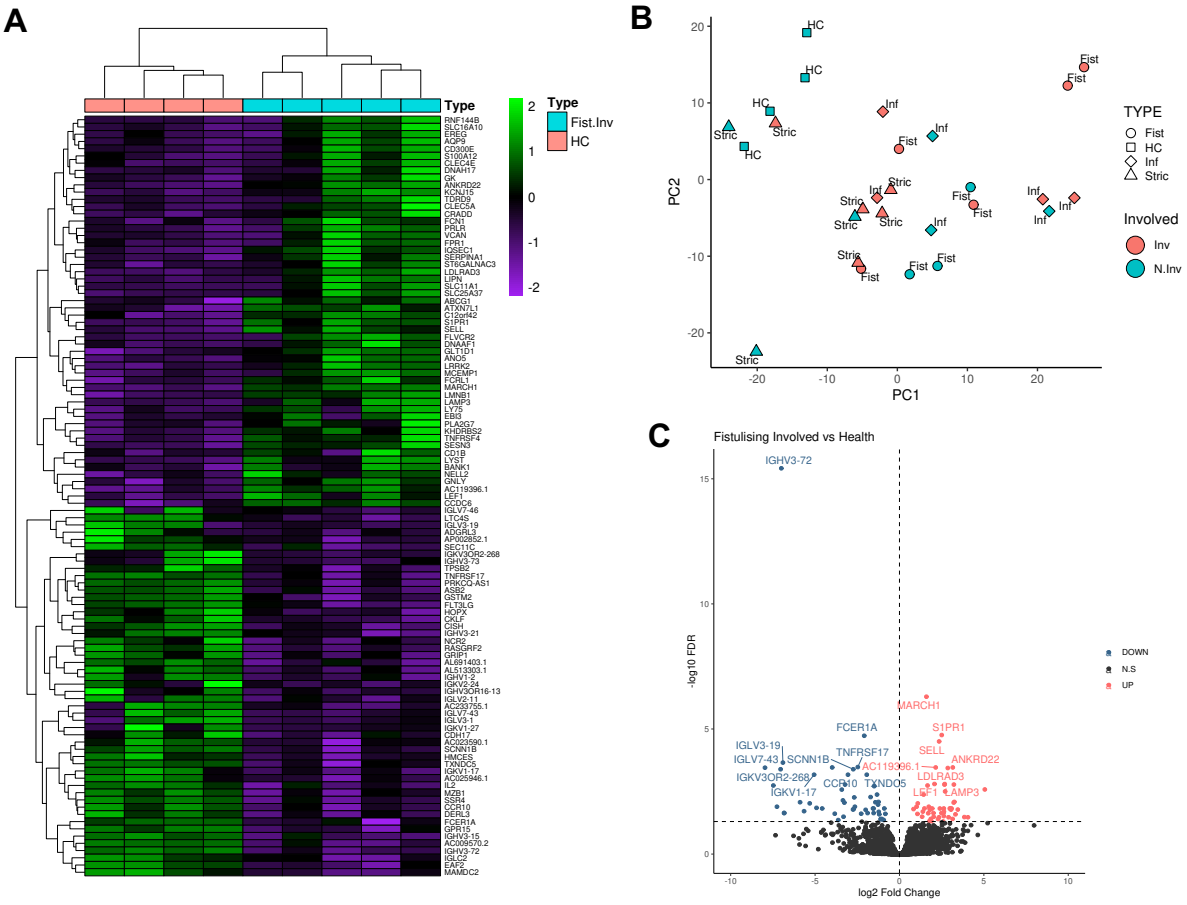


Figure 4.14: Differentially Expressed Genes from Immune Cells Enriched in Fistulating-Involved Tissue Compared to Healthy Control Tissue. (A) Heatmap visualising relative expression of selected differentially expressed genes, identified using the Wald test and corrected for false discovery rate by the Benjamini-Hochberg method, in fistulating-involved tissue (blue; n=5 per group) and healthy control tissue (pink; n=4 per group). (B) Principal component analysis (PCA) based on condition type and involvement status (CD45+ cells only). (C) Volcano plot displaying immune cell-related genes that exhibit differential expression between fistulating-involved tissue (positive fold change, log scale) and healthy control tissue (negative fold change, log scale). The black horizontal line represents a 1% false discovery rate, with the most significant genes highlighted.

4.6.3 T-cell proportional abundance, dynamics and clonality in fistulating CD

TCR libraries were generated from immune cells isolated from both the stromal fraction and IELs of the epithelial fraction. Notably, our analysis revealed distinct patterns of clonal expansion between the two fractions, with epithelial samples exhibiting fewer recovered clonotypes, particularly evident in samples from fistulating and inflammatory involved CD samples (**Fig. 4.15a(i)**). Consequently,

this discrepancy in TCR recovery in the epithelial fraction may pose a potential limitation, leading to underpowered analyses.

Despite these differences in TCR recovery, the extent of clonal expansion within stromal compartments remained comparable across all samples. Interestingly, upon comparing the clones shared between stromal and epithelial compartments, I observed a notable decrease in TRM cell populations in both stromal and epithelial compartments of involved samples, except for stricturing samples. Conversely, healthy and non-involved tissues exhibited many expanded populations in the stroma adopting Th1/TRM phenotypes in the epithelium (**Appx. Fig. A.5**).

Furthermore, the clonal expansion of epithelium-shared stromal clonotypes was particularly prominent in fistulating samples, where gamma-delta (2), Th1, and TRM clones in the epithelium often overlapped with expanded clones in the stromal compartments (**Fig. 4.15c(i)**). Additionally, cycling in the epithelium displayed shared dynamics with GZMK$^+$ cells in the stroma in fistulating tissue. To analyse clonal dynamics, directional overlap and Morisita's index were employed. In healthy tissue, high CD8 naïve, effector and cycling clonotypes were generally shared with TRMs, along with some sharing with Th1, Tc17, and Th17 populations (**Fig. 4.15b(i)**).

In fistulating involved tissue, the greatest GZMK$^+$ effector acquisitions came from FGFBP2$^+$, gamma-deltas, Th17 and TRMs of fistula uninvolved tissue. Additionally, Th17(1)s transitioned to a more cytotoxic Tc17 phenotype in involved tissue. GZMK$^+$ cells occupied a central position in the dynamics, sharing many clonotypes with multiple T-cell populations (**Fig. 4.15b(ii), c(iii)**).

Comparatively, uninvolved fistulating tissue exhibited more TRM incomings from different populations and fewer GZMK$^+$ cells at the centre of dynamics compared to fistulating involved tissue. These findings highlight the intricate T-cell dynamics and clonal relationships within the stromal and epithelial compartments of fistulating CD tissue (**Fig. 4.15b(i-ii)**).

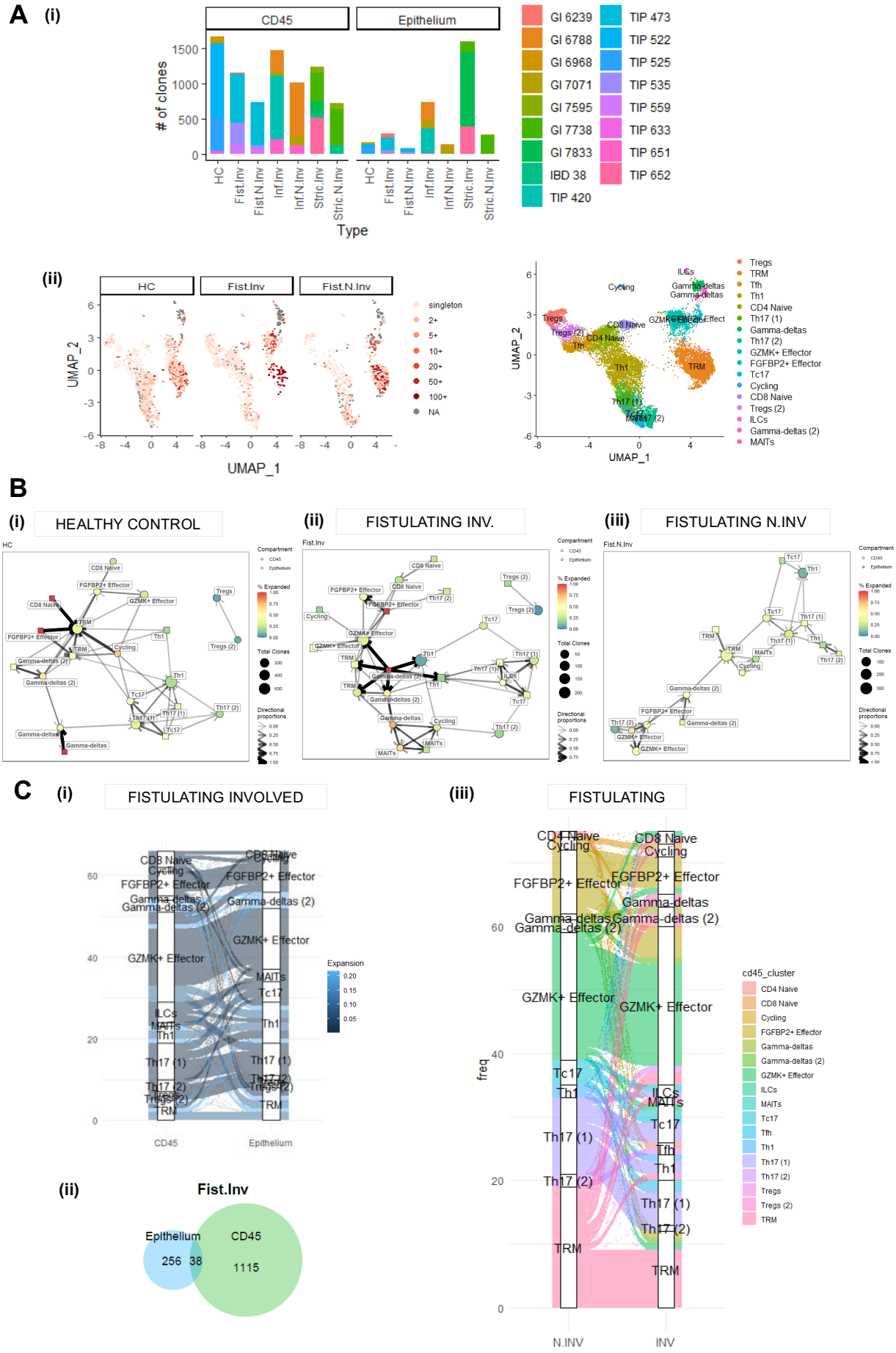

A
(i)
CD45
Epithelium
# of clones
Type
HC
Fist.Inv
Fist.N.Inv
Inf.Inv
Inf.N.Inv
Stric.Inv
Stric.N.Inv
GI 6239
GI 6788
GI 6968
GI 7071
GI 7595
GI 7738
GI 7833
IBD 38
TIP 420
TIP 473
TIP 522
TIP 525
TIP 535
TIP 559
TIP 633
TIP 651
TIP 652
(ii)
singleton
2+
5+
10+
20+
50+
100+
NA
UMAP_1
UMAP_2
Tregs
TRM
Tfh
Th1
CD4 Naive
Th17 (1)
Gamma-deltas
Th17 (2)
GZMK+ Effector
FGFBP2+ Effector
Tc17
Cycling
CD8 Naive
Tregs (2)
ILCs
Gamma-deltas (2)
MAITs
B
(i) HEALTHY CONTROL
(ii) FISTULATING INV.
(iii) FISTULATING N.INV
Compartment
% Expanded
Total Clones
Directional proportions
C
(i) FISTULATING INVOLVED
Expansion
(ii) Fist.Inv
Epithelium
CD45
256
38
1115
(iii) FISTULATING
freq
N.INV
INV
cd45_cluster

Figure 4.15: T-cell Dynamics in Fistulating CD (*previous page*). (A) Bar chart illustrating number of T-cell clones per condition and compartment (i). UMAP embedding visualising T-cell cluster annotations from the immune scRNA-seq data, overlaid with the number of TCR clones per cluster and condition (healthy controls, fistulating CD involved, and fistulating CD non-involved). The colour gradient indicates the number of TCR clones, with darker colours representing a higher number of clones (ii).
(B) Clonal dynamics of the T-cell repertoire, assessed using directional overlap analysis, comparing healthy controls to fistulating CD involved and non-involved tissues.
(C) Sankey diagram illustrating the distribution and flow of T-cell clones between stromal (CD45+) and epithelial compartments in fistulating CD (i). The diagram also shows the number of unique and shared T-cell clones in the epithelium and stroma of fistulating involved tissue (ii). Another Sankey diagram presents the distribution of T-cell clones between involved and non-involved fistulating tissues (iii).

4.7 Interim discussion

A novel population expressing IL11 was identified, predominantly localised in involved tissue of fistulating CD, present in both the epithelial and stromal compartments. This population exhibited expression patterns associated with tissue repair, remodelling, inflammation, and altered metabolism. The shared transcriptional profile observed in both compartments suggests a potential process of EMT, indicating a possible role in the pathogenesis of fistulating disease. However, the limited size of this cluster, especially in the epithelial compartment, necessitates further validation to confirm this hypothesis.

In addition to the possible EMT process, clues of aberrant wound healing repair processes contributing to fistula development were evident in the single-cell data. Aberrant expression of *WNT* and frizzled receptors was identified in fistula-associated stromal cells, implicating the involvement of *WNT* signalling pathways in this context. The dysregulation of these pathways may disrupt the orchestrated activity of these essential morphogens, critical for the maintenance of intestinal homeostasis.

The role of *WNT* signalling in gastrointestinal physiology and disease pathogenesis is multifaceted. Initiated during foetal development, *WNT* activity is crucial for the continual renewal of intestinal crypt-villus structures, essential for maintaining tissue integrity and function (231, 242). Furthermore, *WNT* signalling plays a pivotal role in facilitating repair and regeneration following injury,

highlighting its importance in maintaining intestinal homeostasis. Central to the maintenance of intestinal homeostasis is the preservation of an intact epithelial barrier, which is sustained by intestinal epithelial stem cells. *WNT* signalling, mediated through frizzled receptors and coreceptors, LRP5/6, serves as a key regulator in initiating stem cell differentiation into vital cell types crucial for barrier function (242).

WNT signalling pathways have been implicated in various pathological processes, including EMT in fistulating CD intestinal tissue. Studies have demonstrated the involvement of specific WNT ligands, such as WNT2b, and receptors, such as FZD4, in inducing EMT (99). Additionally, aberrant *WNT* pathway activation has been linked to stricturing CD (243) and IBD-associated colorectal carcinoma (244) pathogenesis, further emphasising its significance in disease progression.

While stromal cells, such as those within the crypt niche, are essential sources of WNT ligands in healthy intestinal tissue, their role shifts in disease states (245). In CD, stromal cells contribute to stromal remodelling of the intestine during inflammation, indicating a dynamic response to pathological stimuli. The spatial distribution of the newly identified fistula-associated stroma (FAS) fibroblasts throughout the intestinal layers poses challenges in characterisation solely through single-cell analysis. Understanding the cellular interactions and signalling cascades involving these fibroblasts also requires complementary approaches beyond single-cell analysis.

Furthermore, aberrant expression of regenerative morphogen pathways within fistula stroma, coupled with the overexpression of ECM proteins and developmental transcription factors like *TWIST1*, suggests dysregulation of normal wound healing processes. This dysregulation may lead to the development of fistulae, highlighting the interplay between developmental pathways and disease pathogenesis.

In the epithelial compartment, an increased abundance of undifferentiated, proliferative cells and transcriptional regulators of epithelial development, such as *SPDEF*, is observed in fistulating

involved tissue compared to controls. This heightened regenerative activity reflects the epithelium's adaptive response to chronic inflammation-induced epithelial damage. Specifically, the expansion of stem cells within fistulating involved tissue aims to replenish the epithelial cell population and preserve tissue integrity amidst ongoing inflammatory insult. Collectively, these findings mark the complex interplay between *WNT* signalling, epithelial-stromal interactions and tissue regeneration in the context of fistulating CD.

Histopathologically, fibrosis is a recognised co-existing feature of fistulating CD, which was also evident in the single-cell data; characterised by the enrichment of pro-fibrotic effector molecules in fistula stroma and epithelium. Changes in immune composition also contributed to fibrosis deposition, with expansions observed in myeloid cells and effector T-cell (e.g., Th17) populations. TCR analysis further unveiled distinct patterns of clonal expansion, highlighting the complex dynamics of T-cell populations within the diseased microenvironment.

The discovery of the FAS population emerged as particularly interesting. However, despite multiple optimisations to the stromal workflow, the FAS population remained relatively small. Upon reflection, I questioned whether our single-cell methods adequately captured the nuanced pathology of the fistula tract. It is possible that technical limitations, such as sampling variability or tissue dissociation-induced cell loss, hindered our ability to detect this population most effectively. Additionally, a key limitation of droplet-based scRNA-seq is its inability to capture certain cell types, such as neutrophils and Paneth cells. Neutrophils have inherently low RNA content and high RNase activity, leading to RNA degradation during the workflow and requiring highly sensitive scRNA-seq methods for detection. Similarly, the dense granules characteristic of Paneth cells can interfere with cell lysis, limiting RNA capture.

In response to these limitations, I decided to refine our approach to pathology characterisation. Specifically, I opted to employ ST to precisely map the location of this pathological cluster within the tissue and elucidate their co-localisation patterns. The integration of single-cell with ST data

would facilitate a deeper exploration into the emerging signals of epithelial-stromal crosstalk present in the data.

Furthermore, the generation of this data serves as a reference dataset to integrate and inform spatial analyses, present in my final results chapter.

Chapter 5: Spatially Resolved Insights into Fistulating Crohn's Disease Pathogenesis

5.1 Introduction

In this final part of the project, the aim was to spatially resolve the cellular heterogeneity of fistulating CD identified at the single-cell level using multimodal ST. Specifically, the focus was on spatially localising the fistula-specific clusters identified at single-cell, particularly the fistula associated stroma (FAS) fibroblast cells and their interactions with epithelial and immune compartments.

While scRNA-sequencing systematically profiled fistulating tissue, technical and sampling limitations meant that the actual fistula tract was not always clearly profiled in each instance. Spatial transcriptomics, however, overcame this limitation by enabling a detailed examination of the anatomy of the fistula tract and its gene expression. This approach is particularly well-suited for studying pathologies with such disrupted tissue architecture as found in fistulating CD.

The initial focus was to spatially map the entire transcriptome of fistulating tissue in an unbiased fashion, using the scRNA-sequencing data to complement these findings. Toward the latter stages of the project, novel imaging-based ST methods employing in-situ hybridization technology, such as

MERFISH and Xenium, emerged (246). These advancements enabled high-resolution, subcellular ST, providing the opportunity to interrogate precisely how tissue structurally remodels in fistula formation and to identify perturbations in cellular signalling, particularly within stromal-epithelial crosstalk.

This spatial cohort included an expanded set of samples to enable the study of perianal fistulating CD and non-CD fistulating controls, allowing for more robust comparisons. In this chapter, I will describe the comprehensive characterisation undertaken to create the first large-scale spatial atlas of fistulating CD pathology. This atlas provides a multimodal resource that catalogues the heterogeneity of fistulae while demonstrating the cellular microdomains shared across fistulating tissue.

5.2 Experimental overview

Spatial transcriptomic analysis was conducted across a diverse cohort of fistulating CD samples and controls (**Fig. 5.1**). The workflow involved selecting suitable FFPE intestinal tissue blocks, confirmed to contain fistula tract pathology by gastrointestinal histopathologists (in most cases, verified by two independent histopathologists). It was important to ensure that the fistula tract, in its entirety or in part, was captured within the area profiled transcriptomically. Utilising FFPE blocks allowed for the sampling of tissue collected retrospectively and prospectively.

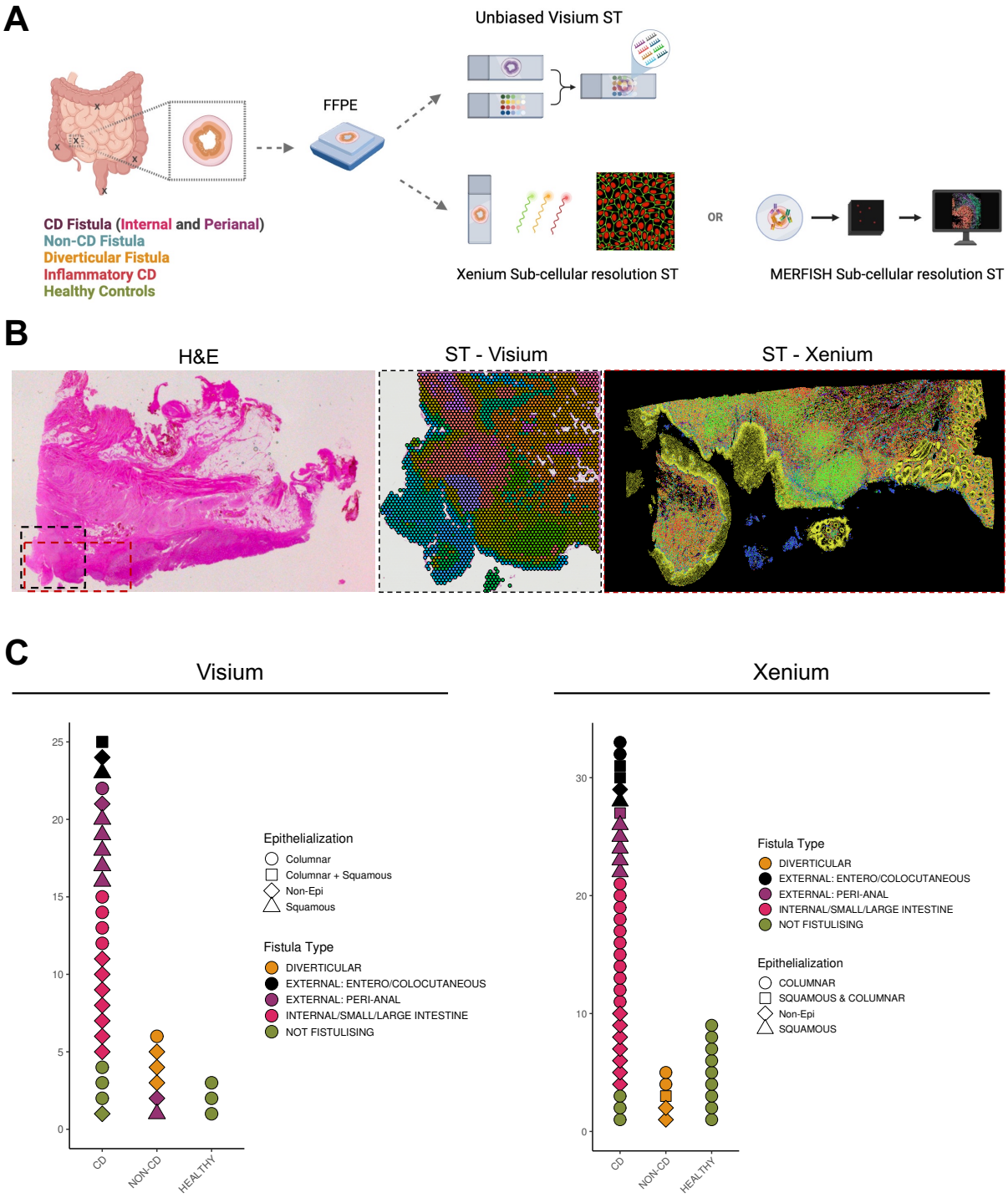


Figure 5.1: Spatial Transcriptomics Experimental Overview. (A) Schematic overview of the experimental design for ST analysis in fistulating (CD and non-CD), inflammatory CD and healthy controls. Formalin-fixed, paraffin-embedded surgically resected tissues underwent unbiased Visium ST and/or subcellular ST (Xenium or MERSCOPE). (B) Representative fistulating section stained with H&E, with corresponding Visium (black box) and MERSCOPE (red box) cluster overlays. (C) Dot plot overview of Visium and Xenium ST experiments, showing sample distribution across condition type, epithelialisation status, and fistula type. [FFPE, Formalin-fixed paraffin-embedded; H&E, haematoxylin & eosin; Non-Epi, Non-epithelialised]

RNA quality assessment of FFPE blocks, evaluated by DV200 assessment, served as another checkpoint before selection for ST (**Methods 2.10**). The samples (n= 38) then underwent unbiased ST using Visium (10x Genomics) (**Methods 2.10**). Briefly, tissue sections were mounted onto oligo-barcoded slides, where they underwent permeabilisation, allowing the release of RNA for reverse transcription and subsequent amplification, enabling sequencing. This workflow generated a total of 93,075 QC-passing tissue covered spots, with an average gene detection per spot of 2045, mean 5777 UMIs per spot (**Appx. Fig. A.9**).

Following UMAP embedding visualisation, 24 clusters were identified across all slides. To validate cluster analysis, initial alignment with corresponding H&E images was conducted to provide spatial context. Computational deconvolution of each spot was employed, with factor analysis based on single-cell data serving to determine the likely single-cell composition of each spot. This process enabled the spatial localisation of clusters identified at single-cell level.

Analysis of data generated from Visium, together with reference scRNA-seq, facilitated the curation of a customised gene panel for subcellular resolution ST, including targets specific to CD and fistulae. While some cases were initially profiled by MERFISH, due to challenges in consistent data recovery, a decision was made to discontinue further MERFISH (**Chapter 3.5.2**) and trial Xenium (10x Genomics) as an alternative.

High-resolution subcellular ST were performed either solely profiled on Xenium or with corresponding paired, unbiased analysis by Visium on selected cases. For Xenium, tissue sections were mounted on slides and processed to release sequestered mRNA, followed by probe hybridisation, ligation, amplification and imaging. An optical signature unique to each gene was then decoded and allocated to cells through cell segmentation. A total of 6,334,604 cells were recovered across 11 slides and 47 samples. In a similar approach to Visium, scRNA-seq data was utilised as a reference for gene subset and label transfer workflows to annotate cell types and clusters.

5.3 Clinical cohort

The spatial cohort profiled 62 fistulae obtained from 42 patients, comprising both prospective and retrospective collections spanning the period from 2011 to 2023. Among these, 51 fistula sections were identified as CD-associated, while 11 were categorised as non-CD-associated (**Fig. 5.1c**). Additionally, specimens from individuals diagnosed with non-fistulating CD and healthy, non-IBD, individuals were included to provide comparative context.

All patients diagnosed with CD had established diagnoses prior to undergoing surgical intervention. Tissue resections were meticulously evaluated, with macroscopic assessment performed by the operating surgeon and subsequent microscopic confirmation by a senior GI pathologist. The cohort of patients with CD comprised 17 males and 22 females, with a mean age of 42 years (IQR 24.5). The percentage of patients receiving biological therapies at the time of surgery was 37.5% for individuals with fistulating CD and 85.7% for those with non-fistulating CD. Among the CD patients with fistulae, 56.0% were female.

The majority of CD-related fistulae originated from the small or large bowel and terminated internally (59.4%), followed by perianal fistulae (31.3%). Fistula terminating externally represented the least prevalent subtype (enterocutaneous 6.3%, colocutaneous 3.1%) within the cohort, accounting for 10.6% of cases. Control cases of fistulae encompassed diverticular fistulae (75%) and idiopathic cryptoglandular perianal fistulae (25.0%). The mean age of control fistulae cases was 64 years (IQR 19), with 71% being male. Malignant fistulae were not included in the project.

Further classification of fistulae delineated them into epithelialised (54.8%) and non-epithelialised (45.2%) tracts, categorised by the type of epithelium present (squamous *vs.* columnar *vs.* squamo-columnar, 26.2% *vs.* 23.8% *vs.* 4.8%) and by inflammatory status. Among the non-fistulating controls, 7 were identified as CD cases (inflammatory, 86.0%; stricturing, 14.0%), while 9 specimens were

derived from healthy individuals, including both ileal and colonic samples. A comprehensive summary of the sample cohort is provided in the appendix (**Appx. Table A.3**).

5.4 Histopathological features of fistulating CD cohort

On histological examination, 82.0% of the studied fistulae were associated with CD. The majority of these specimens exhibited distinct fistula tracts that extended deeply from the lamina propria and muscularis mucosae into the muscularis propria and serosal layers of the intestinal wall.

Approximately 45.2% of the observed fistula tracts were non-epithelialised, lacking a covering epithelium. Conversely, epithelialised fistula tracts typically presented with either squamous or columnar epithelium, corresponding to the location of the fistula. Transition in epithelial type, from squamous to columnar, was identified in cases where the intestinal fistula communicated with another organ, most commonly the skin (e.g., enterocutaneous). Perianal fistulae with stratified squamous epithelium exhibited histological features of hyperkeratosis, hypergranulosis, and mild acanthosis.

Areas of mucosal erosion within fistula tracts revealed evidence of regenerative mucosa, characterised by reactive, flattened epithelium with large nuclei. Non-epithelialised fistula tracts were lined with granulation tissue showing features of neovascularisation and dense inflammatory infiltrate. Chronic fibrosis was evident in nearly half of the cases, characterised by hypertrophied muscularis propria, submucosal fibrosis, and neural hyperplasia, often surrounding the fistula tract.

The extent of fistula penetration varied considerably among cases, suggesting a temporal progression in fistula development. In certain instances, the fistula tract was confined to the muscularis propria, where disruption of muscle fibres was evident. Conversely, in other cases, the fistula tract extended through the layers of the intestinal wall, reaching the serosa. This deeper penetration was often associated with serositis and evidence of haemorrhage. Additionally, some established fistula tracts

were found to be superficial, running parallel to the muscularis mucosae without penetrating the submucosa. Fissuring ulcers were frequently observed in conjunction with established fistula tracts, and are a recognised precursor for fistula formation (247).

The presence of inflammation was observed in the majority of fistulae, with features of both acute and chronic inflammation. Chronic inflammatory features included the formation of lymphoid follicles, the presence of non-caseating granulomas and lymphoplasmacytic infiltrate. Acute inflammation was characterised by neutrophilic infiltrate and abscess formation. Mural abscess formation was frequently observed, particularly adjacent to the fistula, with extensions deep within the muscularis propria. Mucin and debris could also be appreciated within the lumen of the fistula tract, along with non-specific features of acute inflammation.

Inflammation distribution varied from dense infiltrate in the lamina propria to transmural involvement, with minimal inflammation noted in some cases. Additionally, adipose deposition and oedema within the submucosa were common features observed in fistulae **(Fig. 5.2)**.

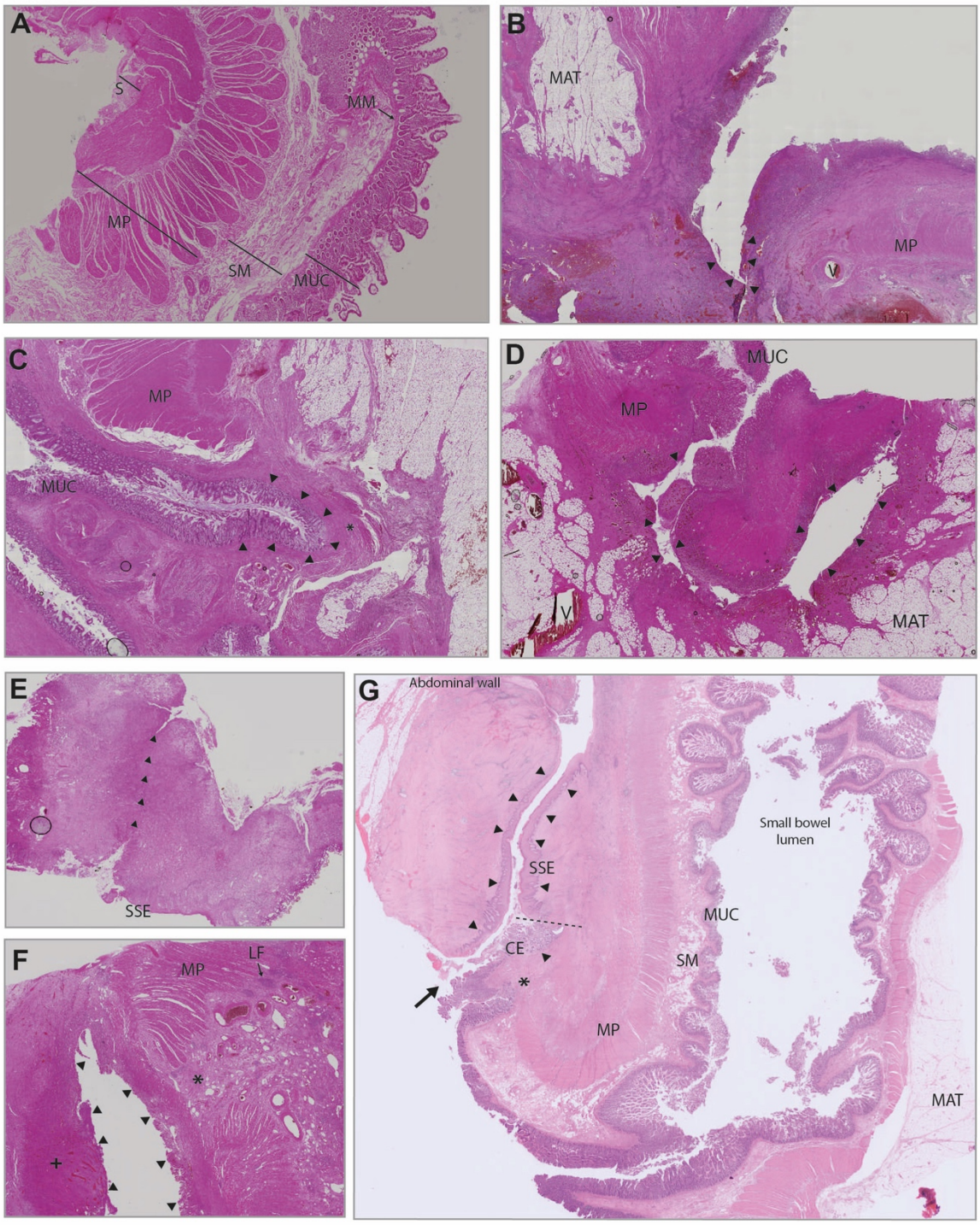


Figure 5.2: Representative Histopathological Features of Crohn's Disease Fistulae, including non-epithelialised *vs.* epithelialised *vs.* non-epithelialised (B, C, D), non-perianal and perianal (D, E), and immune-rich *vs.* quiescent (F, G) presentations.

(A) Normal control ileum displaying representative layers of healthy full-thickness bowel (haematoxylin and eosin, H&E, ×0.88 original magnification) (Patient GI 9676).

(B) Small bowel CD fistula lacking covering epithelium (arrowheads). Granulation tissue lines the fistula tract with neovascularisation and dense inflammatory infiltration. Fibrotic scarring is also evident in the muscularis propria (H&E, ×0.44) (Patient JR 31183_22).

(C) Epithelialised colonic CD fistula (arrowheads) originating from the ileocaecal valve. The fistula tract branches deep within the muscularis propria, interrupting the muscle fibres (*), but does not fully penetrate all the intestinal layers,

indicating an early stage of fistula development. No acute inflammation is observed (H&E, ×0.44) (Patient JR 17451_20).

(D) Colonic CD fistula demonstrating loss of lining columnar epithelium along fistula tract (arrowheads) (H&E, ×0.3) (Patient BAY 86068_17).

(E) Non-epithelialised perianal CD fistula. The distal edge of the fistula at the anal side is delineated by stratified squamous epithelium. The leading direction of the fistula is indicated by arrowheads. The proximal edge of the fistula is more proliferative with associated inflammation (H&E, ×0.88) (Patient BAY 48715_20).

(F) Ileocolonic CD fistula lacks lining epithelium (arrowheads) with inflammatory infiltrate and lymphoid follicle formation. Abscess formation is seen adjacent to the fistula (+), along with muscularis propria disruption (*). (H&E, x0.44) (Patient BAY 20986).

(G) CD Enterocutaneous fistula with a fistula tract (arrowheads) traversing between the anterior abdominal wall and ileum, including an external opening (arrow). On the ileal side of the fistula, the mucosa is characterised by flattened columnar epithelium due to reactive changes (*). On the cutaneous side of the fistula, reactive, narrow squamous epithelium is observed. A small focus of transitional epithelium from columnar to squamous is present (dashed line). Chronic fibrosis is evident, characterised by hypertrophied muscularis propria and fibrotic submucosa within the ileum. Fat wrapping is also observed in this specimen (H&E, ×0.25) (Patient JR 28598_21).
[CE, columnar epithelium; LF, lymphoid follicle; MAT, mesenteric adipose tissue; MM, muscularis mucosae; MP, muscularis propria; MUC, mucosa; SM, submucosa; S, serosa; SSE, stratified squamous epithelium].

5.5 Unbiased spatial transcriptomic analysis of fistulating CD

5.5.1 Quality control, gene recoveries and cluster distribution

Out of the initial 38 samples processed using Visium, 34 samples were included in the final analysis. The exclusion of samples was primarily due to issues with low gene recovery, which often stemmed from factors such as tissue detachment, overall sample quality or the type of tissue being analysed.

To ensure a comprehensive analysis, the data from these samples were merged for cohort analysis and then examined slide by slide (**Appx. Fig. A.6-8**). Utilising dimensionality reduction techniques such as principal component analysis, we observed distinct separation based on different conditions, as depicted in **Figure 5.3**. Furthermore, the data were stratified based on the fistula epithelial type, distinguishing between columnar, squamous, and non-epithelialised regions.

The most significant separation was observed between columnar and squamous epithelium. This separation also correlated to the type of fistula; for example, peri-anal fistula samples exhibited markers consistent with squamous epithelium, as expected for cutaneous regions.

UMAP embedding visualisation identified a total of 31 clusters. To facilitate easier comparison across the merged dataset, we defined eleven broad clusters by integrating the more specific ones. For instance, the integrated cluster representing columnar epithelium showed enrichment in stem cells, transit amplifying cells, goblet cells, BEST4 cells and enterocytes. The FAS cluster originally identified in the single-cell data was also identified in the Visium data and could be stratified into two clusters: FAS 1 and FAS 2. FAS 1 primarily consisted of immune-rich regions, including $IL11^+$ cells, myeloid cells, and neutrophils, whereas FAS 2 comprised stromal subtypes along with $IL11^+$ cells.

The heterogeneity of fistula samples was evident from the distribution of clusters per sample, as illustrated in **Figure 5.4**. While most CD fistula samples were dominated by immune-rich regions, variations were observed within CD subtypes, such as internal fistulae, which showed varying proportions of columnar epithelium and degrees of haemorrhage. In contrast, healthy controls and non-CD perianal samples were more consistent in their composition, with minimal to no immune-rich regions and lymphoid follicles. The reduced proportion of muscularis in perianal samples could be attributed to the incomplete capture of full bowel thickness in some of these samples.

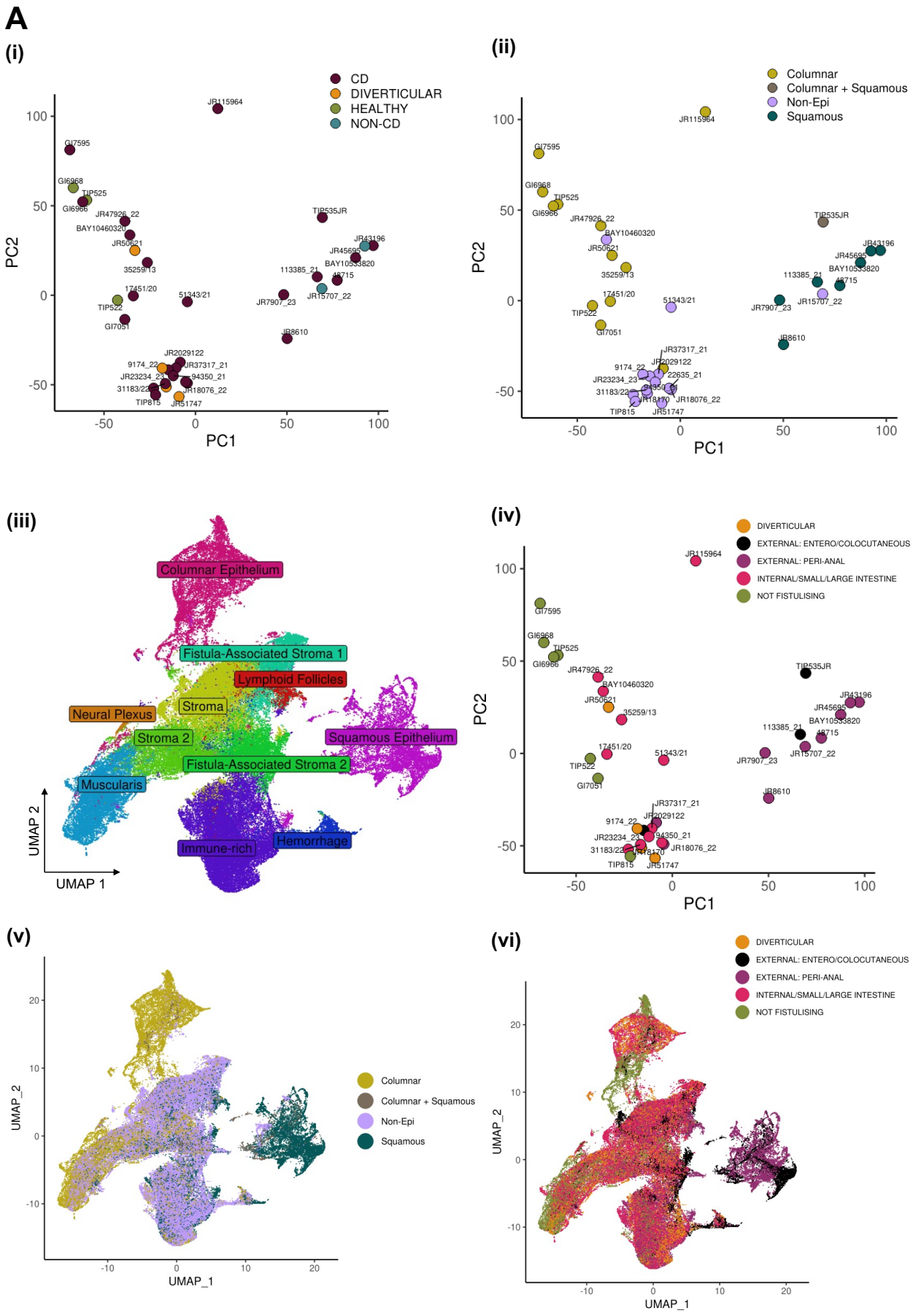

A
(i)
CD
DIVERTICULAR
HEALTHY
NON-CD
PC1
PC2
(ii)
Columnar
Columnar + Squamous
Non-Epi
Squamous
PC1
PC2
(iii)
Columnar Epithelium
Fistula-Associated Stroma 1
Lymphoid Follicles
Neural Plexus
Stroma
Stroma 2
Squamous Epithelium
Fistula-Associated Stroma 2
Muscularis
Immune-rich
Hemorrhage
UMAP 1
UMAP 2
(iv)
DIVERTICULAR
EXTERNAL: ENTERO/COLOCUTANEOUS
EXTERNAL: PERI-ANAL
INTERNAL/SMALL/LARGE INTESTINE
NOT FISTULISING
PC1
PC2
(v)
Columnar
Columnar + Squamous
Non-Epi
Squamous
UMAP_1
UMAP_2
(vi)
DIVERTICULAR
EXTERNAL: ENTERO/COLOCUTANEOUS
EXTERNAL: PERI-ANAL
INTERNAL/SMALL/LARGE INTESTINE
NOT FISTULISING
UMAP_1
UMAP_2

Figure 5.3: **Merged Visium Cohort Using Broad Clustering *(previous page)*.** (A) Principal component analysis (PCA) based on condition type (i), epithelialisation status (ii) and fistula type (iv). UMAP embedding visualising Visium broad cluster grouping (iii) further stratified by epithelialisation status (v) and fistula type (vi). [Non-epi, non-epithelialised]

5.5.2 Merged Visium cohort analysis

Differential abundance analysis was performed using integrated neighbourhoods to compare fistulating subtypes and controls within the Visium cohort. Meta cluster annotation was employed to simplify the representation of the large data and provided further delineation of the clusters, with the identification of two immune-rich fibroblast populations and two further FAS populations (FAS 3 and FAS 4) (**Fig. 5.4**). Upon comparing all fistulating samples (both CD and non-CD) to non-fistulating controls, a notable enrichment in stromal subsets—specifically the FAS 1, FAS 2, and FAS 4 clusters - was observed in fistulating samples. Additionally, immune-rich stroma 1 and 2 exhibited significant enrichment in fistulating samples. Squamous epithelium was differentially enriched in fistulating samples, attributable to the presence of cutaneous tissue absent in non-fistulating controls.

Conversely, clusters associated with mucosa, neural plexus, submucosa, and muscularis were reduced in abundance in fistulating samples. This reduction likely reflects sampling differences rather than true biological variation, as the deeper tissue layers are less prevalent in perianal fistulating samples.

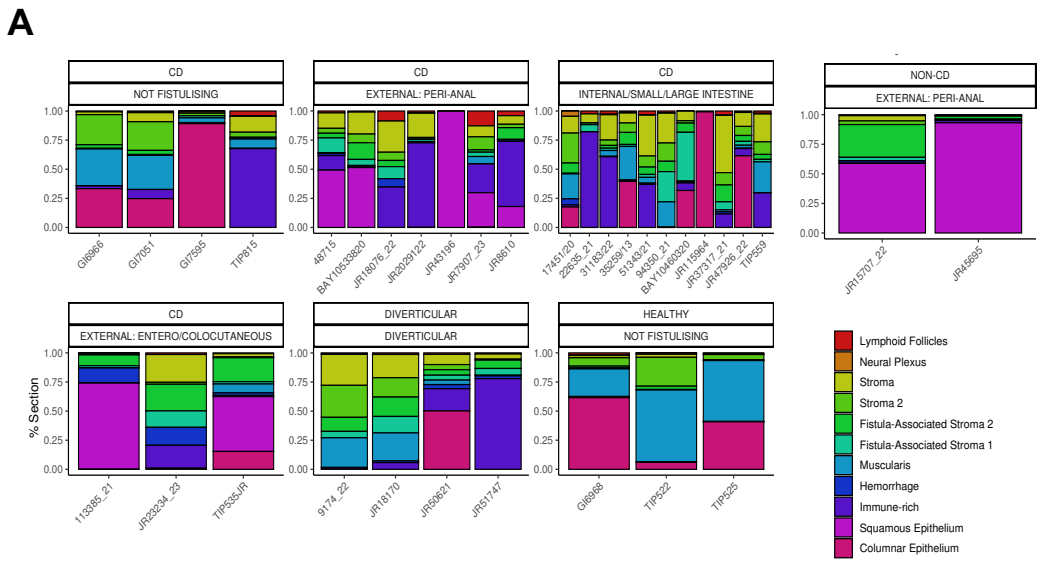
While the FAS population was identified in non-CD controls (e.g., diverticular), differences in the abundance of each FAS subcluster were observed across sample types. These differences reflect the heterogeneity of the clusters and correspond to locational variations of the subclusters. A focused comparison between internal fistulating CD samples and fistulating diverticular controls revealed a significant expansion of the FAS 3 cluster and an increased abundance of follicular cells in fistulating CD samples. Conversely, the FAS 1 cluster was significantly more abundant in diverticular samples.

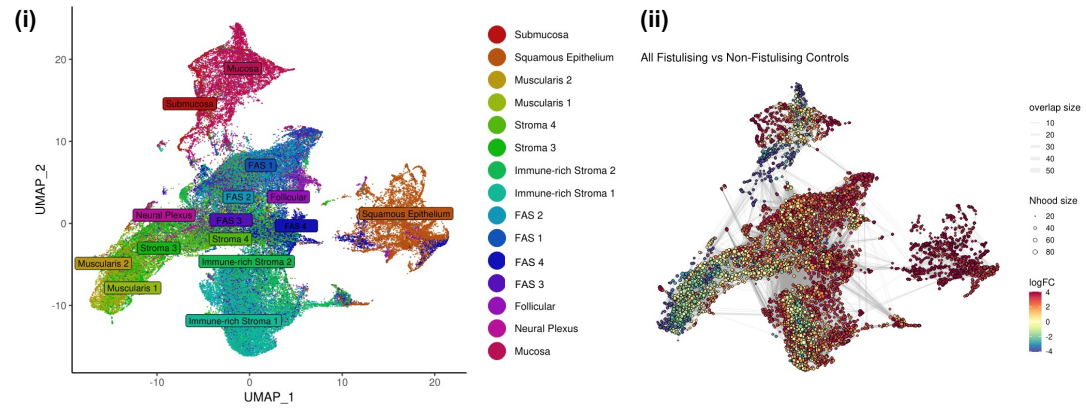
Comparing external (perianal) fistulating CD samples to diverticular controls revealed an abundance of squamous epithelium, follicular cells, and FAS 3. In contrast, FAS 1, stroma 3 and 4 were reduced in abundance, suggesting a distinct stromal and epithelial composition in perianal fistulae.

The region adjacency analysis aimed to identify adjacent domains and regions from the Visium cohort, providing insights into the spatial relationships and dimensions of each region. In healthy controls, the expected absence of FAS subclusters, squamous epithelium, and immune-rich stroma was noted. Deeper tissue layers, such as muscularis, fibroblast-rich stroma (stroma 3 and 4), and neural plexus, were observed to relate to each other, while more superficial structures, such as mucosa and submucosa, clustered in close proximity.

In internal fistulating CD, close proximity and interrelationship between FAS 2 and immune-rich stroma 1 were observed, with bi-directional signalling. Additionally, follicular signals were directed towards both FAS 2 and immune-rich stroma 1. Immune-rich stroma 2 was also in close proximity to FAS 3, which received signals from both FAS 3 and FAS 4.

In both internal and external CD fistulating samples, as well as diverticular samples, the proximity and interrelatedness of FAS 1 and FAS 2 to immune-rich stroma were evident. In external CD samples, FAS 3 was observed in close proximity to squamous epithelium, suggesting a potential role for FAS 3 in relation to the cutaneous edge of a fistula, with bi-directional signalling further supporting this relationship **(Fig. 5.5)**.


A
CD
NOT FISTULISING
CD
EXTERNAL: PERI-ANAL
CD
INTERNAL/SMALL/LARGE INTESTINE
NON-CD
EXTERNAL: PERI-ANAL
CD
EXTERNAL: ENTERO/COLOCUTANEOUS
DIVERTICULAR
DIVERTICULAR
HEALTHY
NOT FISTULISING
% Section
Lymphoid Follicles
Neural Plexus
Stroma
Stroma 2
Fistula-Associated Stroma 2
Fistula-Associated Stroma 1
Muscularis
Hemorrhage
Immune-rich
Squamous Epithelium
Columnar Epithelium


B
(i)
UMAP_1
UMAP_2
Submucosa
Squamous Epithelium
Muscularis 2
Muscularis 1
Stroma 4
Stroma 3
Immune-rich Stroma 2
Immune-rich Stroma 1
FAS 2
FAS 1
FAS 4
FAS 3
Follicular
Neural Plexus
Mucosa
(ii)
All Fistulising vs Non-Fistulising Controls
overlap size
Nhood size
logFC

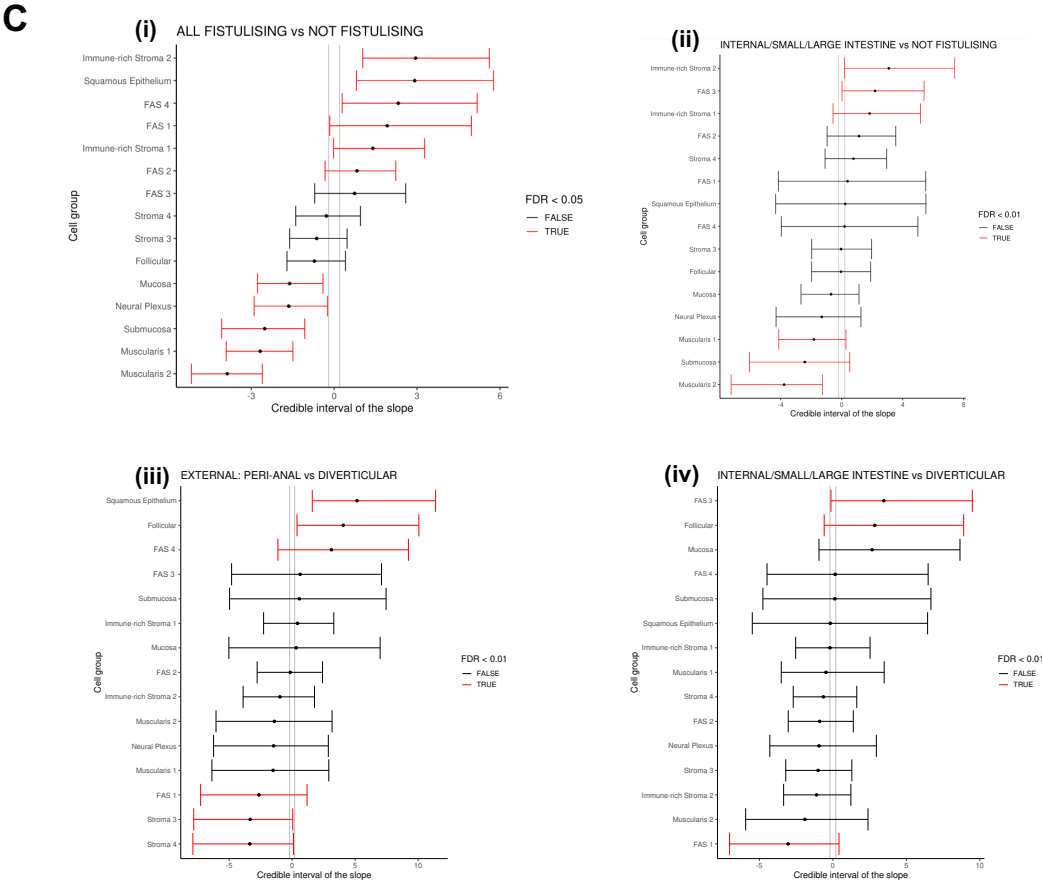
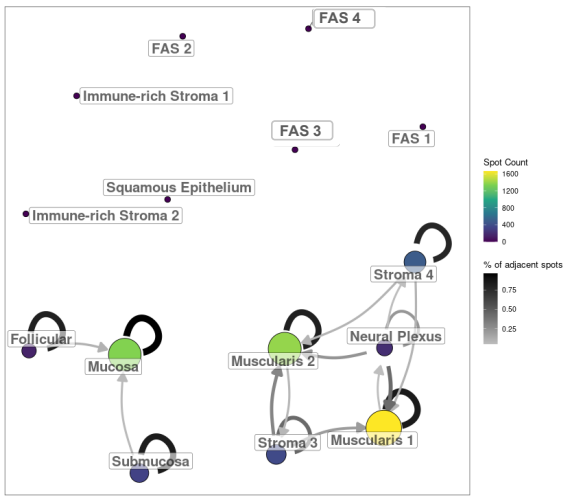

C
(i)
ALL FISTULISING vs NOT FISTULISING
Cell group
Credible interval of the slope
FDR < 0.05
FALSE
TRUE
(ii)
INTERNAL/SMALL/LARGE INTESTINE vs NOT FISTULISING
FDR < 0.01
(iii)
EXTERNAL: PERI-ANAL vs DIVERTICULAR
(iv)
INTERNAL/SMALL/LARGE INTESTINE vs DIVERTICULAR

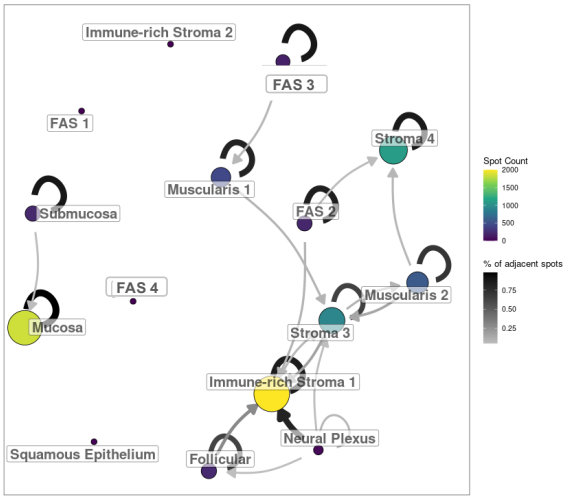
Figure 5.4: Merged Visium Cohort Using Metaclustering *(previous page).* (A) Proportion of samples per broad Visium region, further stratified by condition and fistula type. (B) UMAP embedding visualising Visium metacluster grouping (i) and neighbourhood abundance by metacluster in fistulising versus non-fistulising controls, shown using a Milo network. Statistical significance of the log fold change (logFC) was determined using the Wilcoxon rank-sum test, corrected for multiple comparisons (ii).
(C) Forest plot illustrating metacluster abundance analysis across various conditions: All fistulising vs. non-fistulising (i), internal Crohn's disease (CD) fistulising vs. non-fistulising (ii), external CD fistulising vs. diverticular fistulising (iii), and internal CD fistulising vs. diverticular fistulising (iv). Compositional analysis was performed using sccomp, with statistical significance determined by an FDR threshold of <0.01. [FAS, Fistula-Associated Stroma]

A

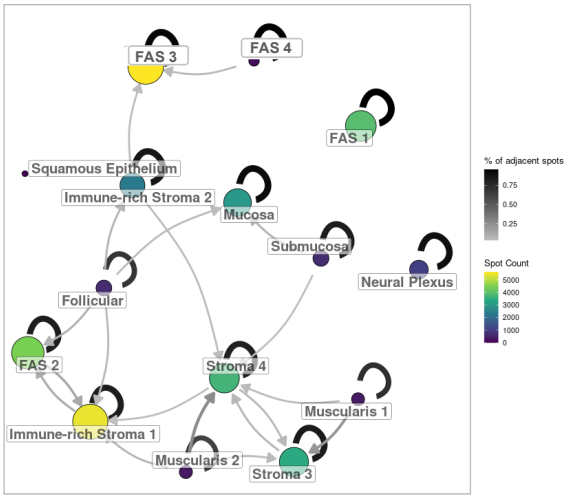
(i) Healthy Controls



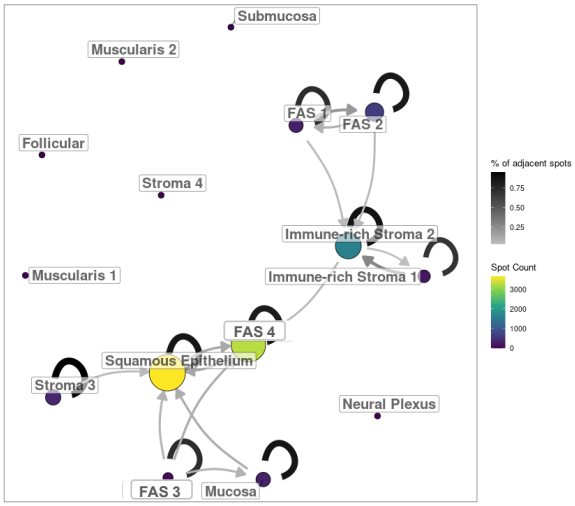
(ii) CD Non-Fistulising Controls



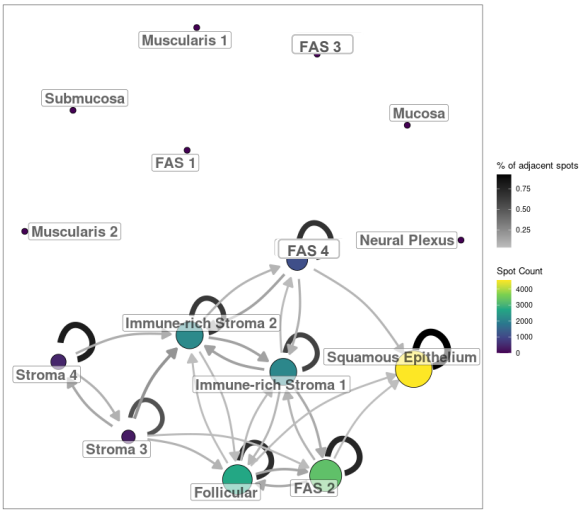
(iii) CD Internal



(iv) CD External



(v) CD Peri-Anal



(vi) Diverticular

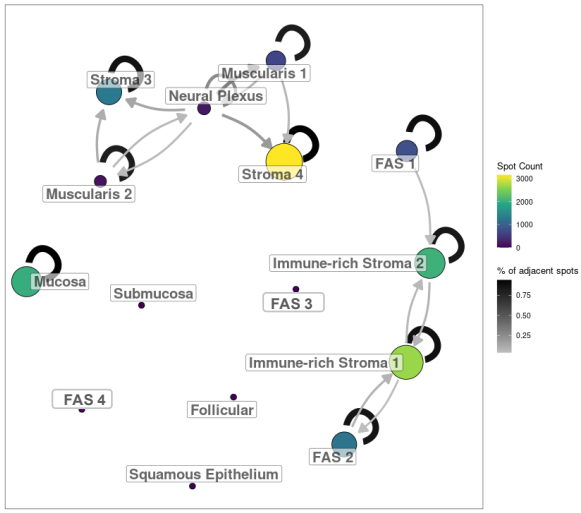


Figure 5.5: Region Adjacency Analysis of Visium Cohort (*previous page*). Region adjacency analysis of the merged Visium cohort, stratified by fistula type: healthy controls (i), Crohn's disease (CD) non-fistulising controls (ii), CD internal (iii), CD external (iv), CD peri-anal (v), and diverticular (vi). Spot size indicates the relative expression level of cell types within each region. Arrows depict interactions between adjacent regions, with a colour gradient representing spot count: darker shades indicate fewer spots, while lighter shades reflect a higher concentration of spots in the given region. Region proximity was determined using the igraph package with the Fruchterman-Reingold layout algorithm. Minor edges representing less than 5% of interactions were filtered out. [FAS, Fistula-Associated Stroma]

5.6 Subcellular resolution spatial transcriptomics – Xenium cohort

A total of 55 samples were run by Xenium. After excluding failed and repeat samples, 44 unique samples from 38 patients were included in the analysis. For sections that had already been profiled by Visium, Xenium data were taken from directly adjacent tissue sections to complement the unbiased full transcriptome analysis. This resulted in 21 paired samples (17 of which were from the same region) and 23 unpaired samples (**Fig. 5.6a**).

Among these samples, 19 were from fistulating CD cases, 5 were from non-CD fistulae, and 8 were healthy controls. The full dataset encompassed 6,334,604 quality control-passing cells. Principal component analysis demonstrated high concordance between Xenium and Visium, with similar separation of clusters by disease, location, fistula type and epithelial subtype (**Fig. 5.6c**).

High resolution clustering analysis of major cell lineages from the Xenium cohort identified approximately 60 cell sub-populations (**Fig. 5.6d**). Broad cell populations in the scRNA-seq dataset were analogous to cell populations in the Xenium dataset and tissue regions in the unbiased Visium dataset. Differential abundance analyses identified consistent enrichment of multiple fibroblast sub-populations across all fistula types, although type specific variations were observed – e.g., squamous epithelium was only detected in perianal fistulae. A strong enrichment of neutrophils as well as multiple fistula-specific macrophage clusters were detected (**Fig. 5.6e**).

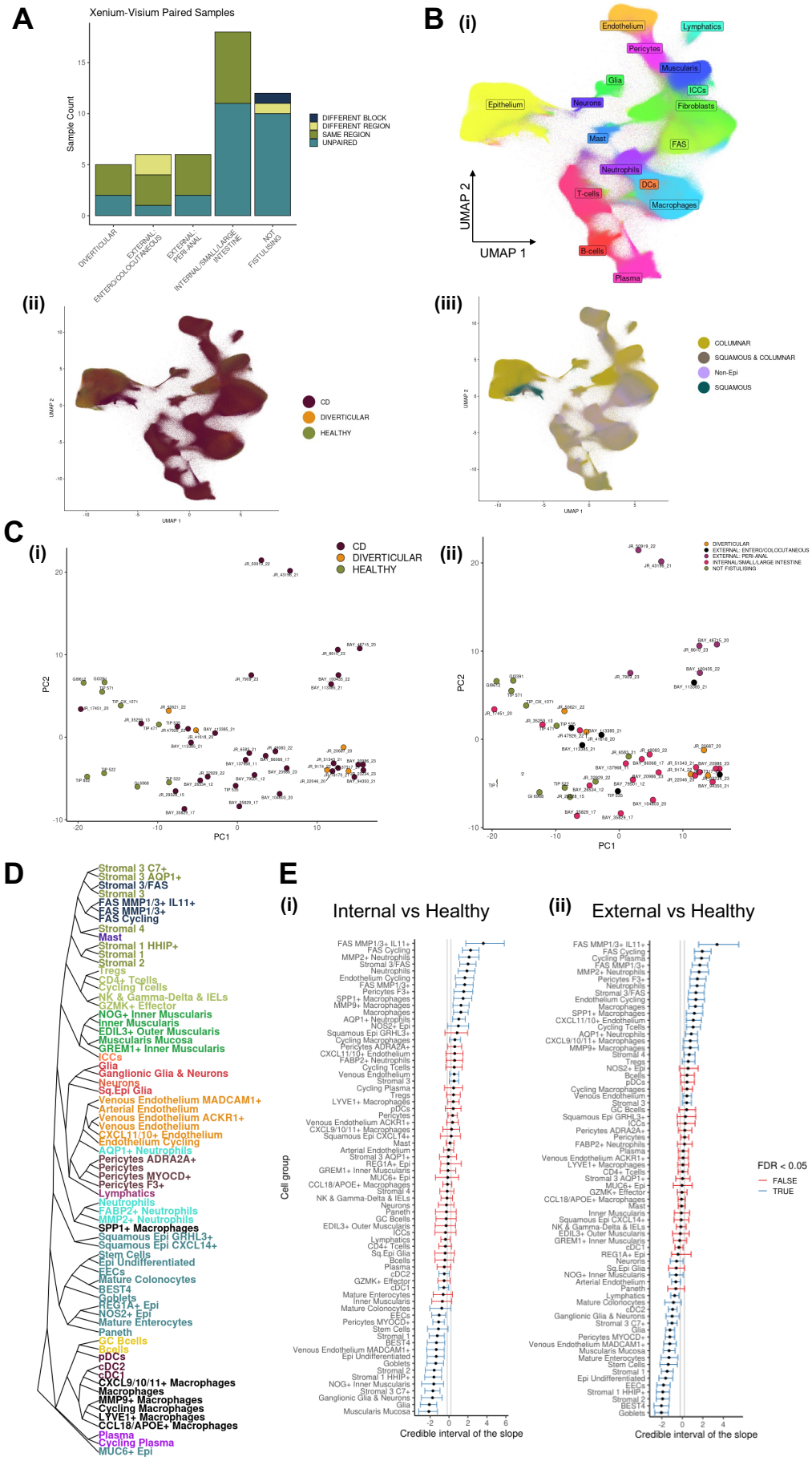

A
Xenium-Visium Paired Samples
Sample Count
DIFFERENT BLOCK
DIFFERENT REGION
SAME REGION
UNPAIRED
DIVERTICULAR
EXTERNAL: ENTERO/COLOCUTANEOUS
EXTERNAL: PERI-ANAL
INTERNAL/SMALL/LARGE INTESTINE
NOT FISTULISING
B (i)
Endothelium
Lymphatics
Pericytes
Muscularis
Glia
ICCs
Epithelium
Neurons
Fibroblasts
Mast
FAS
Neutrophils
DCs
T-cells
Macrophages
B-cells
Plasma
UMAP 2
UMAP 1
(ii)
CD
DIVERTICULAR
HEALTHY
(iii)
COLUMNAR
SQUAMOUS & COLUMNAR
Non-Epi
SQUAMOUS
C
(i)
(ii)
PC1
PC2
D
Stromal 3 C7+
Stromal 3 AQP1+
Stromal 3/FAS
Stromal 3
FAS MMP1/3+ IL11+
FAS MMP1/3+
FAS Cycling
Stromal 4
Mast
Stromal 1 HHIP+
Stromal 1
Stromal 2
Tregs
CD4+ Tcells
Cycling Tcells
NK & Gamma-Delta & IELs
GZMK+ Effector
NOG+ Inner Muscularis
Inner Muscularis
EDIL3+ Outer Muscularis
Muscularis Mucosa
GREM1+ Inner Muscularis
ICCs
Glia
Ganglionic Glia & Neurons
Neurons
Sq.Epi Glia
Venous Endothelium MADCAM1+
Arterial Endothelium
Venous Endothelium ACKR1+
Venous Endothelium
CXCL11/10+ Endothelium
Endothelium Cycling
AQP1+ Neutrophils
Pericytes ADRA2A+
Pericytes
Pericytes MYOCD+
Pericytes F3+
Lymphatics
Neutrophils
FABP2+ Neutrophils
MMP2+ Neutrophils
SPP1+ Macrophages
Squamous Epi GRHL3+
Squamous Epi CXCL14+
Stem Cells
Epi Undifferentiated
EECs
Mature Colonocytes
BEST4
Goblets
REG1A+ Epi
NOS2+ Epi
Mature Enterocytes
Paneth
GC Bcells
Bcells
pDCs
cDC2
cDC1
CXCL9/10/11+ Macrophages
Macrophages
MMP9+ Macrophages
Cycling Macrophages
LYVE1+ Macrophages
CCL18/APOE+ Macrophages
Plasma
Cycling Plasma
MUC6+ Epi
E
(i) Internal vs Healthy
(ii) External vs Healthy
Cell group
Credible interval of the slope
FDR < 0.05
FALSE
TRUE

Figure 5.6: Merged Xenium Cohort Overview *(previous page).* (A) Bar plot showing the distribution of paired Xenium-Visium samples across conditions.
(B) UMAP embedding visualising Xenium clusters of single cells (i) stratified by condition (ii) and epithelialisation type (iii).
(C) Principal component analysis (PCA) highlighting variations based on condition type (i) and fistula type (ii).
(D) Lineage tree illustrating the sub-clustering exploration of major cell lineages derived from single-cell Xenium data.
(E) Forest plot showing the differential abundance analysis across Xenium subpopulations in CD Internal vs. Healthy (i) and CD External vs. Healthy (ii). Compositional analysis was performed using sccomp, with statistical significance determined by an FDR threshold of <0.05. [FAS, Fistula Associated Stroma; PC, Principal Component]

5.7 Spatial microdomains of fistulating CD

In an effort to address the inherent variability among fistulae, a comprehensive profiling of multiple fistulae representing various subtypes was undertaken (**Appx. Fig. A.7 and A.10**). In doing so, I was able to identify shared cellular microdomains across different types of fistulae, elucidated as spatial microdomains, providing a more detailed analysis of the microenvironment within fistulating tissue (**Table 5.1**). These spatial microdomains exhibit interconnectedness, suggesting functional correlations and dependencies. Subsequently, the following subchapters delineate these distinct spatial microdomains and explore their interrelationships, in an attempt to elucidate the complex cellular dynamics inherent to fistulating CD.

5.7.1 Healthy intestinal crypt architecture by spatial analysis

Before defining the epithelial perturbations in CD fistulae, it was essential to spatially resolve normal intestinal crypt architecture. In adults, the maintenance of a self-renewing intestinal epithelium and the determination of cell fate along the crypt-axis is regulated by morphogens. These signalling molecules establish concentration gradients within developing tissues, which are crucial for proper tissue patterning. The morphogen gradients are modulated by antagonists, with the non-canonical WNT/BMP polarised gradient being a well-established example that regulates the expression of intestinal stem cells. Charting the coordinated activity of various morphogens necessary for the normal development of intestinal compartments is imperative to understanding when such pathways become disrupted in disease.

Previously published work from our group enabled the in-silico localisation of cells within the colonic crypt by defining a crypt base-crypt top axis score, or crypt-axis score, using 15 genes expressed in both absorptive and secretory cells. Cells predicted to reside at the top of the crypt were identified as more mature populations, as inferred by pseudotime ordering, and vice versa for cells at the crypt base (167).

Utilising high-resolution ST, we negate the need for in-silico predictions for the localisation of cells, as we can discriminately profile the discrete cellular anatomy of an intestinal crypt. The overlay in **Figure 5.7**, as a representative example, demonstrates the precise location of various cell types within the colonic crypt. Stem cells, marked by Lgr5 and ASCL2, and Paneth cells, marked by DEFA5, are located at the crypt base. TA cells, marked by MKI67, are situated just above the stem cells in the lower third of the crypt.

Goblet cells, marked by MUC2, and Tuft cells, marked by POU2F3, are distributed throughout the crypt and villus, while BEST4^{+} cells are located in the upper crypt and along the villus. Additionally, the distal enterocyte marker, SLC26A3, is expressed along the crypt axis, with the highest expression observed at the crypt top, consistent with the expected localisation of mature enterocytes. Mucosal structural cells (Stromal 1, S1 or ADAMDEC1^{+} fibroblast) support the crypt and villus. Telocytes (Stromal 2, S2 or BMP5/PDGFRA^{+} fibroblasts) are distributed in the peri-cryptal region, and submucosal fibroblasts (Stromal 3, S3, C3^{+} fibroblasts) reside in deeper tissue, more distal to the crypt niche. Glial cells (S100B), which provide support and protection for enteric neurons, are located along the crypt and villus. The muscularis mucosae can be seen delineating the mucosa from the submucosa. This high-resolution mapping of normal crypt architecture provides a critical reference for identifying and understanding epithelial perturbations in CD fistulae.

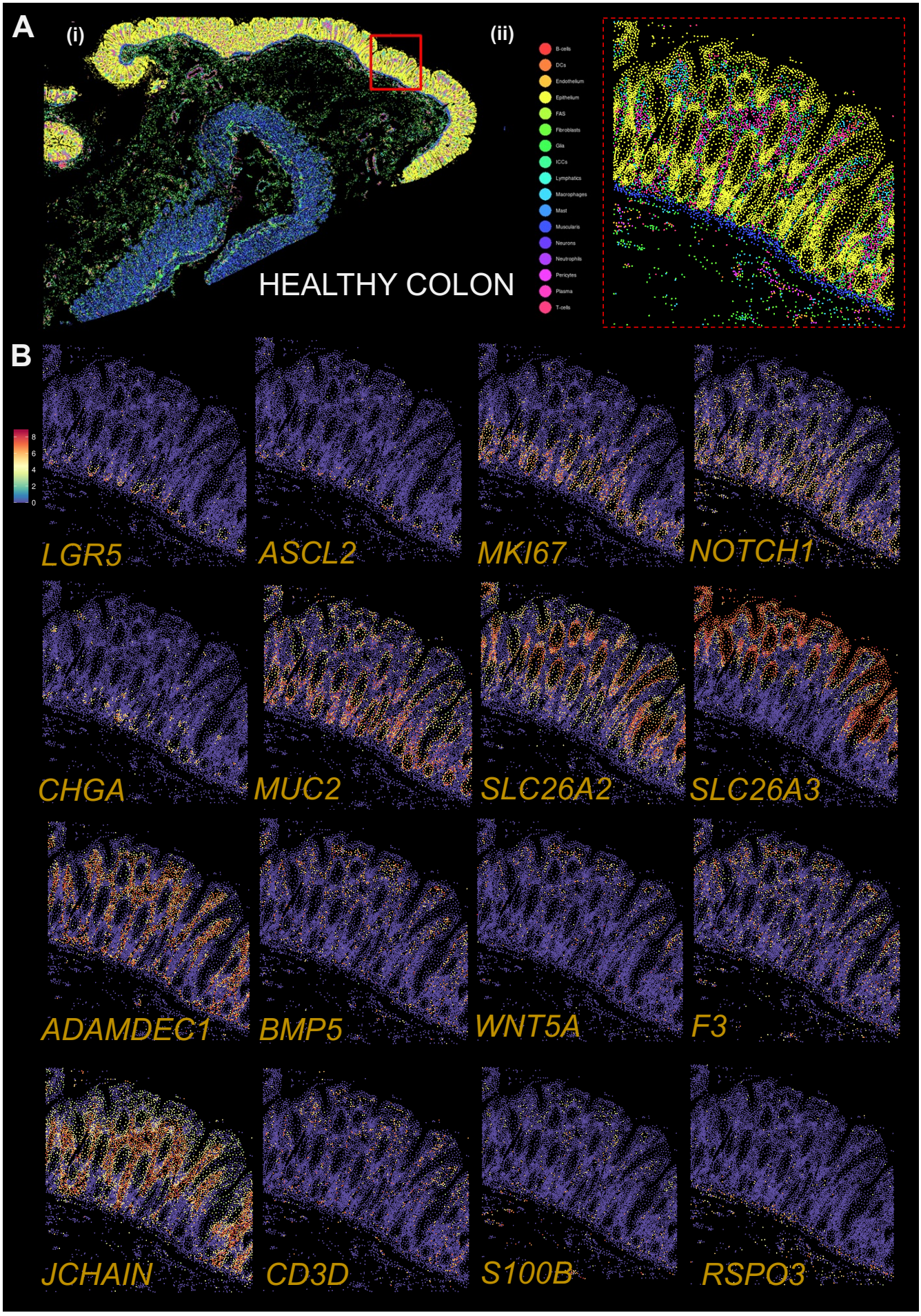

A
(i)
(ii)
HEALTHY COLON
B-cells
DCs
Endothelium
Epithelium
FAS
Fibroblasts
Glia
ICCs
Lymphatics
Macrophages
Mast
Muscularis
Neurons
Neutrophils
Pericytes
Plasma
T-cells
B
8
6
4
2
0
LGR5
ASCL2
MKI67
NOTCH1
CHGA
MUC2
SLC26A2
SLC26A3
ADAMDEC1
BMP5
WNT5A
F3
JCHAIN
CD3D
S100B
RSPO3

Figure 5.7: Subcellular Spatial Transcriptomic Analysis of Healthy Intestinal Crypt *(previous page)*.
(A) Xenium cluster overlay of full-thickness healthy colonic tissue (Patient TIP 1071) (i), with a magnified view of the crypt and crypt niche (ii).
(B) Overlay of expression for the following single genes: *LGR5, ASCL2, MKI67, NOTCH1, CHGA, MUC2, SLC26A2, SLC26A3, ADAMDEC1, BMP5, WNT5A, F3, JCHAIN, CD3D, S100B, RSPO3*. The gene expression is depicted to reflect the localisation of various cell types within the intestinal crypt and crypt niche. The colour intensity of dots (representing single cells) indicates the level of gene expression, with darker dots representing higher expression levels. Data are presented as normalised, log-transformed counts.

Table 5.1. Spatial Cellular Microdomains Overview

		Number of representative sections profiled	
Domain	**Domain descriptor**	**CD Fistula**	**Non-CD Fistula**
1.	Epithelial stem cell collapse	V:3 X:6	X:1
2.	Epithelial regeneration	V2 X:7	-
3.	Deep fistula-associated stromal remodelling	V:19 X:22	V:4 X:3
4.	Superficial fistula-associated stromal remodelling	V:17 X:22	V:2 X:1

V: Visium, X: Xenium

5.8 Epithelial perturbation in epithelialised CD fistula edge

In this subchapter, I aim to dissect the first cellular microdomain shared across some fistulating CD samples: the apparent loss of crypt-villus architecture and the disappearance of Lgr5+ stem cells toward the fistula edge. By examining a representative colonic section from a patient with fistulating CD (Patient JR 47926_22), I will describe the dynamics of epithelial regeneration and the altered stromal-epithelial crosstalk associated with this pathology. Microscopic examination revealed a notable transition in epithelial architecture towards the fistula, marked by a shift from intact 'normal' appearing crypts to a disordered crypt structure in the pre-fistula region (**Fig. 5.8a**). This morphological alteration was accompanied by distinct differences in the spatial transcriptome, highlighting a shift in gene expression profiled between the normal mucosa and the pre-fistula mucosa.

A gradual decline in the expression of epithelial transcription factors, such as *ATOH1*, *CDX1*, *CDX2* and *HNF1B*, along with the loss of the epithelial stem cell marker Lgr5, was evident towards the pre-

fistula mucosa. Concurrently, gene ontology analysis revealed a reduced pathway of "epithelial cell differentiation", as supported by the decreased expression of mature epithelial markers like *FABP1* and *CA4*.

The collapse of the normal crypt architecture appeared to be facilitated by a collapse in the supporting stromal niche, as evidence by the reduced expression of Stromal 1 and 2 markers (*ADAMDEC1, POSTN, F3* and *BMP5*) and therefore a loss of differentiation signals from pericryptal fibroblasts to epithelium.

The combination of loss of stem cell signal and loss of supporting stromal cells eventually results in entire loss of epithelium. Interestingly, a stromal-like signal emerged in the region of the fistula. This observed cluster was identified as FAS cells. This dysregulated stromal signalling is described further in **Chapter 5.10**.

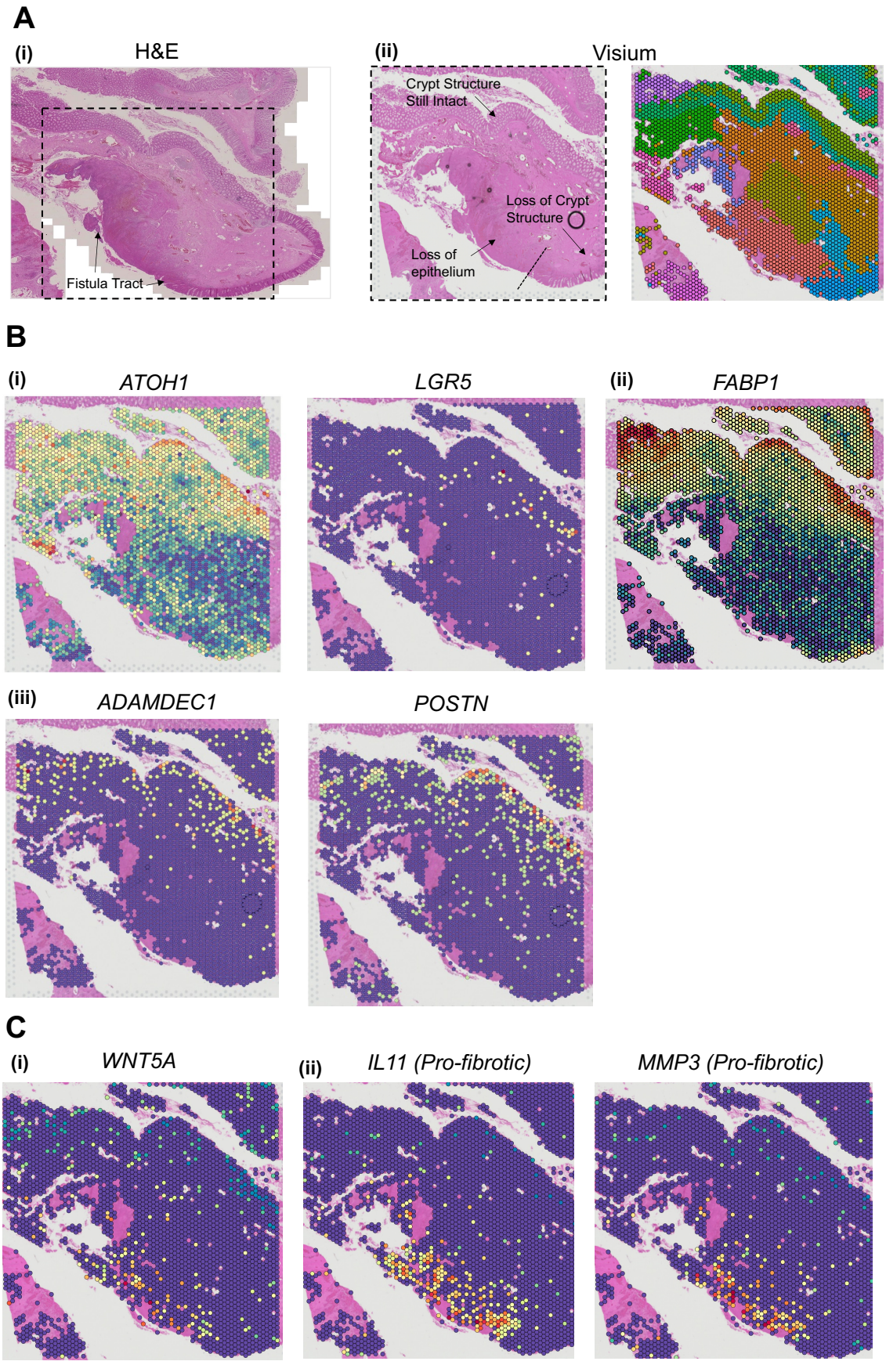

A
(i)
H&E
Fistula Tract
(ii)
Visium
Crypt Structure
Still Intact
Loss of Crypt
Structure
Loss of
epithelium
B
(i)
ATOH1
LGR5
(ii)
FABP1
(iii)
ADAMDEC1
POSTN
C
(i)
WNT5A
(ii)
IL11 (Pro-fibrotic)
MMP3 (Pro-fibrotic)

Figure 5.8: Epithelial Perturbation in Epithelialised CD Fistula by Visium Analysis *(previous page)*.
(A) H&E image of epithelialised CD fistula (Patient JR 47926_22), highlighting the fistula tract (black arrows) (i). Spatial distribution of Visium transcriptome clusters with corresponding H&E image (ii), where black arrows indicate macroscopic alterations in epithelial architecture.
(B) Single gene expression of stem cell markers (*ATOH1, LGR5*) (i) and mature enterocytes (*FABP1*), along with stromal populations (*ADAMDEC1, POSTN*) by Visium.
(C) Expression of genes characterising FAS cells (*WNT5A, IL11, MMP3*) by Visium.
The colour corresponding to relative gene expression levels, with red and orange dots indicating higher expression levels.

To gain deeper insights into these epithelial findings, high-resolution ST techniques were employed to investigate individual cellular interactions. Profiling of the same consecutive section (Patient JR 47926_22) generated 234,530 high-quality individual cells, which were segmented into twenty-one clusters for analysis. This fine-grained analysis shed further light on the aberrant stem cell expression, showing reduced levels of *ASCL2* and *OLFM4* in the pre-fistula mucosa, along with decreased expression of differentiated epithelial cells, such as *BEST4* and *VIPR1*. Changes in the normal crypt-axis distribution of lineage markers were also observed. Namely, the crypt-axis restriction appeared to be lost for *OLFM4*, with its expression extending along the entire length of the crypt in the pre-fistula mucosa. Furthermore, cell proliferation marker *MKi67* was no longer restricted to the TA zone but extended to the crypt top in the pre-fistula region.

High-resolution ST further highlighted the failure of stromal-epithelial crosstalk observed by Visium, with evidence suggesting bidirectional signalling disruptions. Specifically, the reduction in expression of BMP ligands, such as BMP5, in the pre-fistula edge supported a lack of differentiation signals from stromal 2 in this pathological context. Additionally, a reduction in epithelial expression of Indian Hedgehog (*IHH*) in the pre-fistula region and therefore its ability to signal to stromal cells, further demonstrated a failure of epithelial-stromal crosstalk (**Fig. 5.9**).

Immunohistochemical validation of these findings confirmed a gradual reduction in the expression of *LGR5* (ISC) and *F3* (stromal 2), corresponding with the transcriptomic alterations and morphological changes observed (**Fig 5.9b(iii-iv)**). This observation was validated, by IHC, in fistula samples with similar histopathological features (**Appx. Fig A.11**).

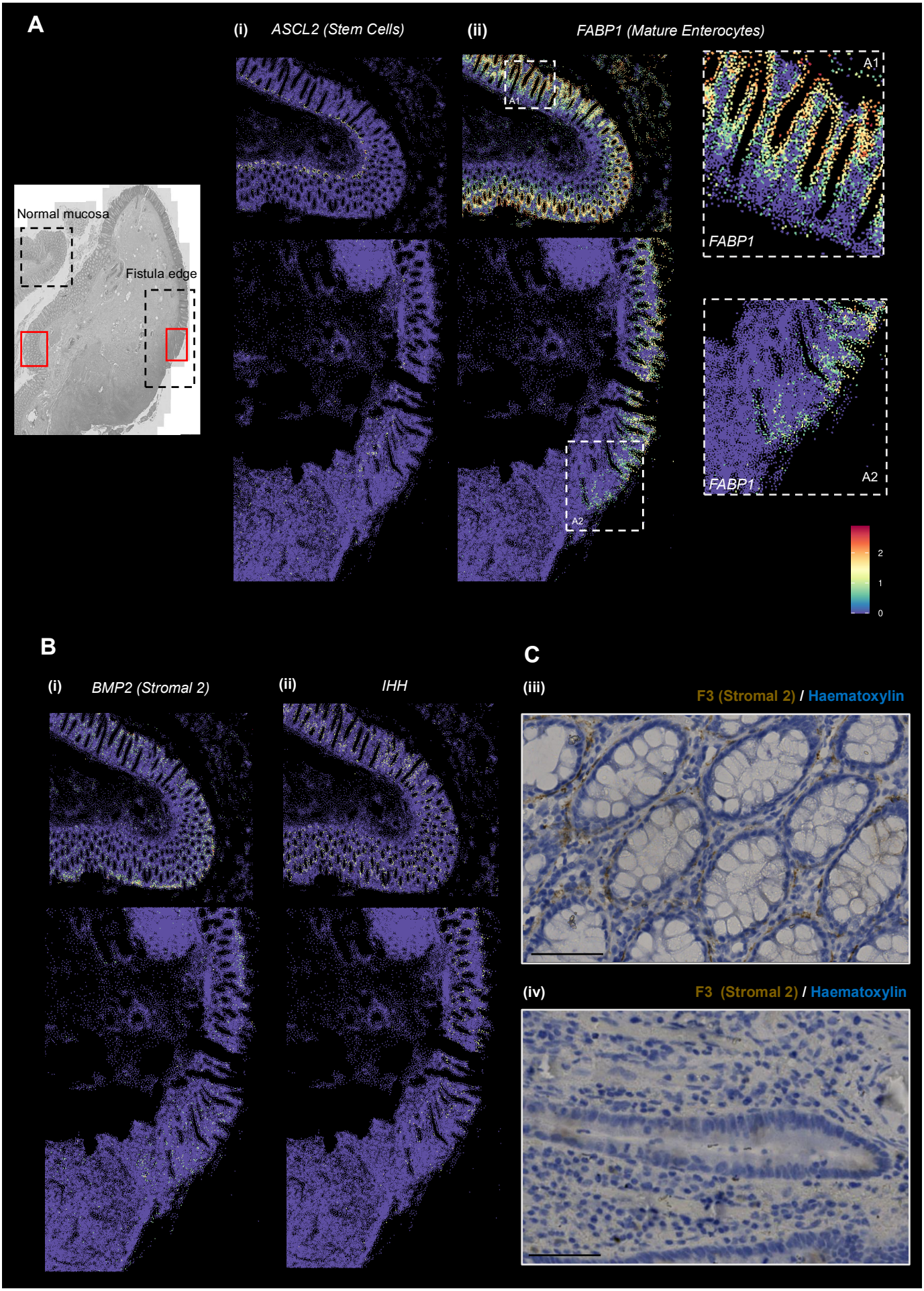

A
(i) ASCL2 (Stem Cells)
(ii) FABP1 (Mature Enterocytes)
Normal mucosa
Fistula edge
A1
A2
FABP1
FABP1
A1
A2
2
1
0
B
(i) BMP2 (Stromal 2)
(ii) IHH
C
(iii)
F3 (Stromal 2) / Haematoxylin
(iv)
F3 (Stromal 2) / Haematoxylin

Figure 5.9: Subcellular Spatial Transcriptomic Characterisation of Epithelial-Stromal Crosstalk Failure in Epithelialised Crohn's Fistula *(previous page)*.
(A) Section of epithelialised CD fistula (Patient JR 47926_22) highlighting areas of normal mucosa and the fistula edge. Xenium spatial distribution of *ASCL2* (i) and *FABP1* (ii) gene expression is shown within these highlighted sections.
(B) Relative gene expression of *BMP2* (i) and *IHH* (ii) in the same regions, with colour gradient indicating the degree of gene activity within the tissue. The colour intensity of dots (representing single cells) indicates the level of gene expression, with darker dots representing higher expression levels. Data are presented as normalised, log-transformed counts. (C) Validation by immunohistochemistry of the highlighted sections from panel (A) (indicated by red boxes), comparing normal mucosa to the fistula edge, with staining for the stromal 2 marker F3 (scale bar is at 50μm). These images are representative of additional validation stains detailed in **Appendix Figure. A.11**.

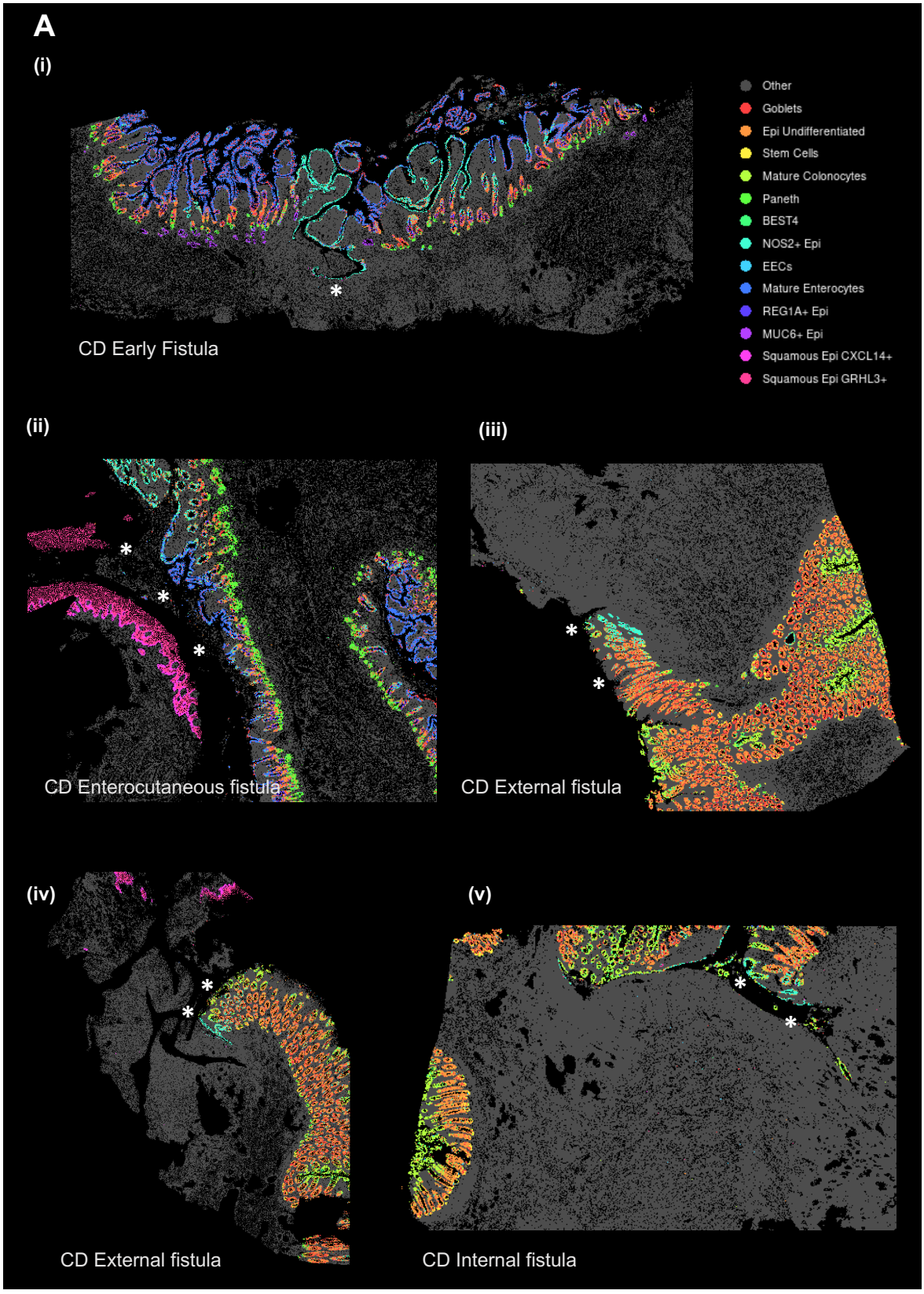


Figure 5.10: Subcellular Spatial Transcriptomic Characterisation of Epithelial Perturbation in Epithelialised Crohn's Fistula. (A) Subcellular spatial transcriptomic distribution of epithelial subclusters in sections of fistulating Crohn's disease (CD) (i) Patient BAY 26534_12, (ii) Patient JR 28598_21 (iii) Patient BAY 113385_21 (iv) Patient BAY 113385_21 (v) Patient BAY 86068_17. The fistula tract is indicated by white asterisks. The different colours represent distinct cell type annotations, illustrating the heterogeneity of epithelial cells in the context of fistula formation. [Epi, Epithelium]

5.9 Epithelial regenerative microdomain in fistulating CD

The perturbations observed in the epithelial mucosa of the fistula tract were consistently replicated among multiple sections of fistulating tissue. In some sections, the tract displayed a partial lining of epithelium followed by a transition to non-epithelialised tissue, while in others, the mucosa prior to the start of the fistula showed disrupted crypt architecture. Fine-grained analysis at the transition in crypt architecture pre-fistula or transition to lack of epithelium within the fistula tract, was performed (**Fig. 5.10**).

This leads on to the second microdomain, where we see an attempt at epithelial regeneration in the absence of Lgr5+ stem cells. Despite the loss of stem cell signal, an effort at repair and regeneration was discerned, characterised by the differential upregulation of regenerative proteins in the pre-fistula region. Regenerating islet-derived proteins, REG1A, REG1B, and REG4, play essential roles in epithelial repair and regeneration (248). Secreted by differentiated intestinal epithelial cells, these multifunctional proteins promote epithelial cell proliferation, migration and survival, resulting in the repair of damaged epithelial surfaces as seen in Crohn's associated inflammation. Secreted REG proteins are downstream effectors of cytokine, IL22. It has been shown that aberrantly expressed *REG1A* has been implicated in inhibiting inflammatory responses and promoting cell proliferation via IL-6/IL22-JAK-STAT3 pathways in patients with IBD (249).

Across several sections, we see a gradient of expression of the regenerating islet-derived proteins, with an increase in their signal correlating to the proximity of the start of the fistula. In this case, in the crypts preceding the fistula region, there was an enrichment of *REG1A*, *REG1B* and *REG4*. However, the regenerative factors appear to lose their crypt-axis restriction with signalling of these regenerative proteins, normally restricted to crypt bases, now expressed right along the crypt length in an attempt to repair mucosa in a compensatory response (**Fig. 5.11a**).

This is coupled with an increase in expression of pro-inflammatory signals, with *DUOX2* and *DMBT1* expression (**Fig. 5.11b**). The increased expression of these genes suggests a switching on of enhanced innate defence mechanisms and active regenerative responses within the pre-fistula mucosa. DUOX2, involved in the production of reactive oxygen species, is required for host defence against invading pathogens. DMBT1 also has antimicrobial properties and confers mucosal protection through the binding and neutralisation of pathogens. Furthermore, DMBT1 promotes epithelial cell proliferation and migration, thereby facilitating wound healing and restoration of the epithelial barrier (250). Both DUOX2 and DMBT1 have been shown to be increased in CD tissue, ameliorating the response to inflammation (251). Variants of both *DUOX2* and *DMBT1* associated with loss of function are linked to CD susceptibility (252, 253). Further, *REG1A* expression, correlated with the upregulation of cytokines CXCL6 and CXCL18, along with a dense immune infiltrate at the fistula edge and deep within the stroma (**Chapter 5.10**).

An increase in the cell signature score for secretory goblet cells (MUC2, FCGBP, ITLN1) was also observed in the pre-fistula region. Goblet cells are specialised epithelial cells that produce and secrete mucus, serving as a physical barrier protecting the mucosa from damage. Additionally, goblet cells have an immunomodulatory role, synthesising various molecules including mucins that can modulate immune cell function and inflammation, participating in tissue repair and resolution of inflammation (254). Therefore, the increase in goblet cells further supports a regenerative response in the pre-fistula region.

Collectively, the upregulation of these sets of genes in the pre-fistula region suggests a dynamic response to tissue damage and inflammation, likely representing the early stages of mucosal healing and regeneration of the fistula tract.

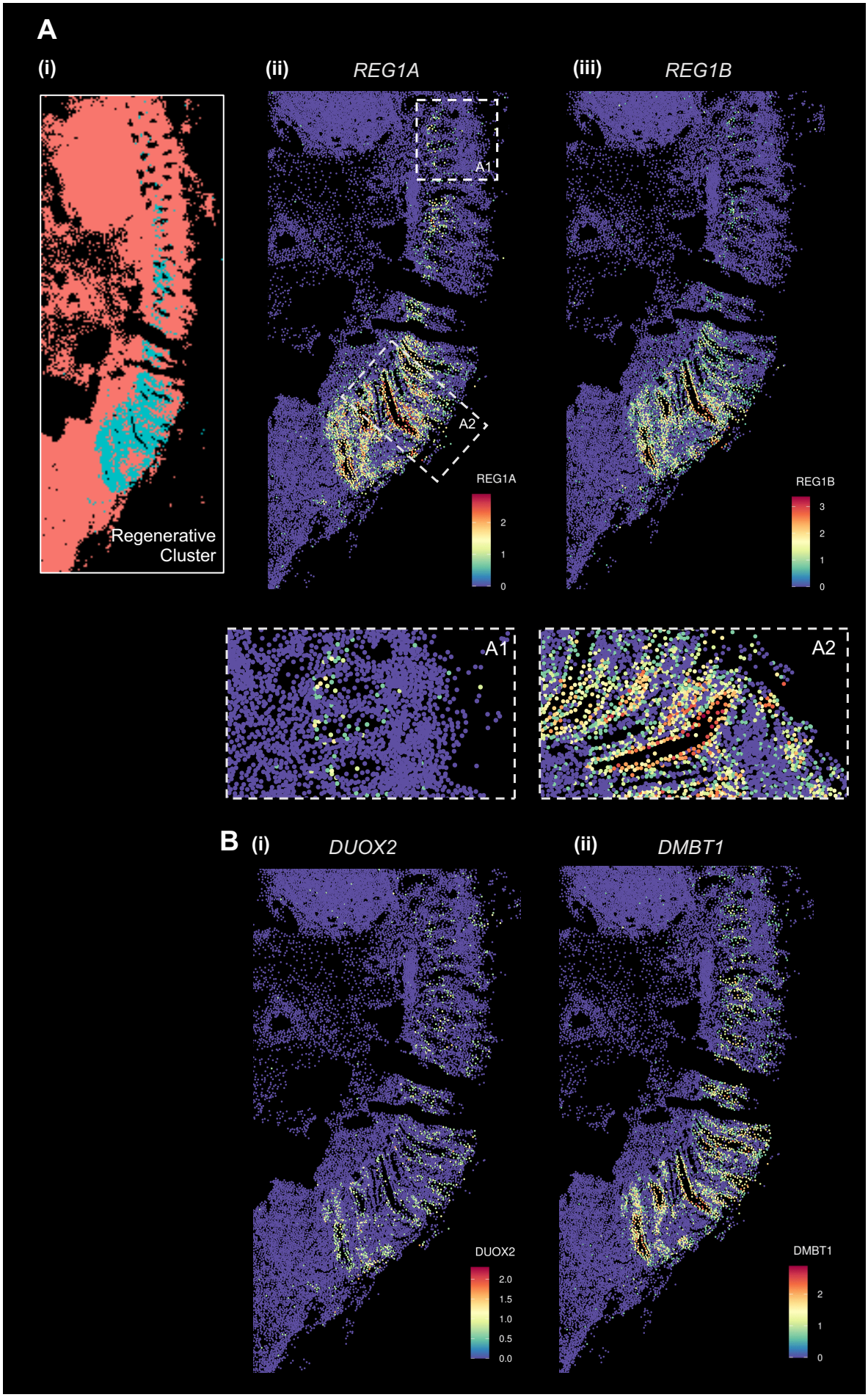

A
(i)
(ii) REG1A
(iii) REG1B
Regenerative Cluster
A1
A2
REG1A
2
1
0
REG1B
3
2
1
0
A1
A2
B
(i) DUOX2
(ii) DMBT1
DUOX2
2.0
1.5
1.0
0.5
0.0
DMBT1
2
1
0

Figure 5.11: Subcellular Spatial Transcriptomic Characterisation of Epithelial Regeneration in Epithelialised Crohn's Fistula *(previous page)*. (A) Section of epithelialised Crohn's disease (CD) fistula (Patient JR 47926_22), showing the subcellular (Xenium) spatial transcriptomic distribution of an identified epithelial "regenerative" cluster (i). Gene expression of regenerative genes is demonstrated with *REG1A* (i), with magnified views of mucosal sections distal (A1) and proximal (A2) to the fistula edge. Additionally, expression of *REG1B* (ii).
(B) Pro-inflammatory signals, *DUOX2* (i) and *DMBT1* (ii) are also presented. Colour gradient indicating the level of relative gene expression. Data are presented as normalised, log-transformed counts.

5.10 Stromal remodelling in fistulating CD

The stromal signal identified in fistulating CD stroma was driven by a subset of fibroblasts, known as FAS, which were originally identified as a fistula-specific population within the single-cell dataset. Projecting this population in spatial terms, this dysregulated stromal signalling could be localised both deep to the fistula tract, constituting the third microdomain and superficially at the fistula edge, representing the fourth microdomain. Further highlighting both the depth and surface interactions of dysregulated fibroblast activity.

WNT signalling driving stromal cell migration in fistulating CD

Aberrant stromal signalling by FAS fibroblasts resulted in an upregulation of morphogenic signalling and a pro-fibrotic response, further contributing to tissue remodelling and fibrosis. The regulation of cellular proliferation is a critical aspect of tissue regeneration. Analysis revealed a close association between WNT signalling, particularly WNT5A, and proliferative stromal signatures, as evidenced by co-localisation with *MKi67* expression in the FAS fibroblasts (**Fig. 5.12**).

Analysis of the cell cycle phase revealed higher proliferation rates in specific clusters, notably in the epithelium, as expected. Significant proliferation was also detected in stromal regions within both external and internal fistulae, with fistulating CD samples displaying more pronounced proliferative activity compared to diverticular fistulae or healthy controls. Specifically, fibroblast proliferation was most concentrated around the fistula tract edges, with high expression of *MKi67*, *STMN1* and *HOPX* enriched at the fistula tract base at subcellular resolution.

Beyond the fistula tract edge, the same fibroblasts exhibited increased IL11 expression. A layering of morphogens with IL11 was observed, where *WNT5A*, typically expressed by fibroblasts around the crypt niche, showed atypical expression beyond the crypt niche and within the deep stromal tissue of the fistula tract. It is known that WNT5A can activate and enhance the fibrotic response of fibroblasts by expressing fibrogenic factors including IL11(255). IL11 can modulate the inflammatory response and promote tissue repair mechanisms, but in excess, it can lead to pathological fibrosis.

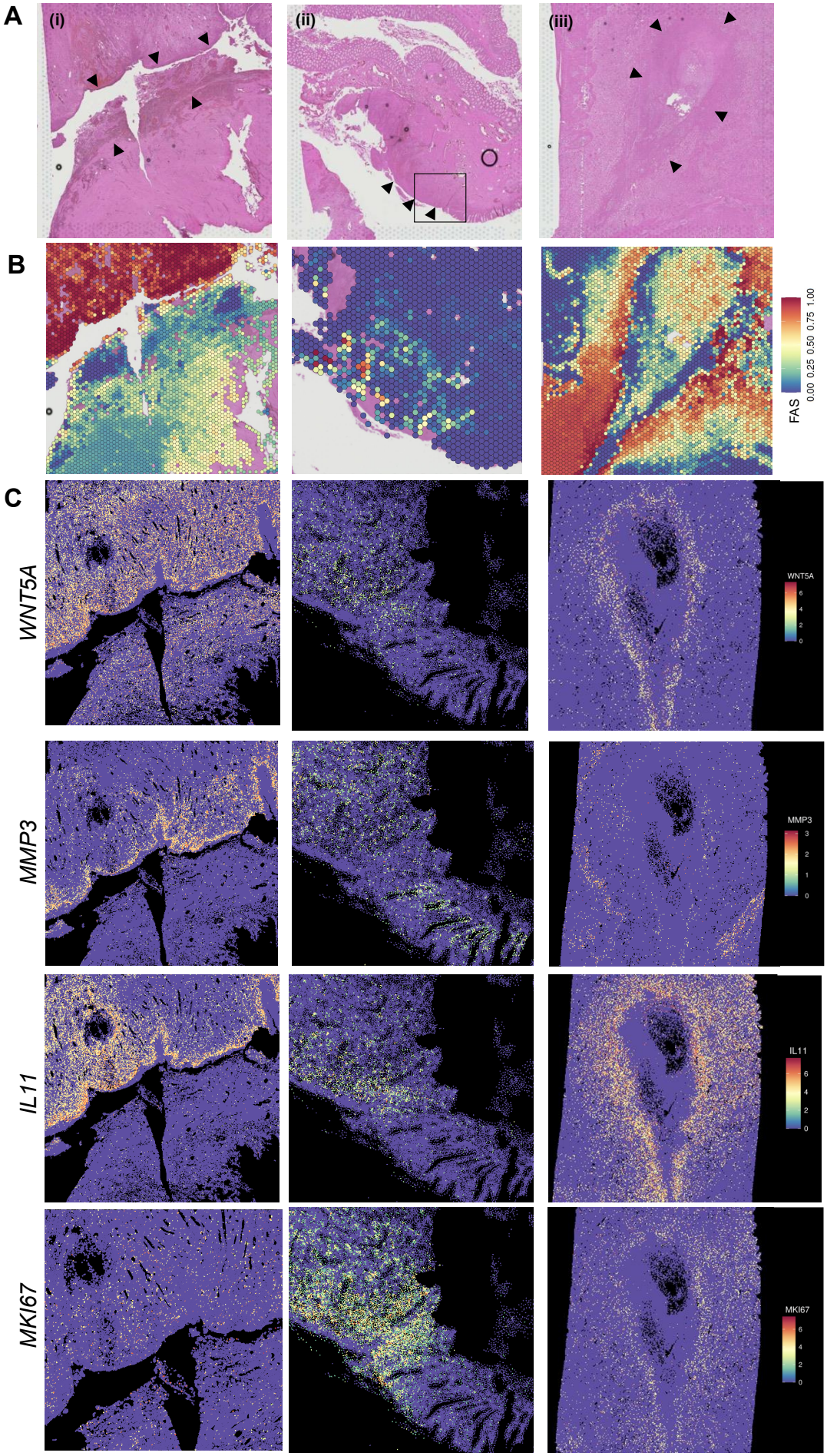

A
(i)
(ii)
(iii)
B
FAS
0.00 0.25 0.50 0.75 1.00
C
WNT5A
WNT5A
MMP3
MMP3
IL11
IL11
MKI67
MKI67

Figure 5.12: Visium and Subcellular Spatial Characterisation of Top Marker Genes in Fistula-Associated Stroma *(previous page)* (A) Corresponding H&E images of Visium sections for Patient JR 23234_23 (i), Patient JR 47926_22 (ii), and Patient BAY 94350_21 (iii). Black arrowheads indicate the fistula tracts, while the black box highlights the magnified region of interest selected for Visium analysis in Patient JR 47926_22.
(B) Visium spatial overlay showing the composite proportion and probability score of fistula-associated stromal (FAS) cells for the corresponding sections.
(C) Subcellular (Xenium) spatial distribution of selected top marker genes (*WNT5A, MMP3, IL11, MKI67*) within the FAS population, as demonstrated in paired patient sections. The colour gradient reflects the degree of relative gene expression. Data are presented as normalised, log-transformed counts.

WNT5A was universally upregulated across the fistulating Visium cohort (**Fig. 5.13**). WNT ligands, WNT1, WNT2, WNT4, and WNT7b, were more specifically upregulated by fistula type. In internal CD fistulae, there was an enrichment in the expression of *WNT1*, whereas *WNT2*, *WNT4*, and *WNT7b* were significantly upregulated in fistulating perianal CD and squamous epithelia of external fistulae. The highest expression of WNT1 ligand was seen in internal fistulating CD, despite overall low expression levels.

TWIST1, identified as a top marker for the FAS population (see **Chapter 4.5.2**), is a known downstream effector of the *WNT1/β-catenin* canonical pathway and *WNT5A* non-canonical pathway. Spatially, *TWIST1* expression was significantly enriched in fistula stroma. *WNT1*, a key activator of the canonical pathway, binds to frizzled receptors 1 (FZD1) and 7 (FZD7) to initiate β-catenin signalling. FZD1 and 7 were highly expressed in fistula-associated fibroblasts.

Additionally, *WNT2* expression, although mainly upregulated in perianal CD fistulae, was also observed in other external fistulae subtypes. Across some tissues, the expression was generalised, while in others, *WNT2* expression was heightened at parts of the fistula tract tunnelling through the submucosa, indicative of fibroblast migration within the fistula. Furthermore, the majority of the planar cell polarity (PCP) pathway components were specifically co-expressed with *WNT4*, indicative of *WNT4* signalling via non-canonical pathways driving fibroblast migration through the fistula. Expression of Cadherin EGF LAG seven-pass G-type receptors, required for PCP signalling, cellular orientation and movement, correlated with the histopathological penetration of the fistulae (**Fig. 5.14**).

To further validate these findings, a multiplexed qPCR experiment was performed on RNA extracted from 84 FFPE samples using the Biomark platform (see **Methods 2.15.2**). The sample cohort encompassed 44 fistulating CD, 11 inflammatory CD, 13 healthy, and 11 diverticular controls (see **Appx. Table A.4**). Validating the findings at a single-cell and spatial level, there was a significant increase in the relative expression of *TWIST1* in internal and external (perianal) fistulating CD compared to healthy controls ($p=0.0027$ and $p=0.0002$ respectively, unpaired T-test) (**Fig. 5.17c(v)**). Perturbations in the WNT pathways held true. FZD7 was significantly upregulated in fistulating samples. The relative expression of FZD5 in fistulating CD and diverticular disease was significantly downregulated compared to healthy controls. While not reaching significance, FZD6 was downregulated in fistulating samples (**Fig. 5.13d**).

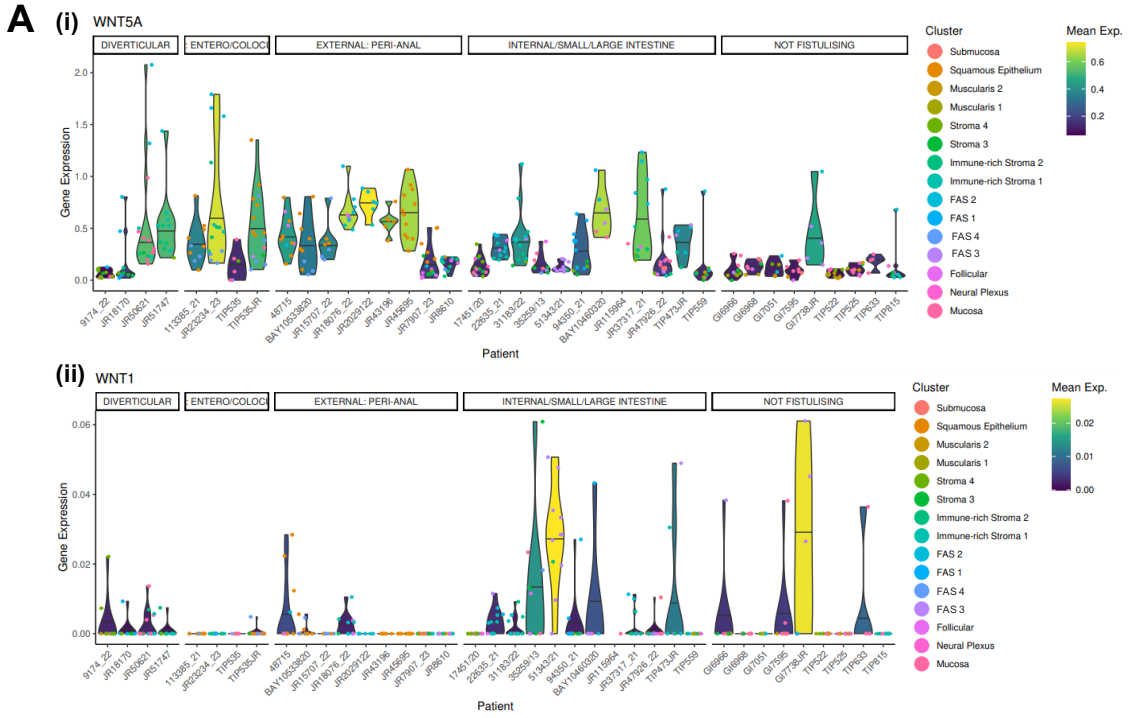
Dissection of these dysregulated WNT signalling pathways points to potential mechanisms for the fistula microdomains captured spatially. Overexpression of WNT ligands, WNT1, WNT2, and WNT4 in FAS fibroblasts, coupled with increased expression of FZD1 and 7, leads to hyperactivation of the canonical WNT/β-catenin signalling pathway. This promotes FAS fibroblast proliferation and ECM production. *TWIST1*, a transcription factor upregulated in these fibroblasts, further exacerbates the process by driving FAS fibroblasts toward a more invasive and fibrotic phenotype, enhancing their ability to migrate into surrounding tissue and produce ECM, facilitating the formation of fibrotic tissue. *TWIST1* can also potentiate the expression of WNT ligands, creating a positive feedback loop and potentially influencing the expression of WNT receptors, including FZD7, further promoting WNT signalling (222).

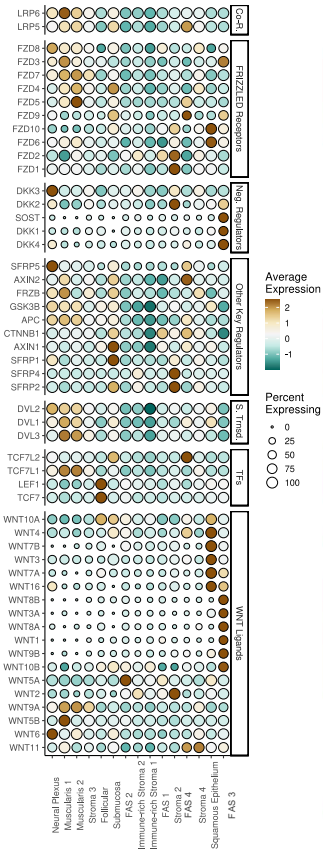
In fistulating perianal CD, hyperactivation of WNT4 signalling via non-canonical pathways, which is crucial for maintaining cell polarity and directing cell movement, leads to stromal cell migration and tissue disorganisation. Additionally, the downregulation of FZD5 and FZD6 in FAS fibroblasts results in inadequate support for the epithelial crypt niche. This failure in stromal-epithelial crosstalk

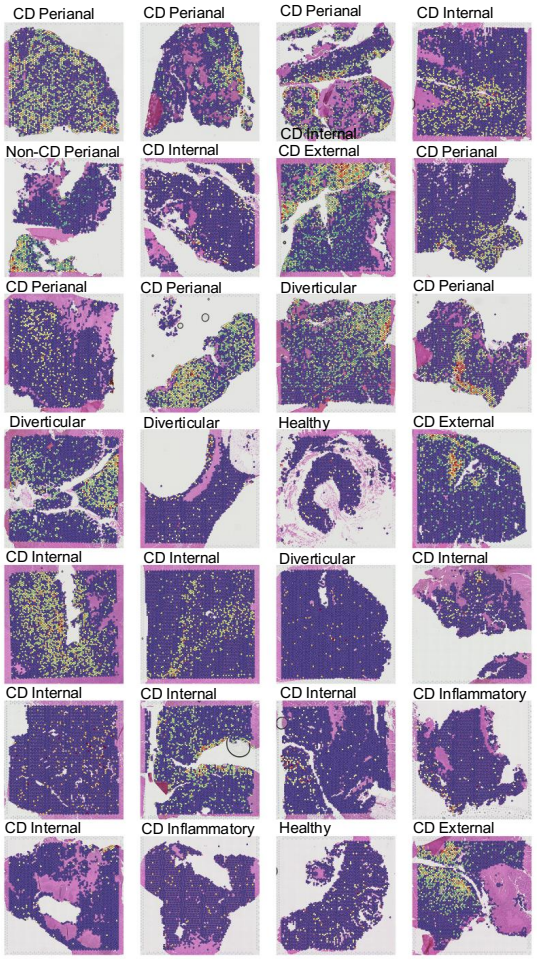
disrupts the maintenance and renewal of the epithelial barrier, leading to its breakdown and perpetuating inflammation, which in turn exacerbates tissue damage and fibroblast activation.

FZD5 also mediates non-canonical Wnt signalling. Overexpression of *WNT5A* non-canonical signalling may be potentiated by reduced FZD5 expression, as FZD5 is a recognised regulator of non-canonical WNT signalling through inhibitory signals (256). Furthermore, an increase in other frizzled receptors (e.g., FZD1) and co-receptors (e.g., ROR2) (**Fig. 5.13b**) was observed, which may be preferentially activated in the presence of FZD5 downregulation, further potentiating pathways that promote fibroblast migration and invasion.

As these activated fibroblasts migrate and penetrate surrounding tissues, they form fistulae. This abnormal tissue remodelling is stabilised by ongoing inflammation and fibrosis, thereby hindering normal wound healing processes and sustaining of fistula tract.


A
(i) WNT5A
DIVERTICULAR
ENTERO/COLOC
EXTERNAL: PERI-ANAL
INTERNAL/SMALL/LARGE INTESTINE
NOT FISTULISING
Gene Expression
Patient
Cluster
Submucosa
Squamous Epithelium
Muscularis 2
Muscularis 1
Stroma 4
Stroma 3
Immune-rich Stroma 2
Immune-rich Stroma 1
FAS 2
FAS 1
FAS 4
FAS 3
Follicular
Neural Plexus
Mucosa
Mean Exp.
(ii) WNT1


B
Co-R
FRIZZLED Receptors
Neg. Regulators
Other Key Regulators
S. Tmed
TFs
WNT Ligands
Average Expression
Percent Expressing


C
CD Perianal
CD Internal
CD External
Non-CD Perianal
Diverticular
Healthy
CD Inflammatory

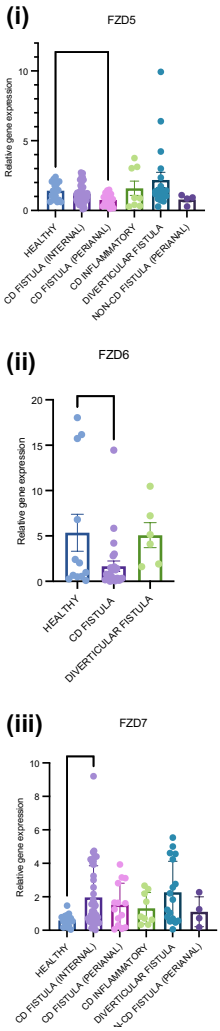
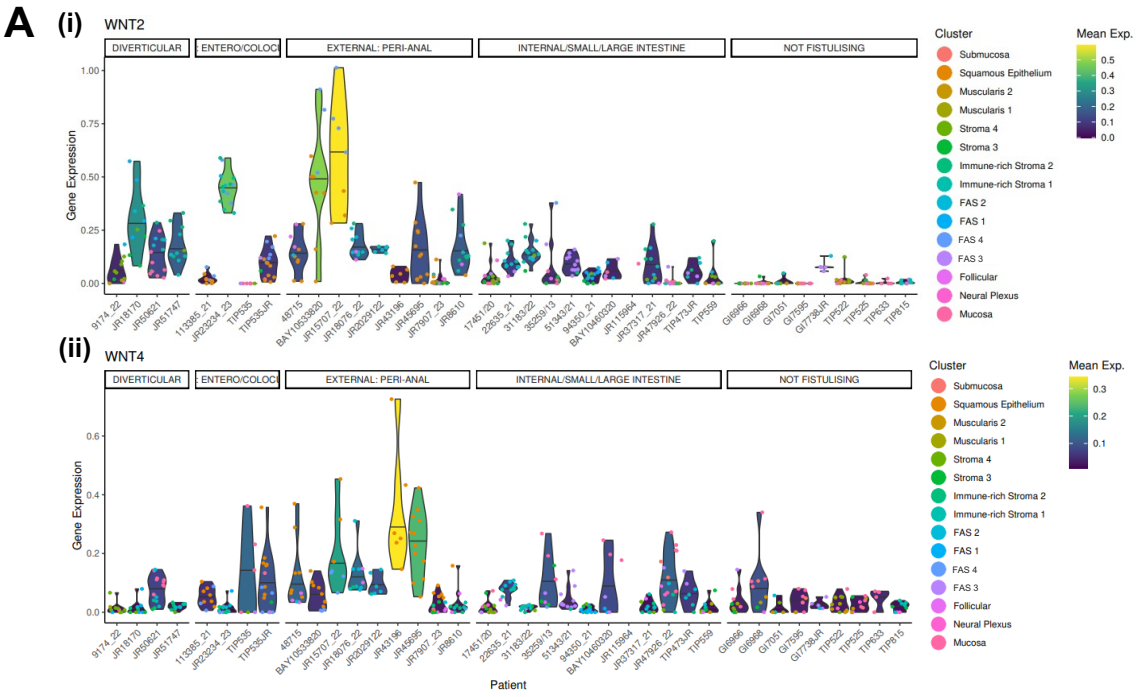
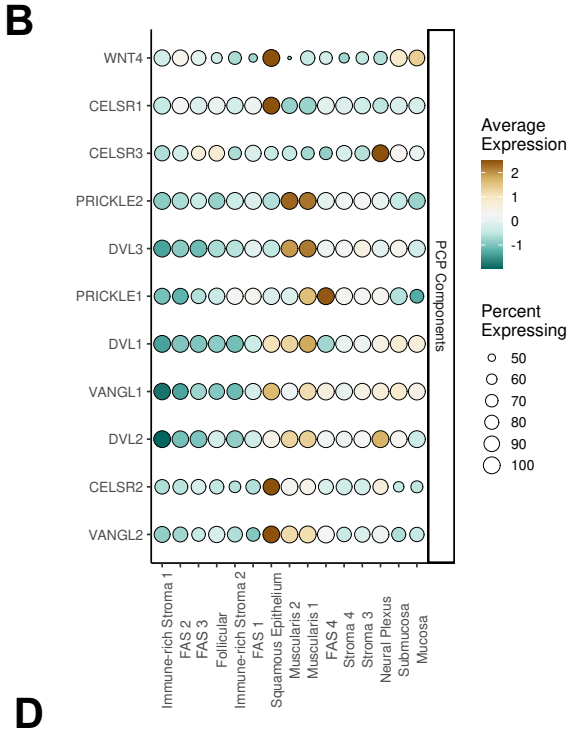

D
(i) FZD5
(ii) FZD6
(iii) FZD7
Relative gene expression
HEALTHY
CD FISTULA (INTERNAL)
CD FISTULA (PERIANAL)
CD INFLAMMATORY
DIVERTICULAR FISTULA
NON-CD FISTULA (PERIANAL)

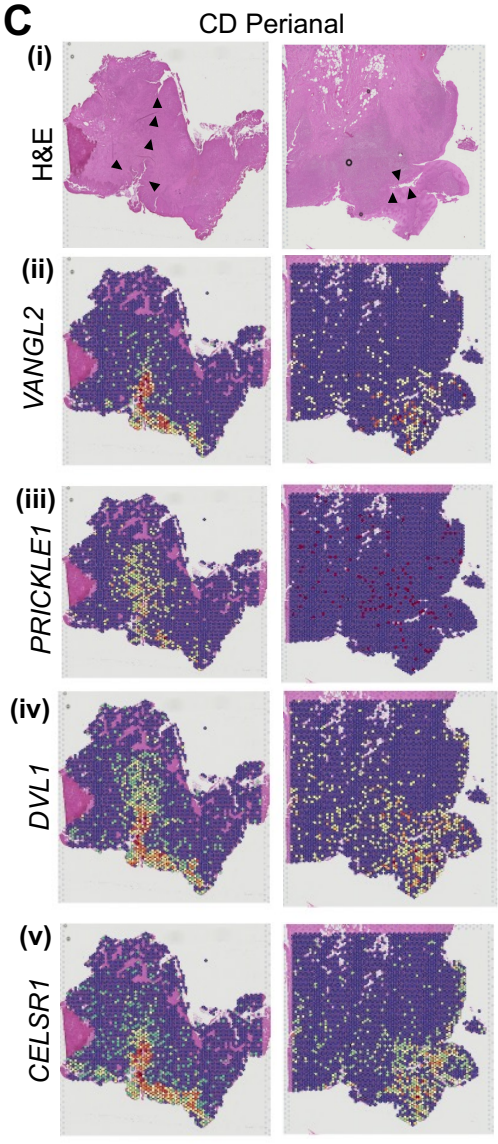
Figure 5.13: Wnt Pathway in Fistulating Crohn's Disease *(previous page)*
(A) Violin plots illustrating the expression levels of *WNT5A* (i) and *WNT1* (ii) across individual patient samples in the Visium cohort, stratified by fistula type and annotated by cell type. The distribution and density of gene expression are depicted across different fistula phenotypes.
(B) Bubble plot showing the expression of *WNT* pathway genes across the total Visium metaclusters. The size of each dot represents the percentage of cells expressing the gene, while the colour gradient indicates the intensity of gene expression.
(C) Spatial transcriptome overlay depicting *WNT4* gene expression across the entire Visium cohort. The colour gradient reflects the relative level of gene expression, with red and orange colours indicating higher expression levels. (See **Appendix Table A.3-4** for patient IDs)
(D) Relative gene expression of frizzled receptors (FZD5 (i), FZD6 (ii), FZD7 (iii)) was assessed using quantitative PCR) in a validation cohort (Biomark data; see **Appx. Table A.3-4**). RNA was extracted from 84 samples of formalin-fixed, paraffin-embedded tissue blocks, including fistulating (CD and non-CD), inflammatory CD, and healthy controls. Gene expression data are presented as mean values with error bars representing the standard error of the mean. Non-parametric data were evaluated using an unpaired t-test, with statistical significance set at $p < 0.05$.

Further analysis was performed using a generalised linear model to adjust for EpCAM expression in the same samples. Frizzled receptor expression was compared between fistula subtypes and healthy controls, with significance determined at $p < 0.05$. The analysis revealed the following statistically significant results:

Reduced FZD6 expression compared to healthy controls: Internal CD fistula (fold change = -1.82, $p = 0.0048$), perianal CD fistula (fold change = -1.60, $p = 0.04$), and diverticular fistula (fold change = -1.79, $p = 0.02$).
Increased FZD7 expression compared to healthy controls: Internal CD fistula (fold change = 0.83, $p = 0.01$), external CD fistula (fold change = 1.23, $p = 0.03$), and diverticular fistula (fold change = 1.02, $p = 0.01$).
Increased FZD5 expression compared to healthy controls: Diverticular fistula (fold change = 0.69, $p = 0.03$).


A
(i)
WNT2
DIVERTICULAR
ENTERO/COLOC
EXTERNAL: PERI-ANAL
INTERNAL/SMALL/LARGE INTESTINE
NOT FISTULISING
Gene Expression
Cluster
Submucosa
Squamous Epithelium
Muscularis 2
Muscularis 1
Stroma 4
Stroma 3
Immune-rich Stroma 2
Immune-rich Stroma 1
FAS 2
FAS 1
FAS 4
FAS 3
Follicular
Neural Plexus
Mucosa
Mean Exp.
(ii)
WNT4
Patient


B
WNT4
CELSR1
CELSR3
PRICKLE2
DVL3
PRICKLE1
DVL1
VANGL1
DVL2
CELSR2
VANGL2
PCP Components
Average Expression
Percent Expressing
Immune-rich Stroma 1
FAS 2
FAS 3
Follicular
Immune-rich Stroma 2
FAS 1
Squamous Epithelium
Muscularis 2
Muscularis 1
FAS 4
Stroma 4
Stroma 3
Neural Plexus
Submucosa
Mucosa


C
CD Perianal
(i)
H&E
(ii)
VANGL2
(iii)
PRICKLE1
(iv)
DVL1
(v)
CELSR1

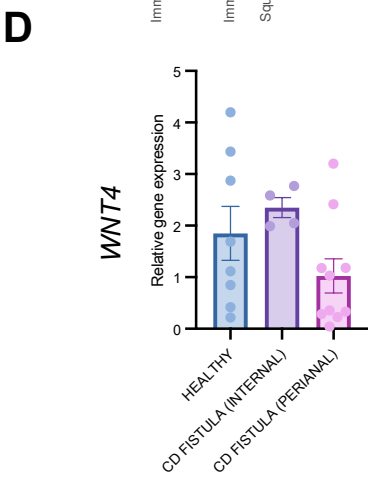

D
WNT4
Relative gene expression
HEALTHY
CD FISTULA (INTERNAL)
CD FISTULA (PERIANAL)

Figure 5.14: PCP Pathway in Fistulating Crohn's Disease *(previous page)*
(A) Violin plots illustrating the expression levels of *WNT2* (i) and *WNT4* (ii) across individual patient samples in the Visium cohort, stratified by fistula type and annotated by cell type. The distribution and density of gene expression are shown for different fistula phenotypes.
(B) Bubble plot displaying the expression of PCP pathway genes across the Visium metaclusters. The size of each dot indicates the percentage of cells expressing the gene, while the colour gradient represents the intensity of gene expression.
(C) Spatial transcriptome overlay depicting the expression of selected PCP pathway genes (*VANGL2* (ii), *PRICKLE1* (iii), *DVL1* (iv), *CE*LSR1 (v)) across two examples of perianal CD fistula (left: Patient BAY 48715_20, right: Patient JR 7909_23) with corresponding H&E images (i). The colour gradient reflects the relative level of gene expression, with red and orange indicating higher expression levels. Black arrowheads denote the fistula tract.
(D) Relative gene expression of *WNT4* was assessed using quantitative PCR (qPCR) in a validation cohort (Biomark data; see **Appx. Table A.3-4**). Gene expression data are presented as mean values with error bars representing the standard error of the mean (SEM). Non-parametric data were evaluated using an unpaired t-test, with statistical significance set at $p < 0.05$.

Further analysis was performed using a generalised linear model to adjust for EpCAM expression in the same samples. *WNT4* expression was compared between fistula subtypes and healthy controls, with significance determined at $p < 0.05$. The analysis revealed the following results:

Reduced WNT4 expression compared to healthy controls: Diverticular fistula (fold change = -0.40, $p = 0.03$).
Increased WNT4 expression compared to healthy controls: External CD fistula (fold change = 0.15, $p = 0.54$), Perianal CD fistula (fold change = 0.12, $p = 0.54$).

Tissue remodelling and pro-fibrosis:

The Visium analysis uncovered significant implications of the FAS population, particularly regarding the dysregulation of matrix metalloproteinases, such as MMP2, 3, and 9, which play crucial roles in the physical remodelling of fistula tissue. Specifically, the presence of MMP3 expression was heightened near the muscularis propria, indicating potential tunnelling of the fistula tract through the muscle layer. This aberrant MMP signal may stem from altered expression of transcription factors like *FOSL1*, a downstream target of *WNT* signalling implicated in MMP regulation (257). Concurrently, increased cell proliferation was observed at the edge of the fistula, further contributing to tissue remodelling.

Five fistula-associated macrophage clusters were identified at subcellular ST resolution, with pro-fibrotic and pro-inflammatory functions (**Fig. 5.15**). SPP1 expressing macrophages were found to densely populate the fistula tract, providing an additional source of MMPs. It is known that SPP1 can stimulate macrophages to secrete MMPs and drive fibroblast activation thereby contributing to ECM remodelling and fibrosis (258). In addition to their pro-fibrotic role, SPP1$^+$ macrophages also recruit macrophages and neutrophils through chemokine expression (CXCL5, CXCL4, CXCL3). These macrophages were one of the strongest signals in fistulating tissue with less enrichment in inflammatory CD controls.

MMP9$^+$ macrophages localised just under the layer of SPP1$^+$ macrophages and co-localised with FAS cells. While also expressing some SPP1, these macrophages were characterised by high MMP expression and pro-fibrotic capabilities.

LYVE1$^+$ macrophages were identified as being enriched near vasculature within the fistula stroma, more distal to the fistula tract. Data indicates that LYVE1$^+$ macrophages play an important role in the morphogenesis or remodelling of the lymphatic vasculature (259). Therefore, these perivascular macrophages within fistula tissue may contribute to tissue remodelling and neovascularisation.

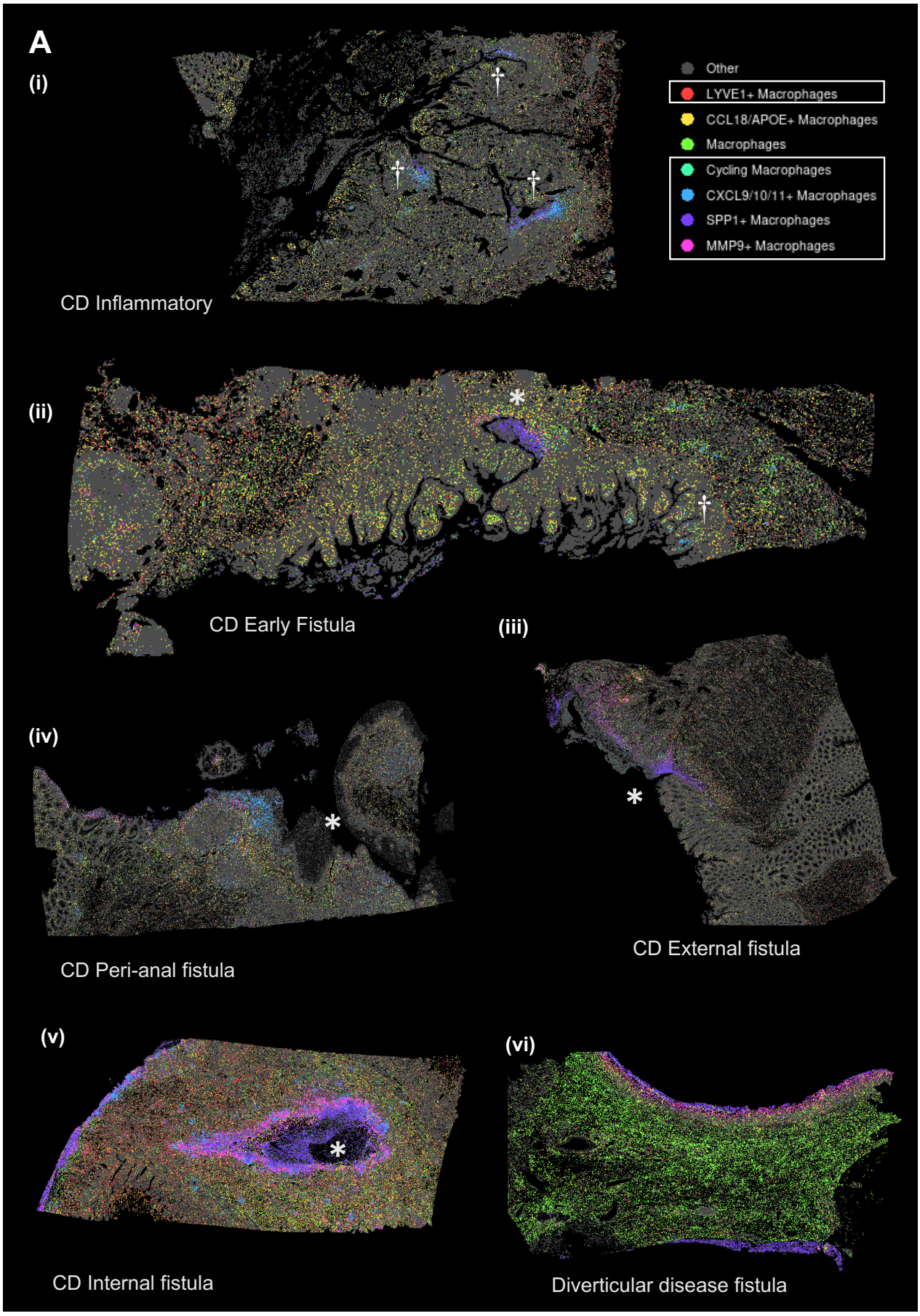

A
(i)
Other
LYVE1+ Macrophages
CCL18/APOE+ Macrophages
Macrophages
Cycling Macrophages
CXCL9/10/11+ Macrophages
SPP1+ Macrophages
MMP9+ Macrophages
CD Inflammatory
(ii)
CD Early Fistula
(iii)
CD External fistula
(iv)
CD Peri-anal fistula
(v)
CD Internal fistula
(vi)
Diverticular disease fistula

Figure 5.15: Subcellular Spatial Transcriptomic Characterisation of Fistula Associated Macrophages (*previous page*) (A) Subcellular spatial transcriptomic distribution of macrophage subclusters in sections of fistulating Crohn's disease (CD), inflammatory CD and diverticular fistulae. The fistula tract is indicated by * and fissuring ulcers by †. The different colours represent distinct macrphage subcluster annotations, illustrating the heterogeneity of macrophage populations in the context of fistula formation. (i) Patient JR 32939_22 (ii) Patient BAY 26534_12 (iii) Patient BAY 113385_21 (iv) Patient JR 7909_23 (v) Patient BAY 94350_21 (vi) Patient JR 18170_21

Immune-rich stromal environment:

A subset of FAS fibroblasts was found to promote a robust immune response within the deep tissue stroma, characterised by strong expression of CXCL13 but not CCL19 (**Appx. Fig. A.12**). This immune microenvironment led to the recruitment of numerous B-cells and T-cells to the deep tissue

areas of the fistula. At the fistula edge, particularly in non-epithelialised tracts, a pronounced neutrophilic signature was observed, characterised by the expression of *CXCL8* and *HCAR3*, indicative of granulation tissue. This layer of granulocytes was found in close proximity to the diverse

population of macrophages and FAS fibroblasts, a consistent feature across multiple sections (**Fig. 5.16**).

Further investigation into niche and crosstalk dynamics is required to elucidate the involvement of stromal-myeloid interactions in fistula pathology. Existing literature suggests that variations in myeloid-stromal crosstalk may contribute to the development of fibrotic and destructive fibroblasts, senescence, and epigenetic memory, which are implicated in perianal CD fistula pathogenesis (260).

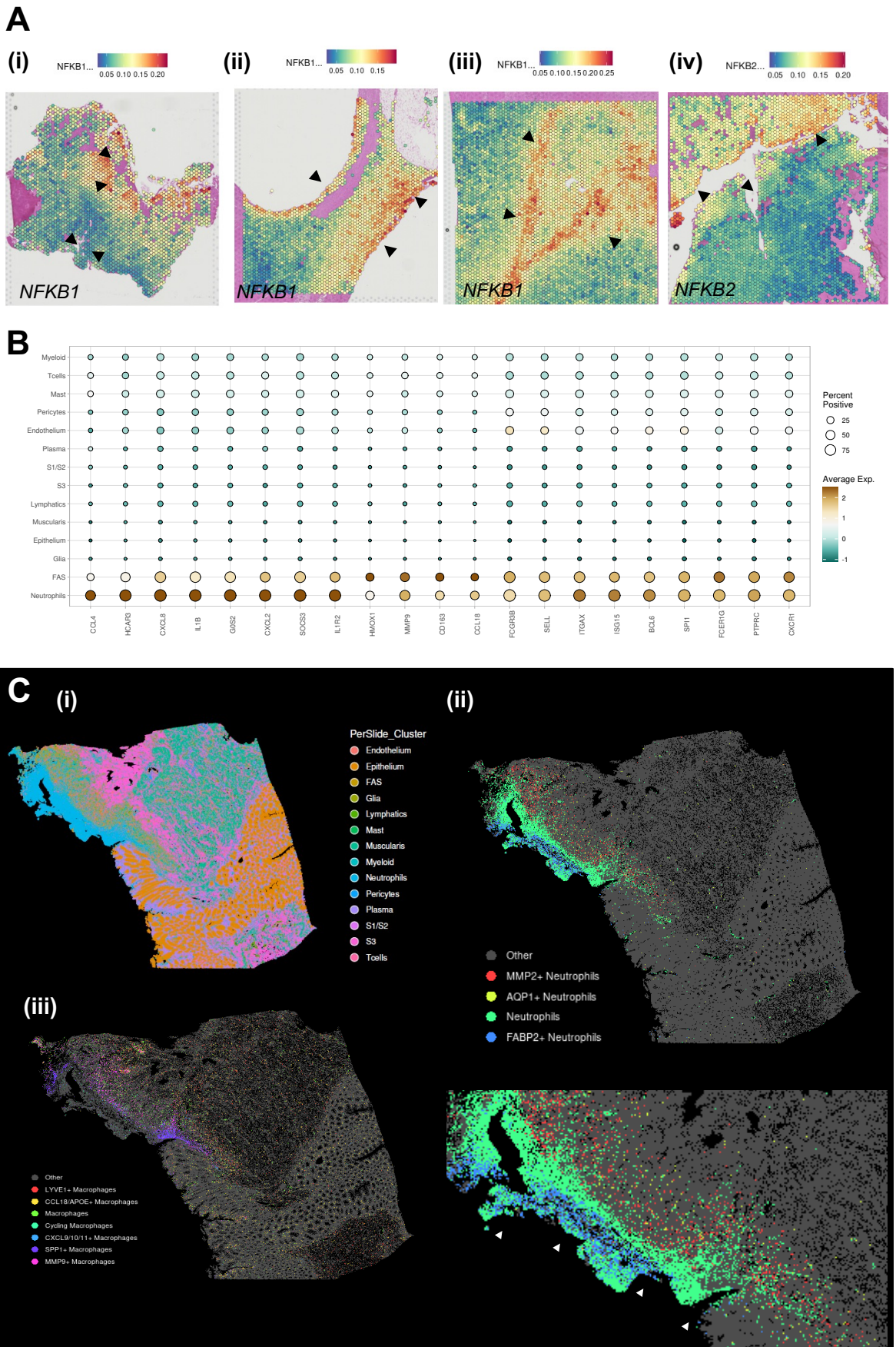

A
(i)
NFKB1...
0.05 0.10 0.15 0.20
NFKB1
(ii)
NFKB1...
0.05 0.10 0.15
NFKB1
(iii)
NFKB1...
0.05 0.10 0.15 0.20 0.25
NFKB1
(iv)
NFKB2...
0.05 0.10 0.15 0.20
NFKB2
B
Myeloid
Tcells
Mast
Pericytes
Endothelium
Plasma
S1/S2
S3
Lymphatics
Muscularis
Epithelium
Glia
FAS
Neutrophils
CCL4
HCAR3
CXCL8
IL1B
G0S2
CXCL2
SOCS3
IL1R2
HMOX1
MMP9
CD163
CCL18
FCGR3B
SELL
ITGAX
ISG15
BCL6
SPI1
FCER1G
PTPRC
CXCR1
Percent Positive
25
50
75
Average Exp.
2
1
0
-1
C
(i)
PerSlide_Cluster
Endothelium
Epithelium
FAS
Glia
Lymphatics
Mast
Muscularis
Myeloid
Neutrophils
Pericytes
Plasma
S1/S2
S3
Tcells
(ii)
Other
MMP2+ Neutrophils
AQP1+ Neutrophils
Neutrophils
FABP2+ Neutrophils
(iii)
Other
LYVE1+ Macrophages
CCL18/APOE+ Macrophages
Macrophages
Cycling Macrophages
CXCL9/10/11+ Macrophages
SPP1+ Macrophages
MMP9+ Macrophages

Figure 5.16: Spatial Characterisation of Neutrophils and Regulatory Mechanisms at the Fistula Edge *(previous page)*. (A) Spatial distribution of *NFKB1* expression across fistulating tissues, visualised using Visium spatial transcriptomics. Samples include perianal CD (Patient BAY 48715_20) (i), diverticular fistula (Patient JR 18170_21) (ii), internal CD (Patient BAY 94350_21) (iii), and external CD (Patient 23234_23) (iv). Black arrowheads indicate the location of the fistula tract.
(B) Dot-plot heatmap illustrating the expression of top neutrophil marker genes across different clusters. The colour gradient reflects the average expression levels, while the size of each dot represents the percentage of cells within each cluster that express the corresponding marker.
(C) Subcellular spatial distribution of transcriptomic clusters in a representative perianal CD fistula (Patient BAY 113385_21) (i). The panel highlights the presence of exclusive neutrophil subclusters (ii), with a magnified view indicating the proximity of these subclusters to the fistula tract edge (white arrowheads) and the adjacent macrophage subclusters (iii).

Heterogeneity of FAS at subcellular resolution:

Subcellular resolution ST analysis enabled spatial localisation of FAS subsets, shedding light on their potential functions. Additional examination at the single-cell level delineated FAS subsets more distinctively: FAS Cycling, FAS MMP1/3$^+$, FAS MMP1/3$^+$IL11$^+$, and Stromal 3/FAS. These subsets exhibited unique spatial distributions around the fistula tract, with those high in MMPs (MMP 1 and MMP3) clustering around the tract's edge. Additionally, FAS cycling cells, marked by *MKi67* expression, were prevalent in these regions, consistent with increased proliferation and ECM remodelling activities.

A specific subset of fibroblasts within the FAS population, identified as Stromal 3/FAS, predominantly occupies the deeper stromal layers distal to the fistula tract, often coexisting with FAS MMP1/3$^+$ cells. This discovery is supported by findings at the single-cell level, where analysis based on transcription factor networks indicated that the FAS subset branches off independently but shares a close relationship with Stromal 3 (submucosal fibroblasts). Additionally, individual classification and lineage prediction of fibroblast cells based on transcription factors reveal that many cells within the FAS subset align with the Stromal 3 lineage, with fewer cells associated with other stromal lineages (**Fig. 5.17**). The expression of these subsets was more pronounced in the more advanced fistula but still present in similar distributions but to a lesser extent in the early fistulae.

Moreover, certain subsets of FAS fibroblasts exhibited irregular patterns of lineage markers, displaying characteristics similar to mucosal fibroblasts (ADAMDEC1$^+$ phenotype) near the fistula

tract. This suggests a potential regenerative response by ADAMDEC1^{+} FAS fibroblasts, aimed at restoring mucosal tissue in the affected region.

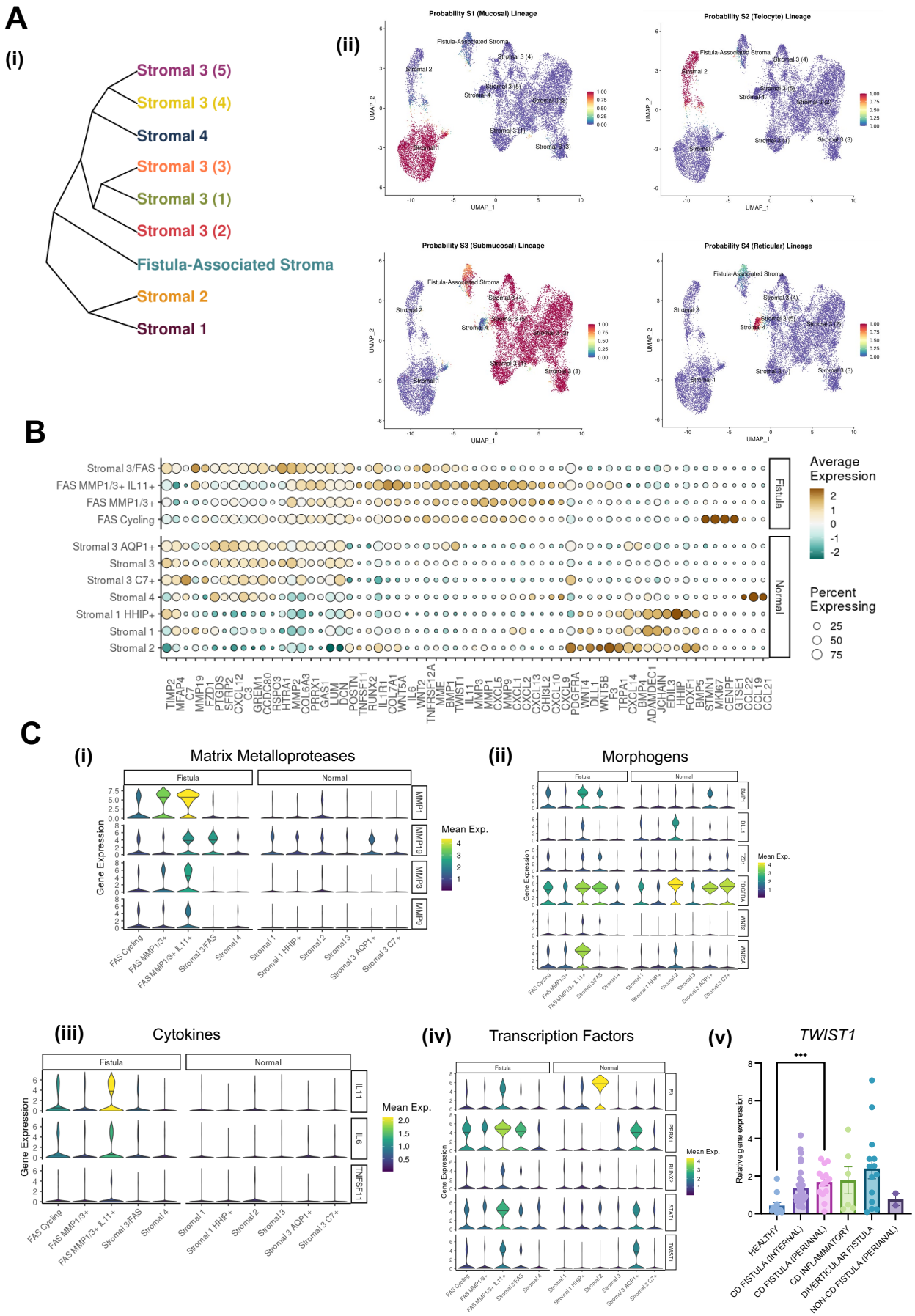

A
(i)
Stromal 3 (5)
Stromal 3 (4)
Stromal 4
Stromal 3 (3)
Stromal 3 (1)
Stromal 3 (2)
Fistula-Associated Stroma
Stromal 2
Stromal 1
(ii)
Probability S1 (Mucosal) Lineage
Probability S2 (Telocyte) Lineage
Probability S3 (Submucosal) Lineage
Probability S4 (Reticular) Lineage
UMAP_1
UMAP_2
B
Stromal 3/FAS
FAS MMP1/3+ IL11+
FAS MMP1/3+
FAS Cycling
Stromal 3 AQP1+
Stromal 3
Stromal 3 C7+
Stromal 4
Stromal 1 HHIP+
Stromal 1
Stromal 2
Fistula
Normal
Average Expression
Percent Expressing
25
50
75
TIMP2 MFAP4 C7 MMP19 FZD1S SFRP5 CXCL12 C3 GREM1 CCDC80 RSPO3 HTRA1 MMP2 COL6A3 PRRX1 GAS1 LUM DCN POSTN TNFSF11 RUNX2 IL1R1 COL7A1 WNT5A IL6 WNT2 TNFRSF12A MME BMP1 TWIST1 IL11 MMP3 MMP1 CXCL5 MMP9 CXCL1 CXCL3 CXCL13 CHI3L2 CXCL10 CXCL9 PDGFRA WNT4 DLL1 WNT5B F3 TRPA1 CXCL14 BMP4 ADAMDEC1 JCHAIN EDIL3 FOXF1 BMP5 STMN1 MKI67 CENPF GTSE1 CCL22 CCL19 CCL21
C
(i)
Matrix Metalloproteases
(ii)
Morphogens
(iii)
Cytokines
(iv)
Transcription Factors
(v)
TWIST1
Gene Expression
Mean Exp.
MMP1
MMP19
MMP3
MMP9
IL11
IL6
TNFSF11
F3
PDGFRA
RUNX2
STAT1
TWIST1
WNT2
WNT5A
FAS Cycling
FAS MMP1/3+
FAS MMP1/3+ IL11+
Stromal 3/FAS
Stromal 4
Stromal 1
Stromal 1 HHIP+
Stromal 2
Stromal 3
Stromal 3 AQP1+
Stromal 3 C7+
Relative gene expression
***
HEALTHY
CD FISTULA (INTERNAL)
CD FISTULA (PERIANAL)
CD INFLAMMATORY
DIVERTICULAR FISTULA
NON-CD FISTULA (PERIANAL)

Figure 5.17: Lineage, Regulation, and Subclassification of Fistula-Associated Stroma Cluster *(previous page)*

(A) Lineage tree illustrating the hierarchical clustering of fibroblast populations based on transcription factors (TFs) derived from scRNA-seq data (i). The corresponding UMAP embedding visualises individual fibroblast cells and their predicted lineage relationships, as inferred from TF expression profiles (ii).

(B) Bubble plot showing the subclassification of fibroblasts into fistula-associated stromal subsets (FAS) and normal-appearing fibroblasts. This classification is based on the expression of top marker genes, as determined from Xenium single-cell data.

(C) Violin plots displaying the transcriptional profiles of key markers within FAS subsets and normal-appearing fibroblasts. These markers include matrix metalloproteinases (MMPs) (i), morphogens (ii), cytokines (iii), and TFs (iv). The colour gradient represents the average expression levels, with lighter colours indicating higher gene expression. Additionally, the relative expression of *TWIST1* was assessed using qPCR in a validation cohort (Biomark data; see **Appendix Table A.3-4**). Gene expression data are presented as mean values, with error bars representing the standard error of the mean (SEM). Statistical analysis was performed using an unpaired t-test for non-parametric data, with significance set at $p < 0.05$ (v).

5.11 Subcellular resolution ST identifies precursor to fistula development in other CD phenotypes

In the clinical progression of complicated CD, chronic inflammation initially results in an inflammatory (non-stricturing, non-fistulating) phenotype. However, in the context of persistent, uncontrolled inflammation, fibrosis and strictures may develop, eventually leading to fistula formation. Histopathologically, an established precursor to fistula formation is fissuring ulceration, as detailed in **Chapter 5.4**.

To explore whether the signals identified in fistulating tissue could also be present in a pre-fistula state, non-fistulating, inflammatory CD samples were profiled. In representative examples (Patient JR 32939_22, Patient BAY 26534_12), multiple fissuring ulcers were observed within the mucosa. At subcellular resolution, two of these ulcers displayed the same transcriptomic signature of FAS fibroblasts. These fibroblasts showed discrete expression of *IL11*, co-localised with *MMP3* and *WNT5A*, specifically arising from the FAS MMP1/3$^+$ subcluster (**Fig. 5.18**). This characteristic was also observed in other inflammatory samples (**Fig. 5.18**).

Although further replication and validation are necessary, this finding might offer an early, predictive signature for fistula formation in CD. Tracing the journey of this population from initial insult to

fistula formation will provide valuable insights. Understanding how these cells become primed during chronic inflammation to assume pathological states is of particular interest.

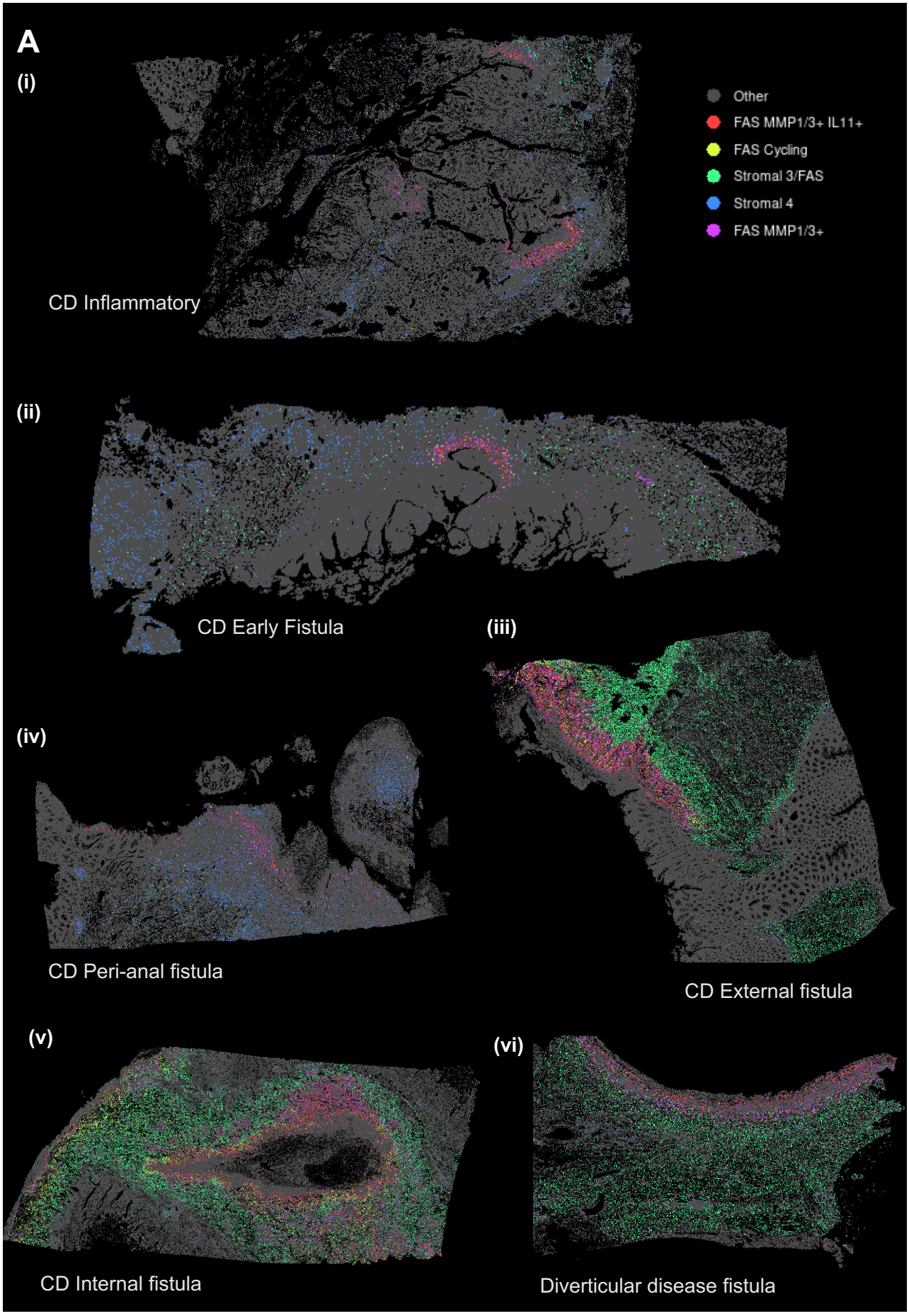

A
(i)
Other
FAS MMP1/3+ IL11+
FAS Cycling
Stromal 3/FAS
Stromal 4
FAS MMP1/3+
CD Inflammatory
(ii)
CD Early Fistula
(iii)
CD External fistula
(iv)
CD Peri-anal fistula
(v)
CD Internal fistula
(vi)
Diverticular disease fistula

Figure 5.18: Subcellular Spatial Characterisation of Fistula Associated Stroma Subclusters *(previous page)*.
(A) Subcellular spatial transcriptomic distribution of fistula-associated stromal (FAS) subclusters across sections of fistulating Crohn's disease (CD), inflammatory CD and diverticular fistulae ((i) Patient JR 32939_22 (ii) Patient BAY 26534_12 (iii) Patient BAY 113385_21 (iv) Patient JR 7909_23 (v) Patient BAY 94350_21 (vi) Patient JR 18170_21). For reference to the denotation of the fistula tract and fissuring ulcer, please see **Figure 5.15**.

5.12 Interim discussion

Crohn's fistulae pose a significant clinical challenge due to their diverse and poorly classified nature. Historically, understanding their molecular origins has been difficult. This concluding chapter reports the first unbiased spatial exploration of adult intestinal fistulating CD pathology across 34 samples. Additionally, it represents the first subcellular spatial transcriptomic exploration of 44 samples and more than six million associated single cells.

A key advantage of this approach was the ability to employ full-thickness intestinal tissue samples, unlike former studies that utilised only intestinal biopsies and samples from the fistula tracts with limited depth. For the first time, this study charts how the fistula penetrates across all layers of the intestinal compartments and delineates the cellular dynamics within, identifying distinct cellular spatial microdomains which hallmark fistula tracts as opposed to descriptive histopathological features.

Common pathological features among diverse fistulae phenotypes were reported as cellular microdomains, including epithelial crypt-villus axis failure, loss of Lgr5+ stem cells, epithelial regeneration, and fistula-associated stromal remodelling. The study identified dysregulated Wnt pathways that hallmark these conserved fistula-associated microdomain pathology.

Subcellular high-resolution ST provided validation of Visium findings and single-cell RNA-sequencing, offering fine-grained analysis of fistula-associated clusters that was not possible at the single-cell level alone. An additional advantage of subcellular spatial characterisation was the ability

to capture cell types not well characterised by single-cell approaches, such as neutrophils. Discrete layering of neutrophils was significantly enriched at the fistula edge, co-localising with macrophages and fistula-associated stroma. Descriptions of heterogeneous macrophage populations (e.g., *SPP1*$^{+}$, *MMP9*$^{+}$, *LYVE1*$^{+}$) enriched in deep and superficial tissue fistula stroma were also defined and further support the purported role of stromal-myeloid crosstalk in the pathogenesis of fistulating CD.

Additionally, subcellular spatial resolution enabled further refinement to our definition of FAS, first identified in our single-cell data and then validated in Visium data. Four novel subsets of FAS fibroblasts were described, including FAS Cycling, FAS MMP1/3^{+}, FAS MMP1/3^{+}IL11^{+}, and Stromal 3/FAS. These subsets preferentially localise at different parts of the fistula and appear to demonstrate phenotypic plasticity.

The emergence of these fibroblasts, incapable of sustaining epithelial stem cell activity, crypt regeneration, and producing excessive ECM, was captured in disease states that precede fistula formation, such as inflammatory CD and early fistula lesions. The ability to chart their emergence temporally and define their chromatin landscape will enable the identification of epigenetic states that predispose fibroblasts to adopt pathological phenotypes, presenting a future direction for this work.

Overall, this multimodal approach has provided an in-depth phenotypic characterisation of fistulating pathology, identifying spatial microdomains that define disease pathology and may serve as a foundation for the molecular classification of fistulae. The integration of spatial transcriptomics, extensive sample profiling, and advanced computational methods establishes this project as a valuable resource with the potential to significantly advance the field. However, this chapter represents an initial exploration of the spatial data - focused on generating, curating, and qualitatively analysing this complex dataset. Future iterations will refine and quantitatively define spatial microdomains, enabling a more robust characterisation of key molecular pathways. A deeper,

quantitative analysis will facilitate the identification of therapeutic targets for functional validation and guide the development of aligned tissue and *in vivo* model systems.

Chapter 6: Concluding Discussion

6.1 Introduction

Fistulating CD is a complex clinical entity that affects 40% of patients with CD and is associated with significant morbidity. Monoclonal antibodies targeting TNF remain the mainstay of treatment; however, up to a third of patients experience fistula recurrence despite treatment and up to 40% fail to respond or lose response to biological therapies. In severe or refractory cases, surgery is necessitated. The mechanisms behind Crohn's associated fistula pathogenesis however are poorly understood.

Historically, the pathogenesis of Crohn's associated fistulae has been described using histopathological observations, immunohistochemical data, and *in vitro* experiments with intestinal cell lines. Existing hypothetical pathways to fistula formation include a combination of EMT and overexpression of MMPs in response to an epithelial defect mediated by the upregulation of pro-inflammatory cytokines. Exploration of molecular pathways underpinning pathology has been limited by access to genetic model systems and by the inherent rarity of cells making up fistula tracts.

The utilisation of single cell technologies opens access to interrogation of the phenotypic properties of rare lesional cells in an unbiased manner. Further, the emergence of ST enables researchers to simultaneously characterise gene expression of the whole transcriptome while retaining information

about the spatial tissue context. A significant advancement during the course of this project was the ability to spatially resolve gene expression signatures at a subcellular resolution using multiplexed, customisable gene panels. This approach facilitates the interrogation of cell-to-cell communication networks and spatially restricted molecular interactions. Such methodologies have recently been employed to demonstrate aberrant stromal-myeloid crosstalk in IBD (186).

The work of my project aimed to advance our understanding of Crohn's fistulae formation through the application of scRNA-seq and ST. In this study, 33 full-thickness ileal tissue samples from 21 donors were sequenced using scRNA-seq, 38 tissue sections using unbiased Visium ST, and 55 tissue sections were imaged using subcellular resolution Xenium ST.

Following data analysis, the scRNA-seq cohort yielded 129,204 high-quality cells from immune, stromal, and epithelial isolation reactions. In the ST cohort, I sequenced 93,075 tissue-covered spots using the unbiased Visium ST platform and imaged 6,334,604 high-quality segmented single cells using subcellular resolution Xenium ST with a fully custom pan-gut 480 gene panel.

The data generated provides a detailed insight into the epithelial, immune and stromal interactions in fistulating pathology. This study offers an uncharted, unbiased and detailed characterisation of the unique stromal subtypes and their transcriptional regulation in fistula pathogenesis.

6.2 Summary of key findings and wider significance to the field

6.2.1 Novel single cell molecular atlas of fistulating CD

scRNA-seq identifies fistula tract specific fibroblast perturbations

Single-cell RNA-sequencing analysis unveiled nine distinct fibroblast subsets within the data, which broadly grouped into previously described core fibroblast subsets. These included mucosal structural cells (Stromal 1, or ADAMDEC1$^+$ fibroblasts), telocytes (Stromal 2, or F3/PDGFRA$^+$ fibroblasts),

submucosal/deep tissue fibroblasts (Stromal 3, C3$^+$ fibroblasts), and follicular reticular cells (Stromal 4, CCL19$^+$ fibroblasts). Additionally, there was an expansion of a subpopulation of pericytes in the involved CD samples, which expressed genes linked to ECM interaction and immune modulation.

Moreover, a novel cluster of fibroblasts termed Fistula-Associated Stroma (FAS) fibroblasts was identified. These fibroblasts expressed several markers associated with fibrosis and tissue remodelling, including unique MMPs (MMP1, MMP3, MMP9, MMP19), collagens (COL7A1, COL1A1), and morphogens (WNT5A, WNT2, BMP1). Notably, these cells also strongly upregulated several cytokines (IL6, IL11, TNFSF11, IL24) and chemokines (CXCL1, CXCL10, CXCL13, CXCL2, CXCL9), indicating potential crosstalk with both lymphoid (e.g., CXCL13) and myeloid (e.g., CXCL1) cells. IL24 has been previously shown to promote wound healing in skin by modulating fibroblast functions and aiding keratinocyte differentiation and re-epithelialisation processes (261). IL11, on the other hand, promotes the activation and proliferation of fibroblasts, which can lead to excessive collagen deposition and fibrosis (226).

Gene regulatory network analysis was conducted, revealing several upregulated or uniquely expressed TFs within the scRNA-seq FAS subset. *TWIST1* and *TWIST2* were notable for their roles in developmental patterning and morphology. Other key TFs included *PRRX1* and *PRRX2*, linked to limb development and craniofacial abnormalities, and *RUNX2*, a regulator of skeletal development. The analysis also suggested that the upregulation of *ELK1* and *STAT1* may contribute to abnormal WNT signalling in these cells thereby driving cell proliferation.

Collectively, these findings suggest that these cells are heavily involved in the tissue remodelling that drives fistula formation in CD. This novel stromal cluster, primarily associated with fistula samples, was absent in healthy controls. This FAS population differed from the IL11$^+$ fibroblast cell state observed in inflammatory Crohn's, termed "inflammation associated fibroblasts (IAFs)" (169, 175, 260). Unlike IAFs, the transcriptional regulatory network of FAS indicated a Stromal 3 origin, sharing only a limited set of markers, including *TWIST2* and *SNAI2* (see **Appx. Fig. A.13**).

Changes in epithelial and immune composition and gene expression in fistulating CD

Significant variations were observed in epithelial and immune cell populations between healthy and fistulating involved tissue. An expansion of undifferentiated epithelial cells in fistulating tissue reflected increased proliferation. The immune cell composition also varied, with an increase in B-cells and an activated myeloid state noted in fistulating involved tissue, suggesting alterations in immune responses.

T-cell subset analysis revealed higher abundances of effector, Th17 and regulatory T-cells in diseased tissues compared to healthy controls, with specific increases in T-follicular and GZMK^{+} effector T-cells in fistulating involved tissue.

6.2.2 High resolution spatially resolved cellular anatomy of diverse fistula microdomains

To understand the precise cellular and molecular determinants driving pathology in CD fistulae, high-resolution spatial analysis of fistula tract areas from CD patients and controls was undertaken. Detailed profiling was conducted across a range of highly heterogeneous disease presentations, ensuring adequate sampling from internal and external fistulae (including enterocutaneous, colocutaneous, and peri-anal types) and capturing both epithelialised and non-epithelialised fistula tract presentations.

Despite the considerable heterogeneity observed, transcriptional commonalities could be appreciated across locations (e.g., perianal *vs.* internal) and fistula types (e.g., internal CD fistulae *vs.* diverticular fistulae). Furthermore, shared cellular microdomains were identified across different types of fistulae, providing a more detailed analysis of the microenvironment within fistulating tissue.

Spatial analysis of the Visium cohort revealed 14 distinct regions, while the high-resolution Xenium cohort identified 17 distinct tissue regions. Differential abundance analyses consistently identified multiple fibroblast sub-populations enriched across all fistula types, although type-specific variations were noted; for example, squamous epithelium was exclusively detected in perianal fistulae. Additionally, there was a strong enrichment of neutrophils and multiple fistula-specific macrophage clusters.

In both the Visium and Xenium cohorts, transcriptionally distinct microdomains of the fistula tract were identified. These microdomains manifest in terms of intestinal epithelial stem cell collapse, regenerative epithelia, Fistula-Associated Stroma both deep within the tissue and at the fistula edge.

High resolution spatially resolved cellular anatomy of fistula microdomains

A gradual transition from intact crypt-villus architecture to disordered crypt architecture was observed in the mucosa toward the fistula edge, accompanied by the loss of Lgr5+ stem cells. This alteration was linked to the downregulation of epithelial TFs (*ATOH1*, *CDX1*, *CDX2*) and markers of mature epithelial cells (FABP1, CA4). This transition was facilitated by the collapse of the supporting stromal niche, indicated by reduced expression of stromal markers (ADAMDEC1, POSTN, F3, BMP5). FAS fibroblasts enriched in these regions expressed high levels of *WNT5A*, *PDGFRA* and *F3*, which are typically associated with telocytes - fibroblasts required for the maintenance of the epithelial crypt niche. This suggested an additional role for these cells in supporting epithelial regeneration, resulting in abnormal epithelialisation of the fistula tract.

This fistula-edge domain was not present in areas deeper into fully-epithelialised fistula tracts, suggesting that once re-epithelialisation is established, these domains break down, establishing a more quiescent "new homeostasis" within the fistula. Further analysis also identified regenerative efforts at the site of epithelial loss in the pre-fistula region in the absence of Lgr5+ stem cells, characterised by upregulated regenerative proteins (*REG1A*, *REG1B*, *REG4*) and pro-inflammatory

signals (DUOX2, DMBT1). These changes suggest a dynamic response to tissue damage, aiming to promote epithelial repair and innate defence mechanisms.

The stromal remodelling in fistulating CD tissue was driven by FAS fibroblasts, which exhibited increased proliferation and aberrant WNT signalling. This morphogenic signalling, particularly involving WNT5A, WNT4 and WNT2, promoted fibroblast migration and tissue remodelling via canonical and non-canonical pathways.

Neutrophil and macrophage (SPP1^{+}, MMP9^{+}) rich regions made up the outermost layer of the tracts in non-epithelialised fistulae and were adjacent to a fistula-edge specific stromal domain. More distally to the fistula edge, specific immune-rich stromal regions distinct from the fistula-edge domain were observed.

High resolution spatial mapping of Fistula Associated Stromal subsets

In situ data identified four fibroblast sub-populations that were consistently over-represented across all broad fistulae subtypes. These sub-populations could be distinguished by the expression of key TFs, cytokines, chemokines, morphogens, and matrix-associated factors. Mapping these to our scRNA-Seq data, these subsets included FAS Cycling, FAS MMP1/3^{+}, FAS MMP1/3^{+}IL11^{+}, and Stromal 3/FAS, each with distinct spatial distributions.

FAS MMP1/3^{+} cells, producing MMPs, clustered around the fistula edges. Stromal 3/FAS fibroblasts were primarily found in deeper stromal layers. FAS MMP1/3^{+}IL11^{+} cells lined the outermost lumen-adjacent zone, closely interspersed with actively proliferating FAS (FAS Cycling) cells. Immune and morphogen signalling and matrix-remodelling pathways were dampened with increasing distance from the fistula edge.

6.2.3 Wider significance in the field of IBD

The findings presented here support the view that there is considerable diversity and functional heterogeneity of intestinal fibroblasts in disease, highlighting the complex interplay between fibroblasts and immune as well as non-immune cells in the intestine in both health and disease. This builds on previous research demonstrating the critical role of these interactions in inflammatory bowel diseases, as noted in key studies by Xavier and colleagues (175) (169), Mao (262) and Kinchen (168), as well as the comprehensive work of Rieder and colleagues (263, 264). Uniquely, the work of this project can now comment on the exact spatial location of these different fibroblast populations.

Recent work from Rieder's group has generated the first full-thickness scRNA-seq atlas for stricturing CD, identifying unique fibroblast populations that are upregulated or transcriptionally active in CD strictures and as the major signal-sending cell type within these strictures. In their study, selective blockade of the shared cell surface receptor, Cadherin 11, demonstrated the prevention of intestinal fibrosis in animal models (265). Their approach, which utilised unbiased analysis of single-cell transcriptional activity in fibroblast subsets to explore therapeutic targets, parallels the identification of pathogenic fibroblast populations in fistulae within my work. This identification presents a potential basis for the development of further fibroblast-selective therapies, which are currently the focus of clinical trials in the field (NCT05843578).

This project has also identified an increase in several populations of myeloid cells that tightly co-localise to the fistula edge. This builds on previous work that has implicated the crosstalk between activated myeloid cells and stromal cells as integral to CD pathogenesis (186, 210) and most recently in perianal fistulae through ligand-receptor analysis (260). Moreover, the significant upregulation of immune pathways at both the fistula edge and deep fistula-associated stroma, including heightened NF-κB signalling, reinforces the mechanistic basis for established and efficacious therapeutic strategies such as anti-TNF therapies in the treatment of fistulating CD.

Additionally, this project has identified changes in epithelial composition along the fistula edge, providing insight into the process of re-epithelialisation in the context of disrupted crosstalk with the supporting stromal niche. This contributes to the field, as there is limited detailed understanding of the regenerative capacity of epithelial cells and the mechanisms of wound healing in patients with CD. Importantly, within the cohort, there was no robust transcriptional evidence for any intermediate epithelial-mesenchymal cell states, which have been previously described in histopathological analyses and hypothesised to drive fistula formation.

Crohn's fistulae are heterogeneous and largely unclassified, making it a top ten research priority in IBD in recent times (266). While clinically based classifications have recently been defined for perianal fistulising CD by consensus (267), the functional and cellular microdomains described in this work provide insights that may contribute to future classification and stratification efforts. Additionally, this study offers a valuable resource for further research, supporting the development of aligned tissue and *in vivo* model systems to aid in the identification of therapeutic targets.

6.3 Limitations of work

There are some limitations to the presented work. Firstly, it was not possible to curate a perfectly controlled cohort. Due to the advanced nature of fistulating pathology, some patients had received multiple biological therapies by the time they came to surgery. As such, it was not possible to power the study enough to control for biological therapies. The biological status of patients might indeed affect the molecular signatures described. To address this in future studies, it would be beneficial to cohort patients based on the presence or absence of biologics and to consider the type of biologic therapy within the cohort.

A significant challenge encountered in scRNA-seq, as previously discussed, was the lack of precise capture of the fistula tract for sampling purposes. To circumvent this, I employed ST, which enabled a precise study of fistula tract anatomy. However, it was not possible to validate some of the cellular

interactions identified by Xenium with the scRNA-seq data due to insufficient isolation of fistula tract cells in the single-cell approach. In the future, an optimisation to the collection process would be to micro dissect out the fistula tract from the tissue before processing for cryopreservation, ensuring a more targeted and accurate capture of the relevant cells.

Although the Xenium gene panel was curated to capture signatures of cell types present in health based on scRNA-seq data. It would be useful to iterate on its design to investigate hypotheses thrown up by unbiased scRNA-seq data, for example in capture of specific regenerative pathways likely to be dysregulated in the tracts. For instance, while WNT4 and the core components of the PCP pathway were well characterised in the Visium data, the Xenium panel included only WNT4 from that pathway. This limitation hindered our ability to validate some key findings across paired Visium-Xenium samples.

6.4 Future directions

In this study, spatial and scRNA-seq were brought together to phenotype fistulating CD. Extensive optimisation experiments were conducted to facilitate the study of FT, intestinal tissue for the first time in this way. While the subcellular ST dataset is still undergoing analysis and curation, the characterisation achieved thus far, despite some limitations, offers an unprecedented level of detail on fistulating Crohn's pathology. In the future, I envisage that spatial platforms may advance to the extent where subcellular resolution spatial profiling can be performed in an unbiased manner across the entire human genome, enabling fine-grained *de novo* discovery of fistulating CD lesions.

With the advent of multiomic approaches, there is a potential to further this work by studying the epigenetic landscape of fistulating CD. This approach could provide deeper insights into the phenotype of fistula tracts, especially given the discovery of key transcriptional regulators associated with patterning and morphogenesis in fistula formation. Within our lab, we have optimised a sc-multiomic technique that simultaneously employs scRNA-seq, single-cell ATAC-sequencing

(scATAC-seq), and CITE-sequencing on FT intestinal tissue. Using this method, I aim to define the transcriptional regulatory networks and open chromatin regions that hallmark the emergence of pathological fibroblast states in contrast to healthy tissue regeneration in response to injury. This in turn will enable the identification of the TFs and signalling pathways that drive the emergence of FAS fibroblasts, which are incapable of sustaining epithelial stem cell activity, crypt regeneration and produce excessive ECM.

The primary goal of this thesis has been to phenotype fistulating CD. Moving forward, my research will build on these findings to better understand the functional biology underlying fistula formation. One hypothesis from the spatial analysis suggests that fistulae develop from fissuring ulcers. In the first instance, this can be investigated further by performing immunohistochemistry on pre-fistula tissue, such as inflammatory Crohn's controls with fissuring ulcers, using markers identified from the FAS population, including key orchestrating TFs like *TWIST1*, *TWIST2*, *PRRX1*, *PRRX2*, and *RUNX2*. Additional validation of these markers can be carried out in FFPE intestinal tissue derived from the SAMP/YitC mouse model with the fistulating phenotype, to determine if this mouse model accurately represents fistulating Crohn's.

It will also be valuable to sample lesional pathology from Crohn's fistulae at various stages (pre-fistula, early fistula and established fistula) to classify and predict the development and chronicity of fistulae at the molecular level.

Having defined the aforementioned TFs as key regulators of FAS fibroblasts, it may be possible to examine their functional contributions. To assess their function, I plan to overexpress these TFs in fibroblasts isolated from healthy ileal tissue. The gene expression and chromatin landscapes of these transfected fibroblasts could be analysed through chromatin immunoprecipitation sequencing (ChIP-Seq) to identify DNA binding sites of these TFs and understand the genes they regulate. This will help to elucidate the downstream effects on gene expression and cellular pathways, providing insights into the overall impact of TF overexpression on fibroblast function and phenotype. In

addition to gene expression analysis, fibroblast motility, migration and invasion will be studied using assays such as scratch assays, comparing manipulated and unmanipulated fibroblasts.

Some fistula tracts are observed to be non-epithelialised. The data suggests that the loss of frizzled receptor 5 disrupts signalling to IESCs, impairing the maintenance and renewal of the epithelial barrier. To further investigate this hypothesis, I plan to co-culture patient-derived fibroblasts from fistulating intestinal tissue with epithelial organoids from healthy intestinal tissue to assess the impact of diseased fibroblasts on epithelial regeneration. Using multiomic analysis and by examining key signalling pathways, I will measure the effects on both the epithelial and stromal compartments. This approach will provide comprehensive insights into how stromal-epithelial interactions contribute to fistula formation and maintenance.

In the latter stages of my project, I initiated the culture of patient-derived intestinal epithelial organoids from fistulating CD tissue, measuring their differentiation efficacy through qPCR of key markers. While further optimisation is needed to reliably culture these organoids from frozen tissue, establishing robust human tissue models that accurately recapitulate fistulating pathology is critical. These models will verify and expand our current understanding while allowing us to explore novel causal pathways and potential therapeutic targets.

In summary, the work presented in this thesis was designed to enhance our understanding of the lesional pathology associated with Crohn's fistulae, rather than to establish causality. By providing a detailed phenotypic characterisation of fistulating CD, this research lays the groundwork for future studies aimed at identifying targets for therapeutic intervention. My hope is that these efforts will pave the way for novel treatments to address this debilitating complication of CD, ultimately improving patient outcomes and quality of life.

Appendix

Contents:

Table A.1: Single cell RNA-sequencing patient metadata

Study Code	Date of Surgery	Run	Experiment	Chemistry	CITE-Seq	Hashing Code	Disease	Phenotype (Montreal)	Location	Involvement Status
TIP 522	30.06.2021	SC RUN 1	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_5	Healthy	NA	Ileum	NA
TIP 473	10.04.2021	SC RUN 1	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_1	CD	Fistulating (B3)	Ileum	Involved
TIP 473	10.04.2021	SC RUN 1	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_2	CD	Fistulating (B3)	Ileum	Uninvolved
GI 6788	25.03.2021	SC RUN 1	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_3	CD	Inflammatory (B1)	Ileum	Involved
GI 6788	25.03.2021	SC RUN 1	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_4	CD	Inflammatory (B1)	Ileum	Uninvolved
GI 6968	07.12.2021	SC RUN 2	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_7	Healthy	NA	Ileum	NA
TIP 535 ECF	15.07.2021	SC RUN 2 (Epi/CD45/Str) + SC RUN 9 (Str)	Epithelial Enrichment + CD45/Stromal Enrichment, Stromal Enrichment	5' 10X scRNA	Y	HTO_1 (R2) , HTO_1(R9)	CD	Fistulating (B3)	Ileum	Involved
TIP 535 ECF	15.07.2021	SC RUN 2	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_2	CD	Fistulating (B3)	Ileum	Involved
TIP 535 FIST TRACT	15.07.2021	SC RUN 9	Stromal Enrichment	5' 10X scRNA	Y	HTO_2	CD	Fistulating (B3)	Ileum	Involved
TIP 535 MUC FST	15.07.2021	SC RUN 4 (Epi/CD45/Str) + SC RUN 7 (Str)	Epithelial Enrichment + CD45/Stromal Enrichment, Stromal Enrichment	5' 10X scRNA	Y	HTO_2 (R4), HTO_7 (R7)	CD	Fistulating (B3)	Ileum	Involved
GI 7051	27.05.2021	SC RUN 2	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_3	CD	Inflammatory (B1)	Ileum	Involved
GI 7051	27.05.2021	SC RUN 2	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_4	CD	Inflammatory (B1)	Ileum	Uninvolved
GI 7595	26.08.2021	SC RUN 2	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_5	CD	Stricturing (B2)	Ileum	Involved
GI 7595	26.08.2021	SC RUN 2	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_6	CD	Stricturing (B2)	Ileum	Uninvolved
TIP 633	22.12.2021	SC RUN 3	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_7	Healthy	NA	Ileum	NA
TIP 559	13.09.2021	SC RUN 3 (Epi/CD45/Str) + SC RUN 9 (Str)	Epithelial Enrichment + CD45/Stromal Enrichment, Stromal Enrichment	5' 10X scRNA	Y	HTO_1 (R3), HTO_3(R9)	CD	Fistulating (B3)	Ileum	Involved

TIP 559	13.09.2021	SC RUN 3	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_2	CD	Fistulating (B3)	Ileum	Uninvolved
TIP 651	13.01.2022	SC RUN 3 (Epi/CD45/Str) + SC RUN 8 (Str)	Epithelial Enrichment + CD45/Stromal Enrichment, Stromal Enrichment	5' 10X scRNA	Y	HTO_3(R3), HTO_6(R8)	CD	Inflammatory (B1)	Ileum	Involved
TIP 651	13.01.2022	SC RUN 3	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_4	CD	Inflammatory (B1)	Ileum	Uninvolved
IBD 38	20.07.2021	SC RUN 3 (Epi/CD45/Str) + SC RUN 8 (Str)	Epithelial Enrichment + CD45/Stromal Enrichment, Stromal Enrichment	5' 10X scRNA	Y	HTO_5 (R3), HTO_2(R8)	CD	Stricturing (B2)	Ileum	Involved
IBD 38	20.07.2021	SC RUN 3	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_6	CD	Fistulating (B3)	Ileum	Uninvolved
GI 6239	31.03.2021	SC RUN 4 (Epi/CD45/Str) + SC RUN 7 (Str)	Epithelial Enrichment + CD45/Stromal Enrichment, Stromal Enrichment	5' 10X scRNA	Y	HTO_1(R4), HTO_2(R7)	CD	Inflammatory (B1)	Ileum	Involved
GI 7738	18.11.2021	SC RUN 4	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_3	CD	Stricturing (B2)	Ileum	Involved
GI 7738	18.11.2021	SC RUN 4	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_4	CD	Fistulating (B3)	Ileum	Uninvolved
TIP 420	23.02.2021	SC RUN 5	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_1	CD	Inflammatory (B1)	Ileum	Involved
TIP 525	03.03.2022	SC RUN 5	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_2	Healthy	NA	Ileum	NA
GI 7833	09.12.2021	SC RUN 5	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_3	CD	Stricturing (B2)	Ileum	Involved
TIP 652	13.01.2022	SC RUN 5	Epithelial Enrichment + CD45/Stromal Enrichment	5' 10X scRNA	Y	HTO_4	CD	Stricturing (B2)	Ileum	Involved
TIP 815	30.06.2022	SC RUN 7	Stromal Enrichment	5' 10X scRNA	Y	HTO_3	CD	Inflammatory (B1)	Ileum	Involved
GI 7812	16.06.2022	SC RUN 7	Stromal Enrichment	5' 10X scRNA	Y	HTO_4	CD	Inflammatory (B1)	Ileum	Involved
TIP 852	26.09.2022	SC RUN 8	Stromal Enrichment	5' 10X scRNA	Y	HTO_3	Healthy	NA	Ileum	NA
GI 8366	11.10.2022	SC RUN 8	Stromal Enrichment	5' 10X scRNA	Y	HTO_4	Healthy	NA	Ileum	NA
GI 6966	04.07.2022	SC RUN 9	Stromal Enrichment	5' 10X scRNA	Y	HTO_4	CD	Stricturing (B2)	Ileum	Involved

Table A.1 Single Cell RNA-sequencing Patient Metadata. [ECF, enterocutaneous fistula; HTO, Hashtag oligo; SC, single-cell; Str, stromal]

Table A.2: Customised Gene Panels for MERFISH and Xenium. List of genes within customised 500-plex MERFISH and 480-plex Xenium panels respectively. Table available here: https://figshare.com/s/51e709939df40eb137c7

Table A.3: Overview of Patient Sample Cohort Characteristics and Clinical Metadata. Comprehensive overview of patient sample cohort characteristics and clinical metadata linked to single-cell, Visium spatial transcriptomics, Xenium ST, and qPCR experiments. Table available here: https://figshare.com/s/09a9ff5f58dd3613cfdb

Table A.4: qPCR Relative Expression Data. This table presents the processed quantitative PCR data, showing relative gene expression for each sample, along with the corresponding TaqMan primer details. Included is an analysis of selected Wnt pathway markers, adjusted for epithelial content using EpCAM expression as a surrogate measure. Comparative expression levels between Crohn's disease samples and healthy controls are provided. Full table available here: https://figshare.com/s/647da019de8595188b9e

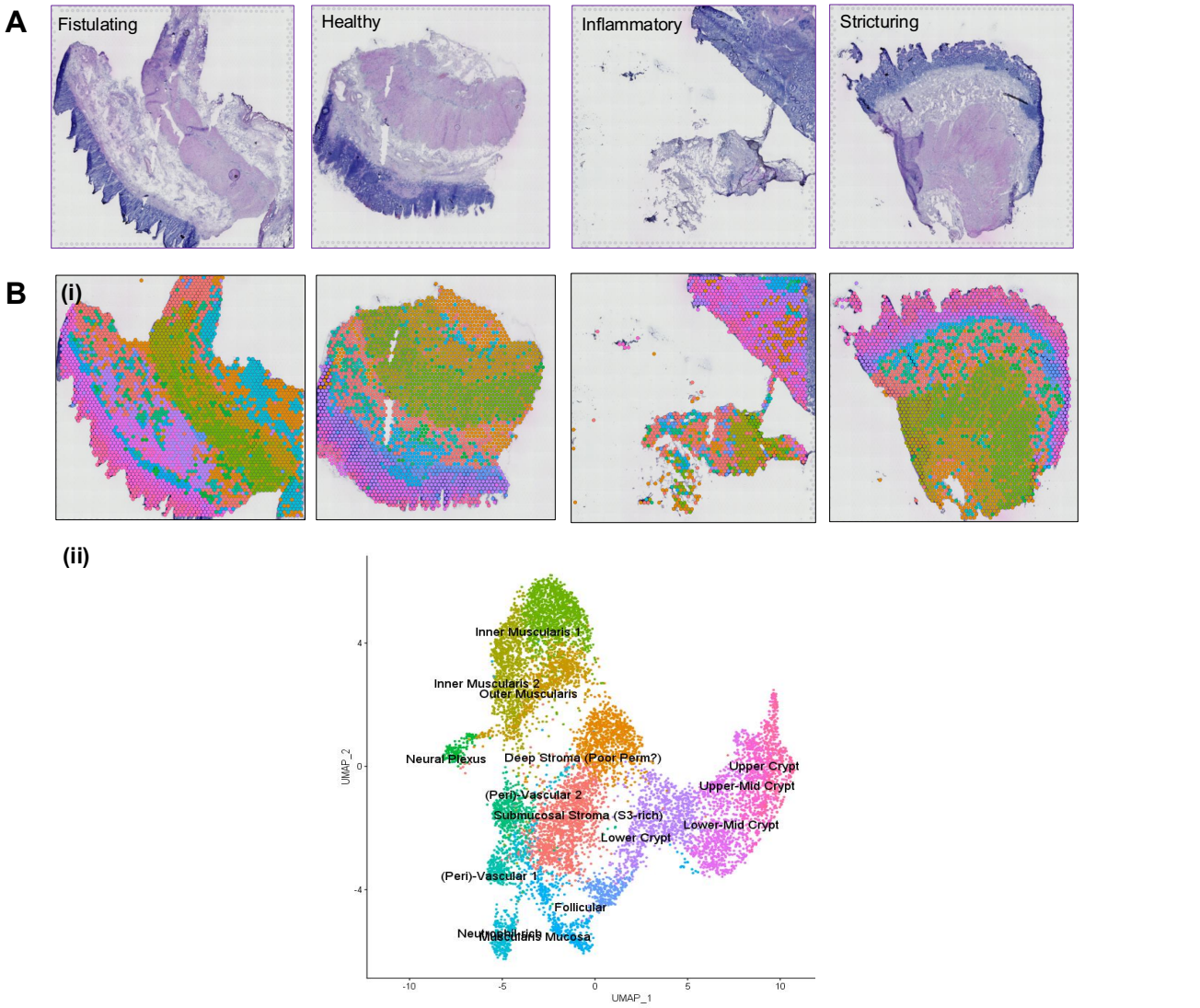
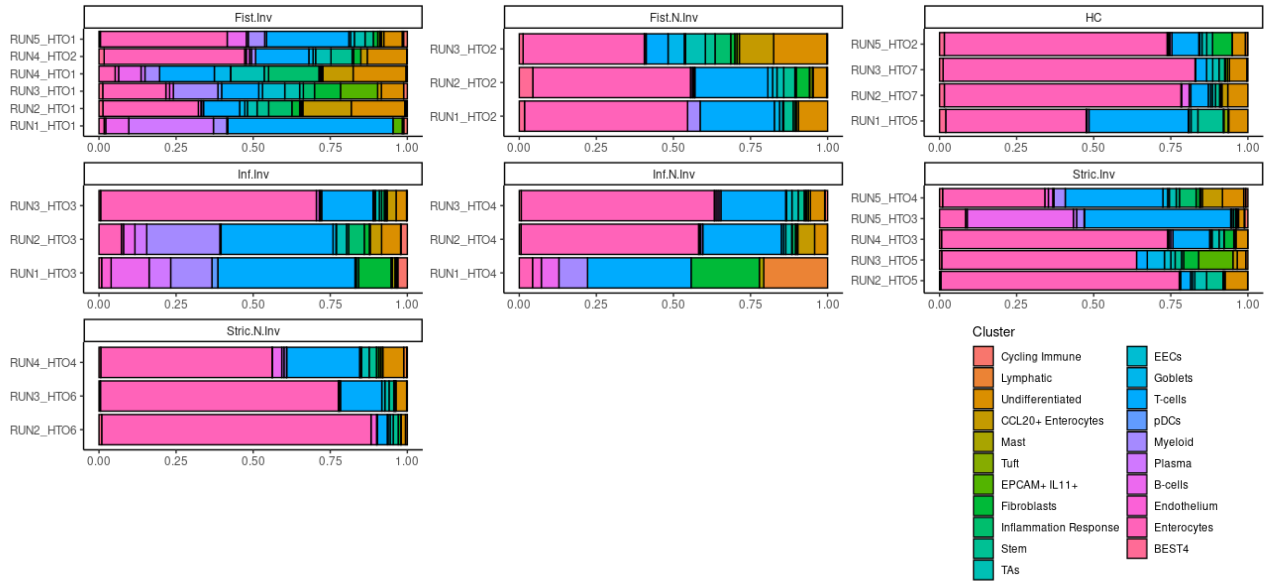
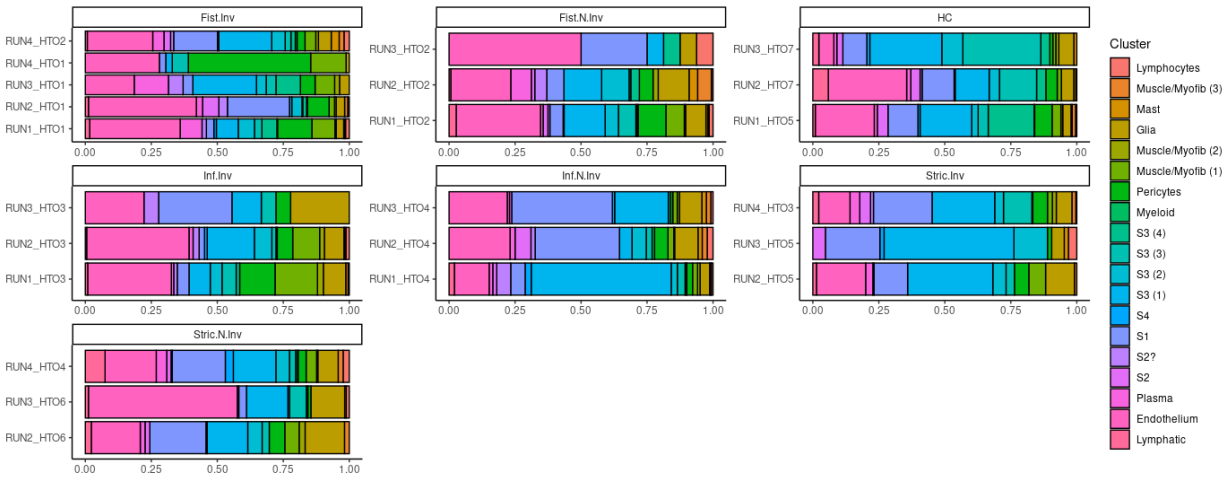

A
Fistulating
Healthy
Inflammatory
Stricturing
B
(i)
(ii)
Inner Muscularis 1
Inner Muscularis 2
Outer Muscularis
Neural Plexus
Deep Stroma (Poor Perm?)
(Peri)-Vascular 2
Submucosal Stroma (S3-rich)
Lower Crypt
Upper Crypt
Upper-Mid Crypt
Lower-Mid Crypt
(Peri)-Vascular 1
Follicular
Muscularis Mucosa
UMAP_1
UMAP_2
C
(i)
Median.Genes.per.Spot
(ii)
Reads.Mapped.to.Genome
(iii)
Fraction.Reads.in.Spots.Under.Tissue
Fistulating
Healthy
Inflammatory
Strictured
Type

Figure A.1: Overview of the Spatial Transcriptomics Fresh Frozen Pilot Run (*previous page*). (A) Experimental overview – Healthy (Patient GI 6968), Fistulating (Patient TIP 535), Stricturing (Patient IBD 38), and Inflammatory Crohn's (Patient GI 7051), as demonstrated by H&E staining. (B) Spatial distribution of transcriptome clusters in a representative ST slide from healthy, fistulating, stricturing, and inflammatory Crohn's resections (i). UMAP embedding visualising joint spatial region spot clusters from ST, annotated according to tissue location, structure, or cell type enrichment (ii). (C) Gene recovery per section on slide (i), number of reads mapping to the genome per section (ii), and fraction of reads falling in spots under tissue (iii).

A


Fist.Inv
RUN5_HTO1
RUN4_HTO2
RUN4_HTO1
RUN3_HTO1
RUN2_HTO1
RUN1_HTO1
Fist.N.Inv
RUN3_HTO2
RUN2_HTO2
RUN1_HTO2
HC
RUN5_HTO2
RUN3_HTO7
RUN2_HTO7
RUN1_HTO5
Inf.Inv
RUN3_HTO3
RUN2_HTO3
RUN1_HTO3
Inf.N.Inv
RUN3_HTO4
RUN2_HTO4
RUN1_HTO4
Stric.Inv
RUN5_HTO4
RUN5_HTO3
RUN4_HTO3
RUN3_HTO5
RUN2_HTO5
Stric.N.Inv
RUN4_HTO4
RUN3_HTO6
RUN2_HTO6
0.00
0.25
0.50
0.75
1.00
Cluster
Cycling Immune
Lymphatic
Undifferentiated
CCL20+ Enterocytes
Mast
Tuft
EPCAM+ IL11+
Fibroblasts
Inflammation Response
Stem
TAs
EECs
Goblets
T-cells
pDCs
Myeloid
Plasma
B-cells
Endothelium
Enterocytes
BEST4

B


Fist.Inv
RUN4_HTO2
RUN4_HTO1
RUN3_HTO1
RUN2_HTO1
RUN1_HTO1
Fist.N.Inv
RUN3_HTO2
RUN2_HTO2
RUN1_HTO2
HC
RUN3_HTO7
RUN2_HTO7
RUN1_HTO5
Inf.Inv
RUN3_HTO3
RUN2_HTO3
RUN1_HTO3
Inf.N.Inv
RUN3_HTO4
RUN2_HTO4
RUN1_HTO4
Stric.Inv
RUN4_HTO3
RUN3_HTO5
RUN2_HTO5
Stric.N.Inv
RUN4_HTO4
RUN3_HTO6
RUN2_HTO6
0.00
0.25
0.50
0.75
1.00
Cluster
Lymphocytes
Muscle/Myofib (3)
Mast
Glia
Muscle/Myofib (2)
Muscle/Myofib (1)
Pericytes
Myeloid
S3 (4)
S3 (3)
S3 (2)
S3 (1)
S4
S1
S2?
S2
Plasma
Endothelium
Lymphatic

C

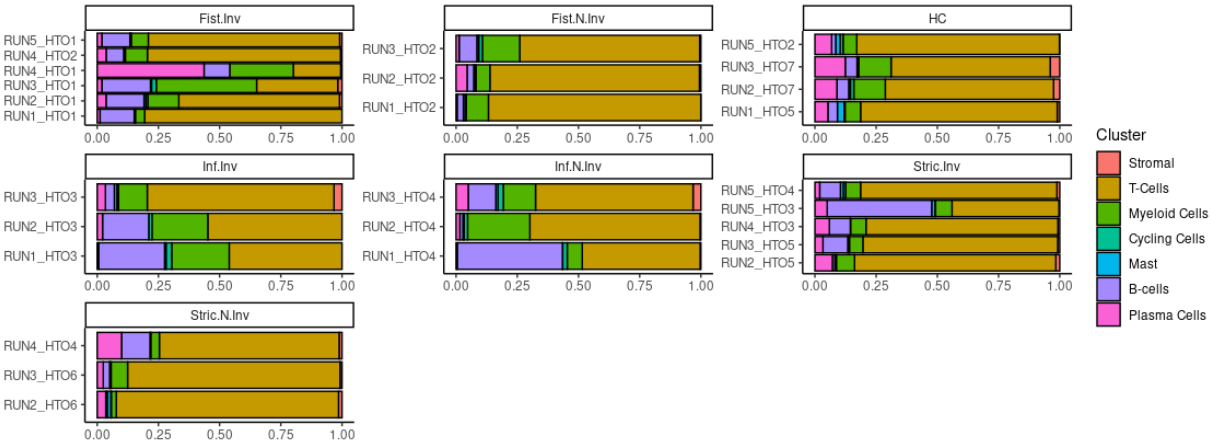

Fist.Inv
RUN5_HTO1
RUN4_HTO2
RUN4_HTO1
RUN3_HTO1
RUN2_HTO1
RUN1_HTO1
Fist.N.Inv
RUN3_HTO2
RUN2_HTO2
RUN1_HTO2
HC
RUN5_HTO2
RUN3_HTO7
RUN2_HTO7
RUN1_HTO5
Inf.Inv
RUN3_HTO3
RUN2_HTO3
RUN1_HTO3
Inf.N.Inv
RUN3_HTO4
RUN2_HTO4
RUN1_HTO4
Stric.Inv
RUN5_HTO4
RUN5_HTO3
RUN4_HTO3
RUN3_HTO5
RUN2_HTO5
Stric.N.Inv
RUN4_HTO4
RUN3_HTO6
RUN2_HTO6
0.00
0.25
0.50
0.75
1.00
Cluster
Stromal
T-Cells
Myeloid Cells
Cycling Cells
Mast
B-cells
Plasma Cells

Figure A.2: Cellular composition differences in epithelial, stromal, and immune isolation reactions (*previous page*). (A) Epithelial cell type proportions per cluster, by condition and run. (B) Stromal cell type proportions per cluster, by condition and run (runs 1-4 only). (C) Immune cell type proportions per cluster, by condition and run. [Fist, fistulating; HC, healthy control; Inf, inflammatory; Inv, involved; N.inv, non-involved; Stric, stricturing]

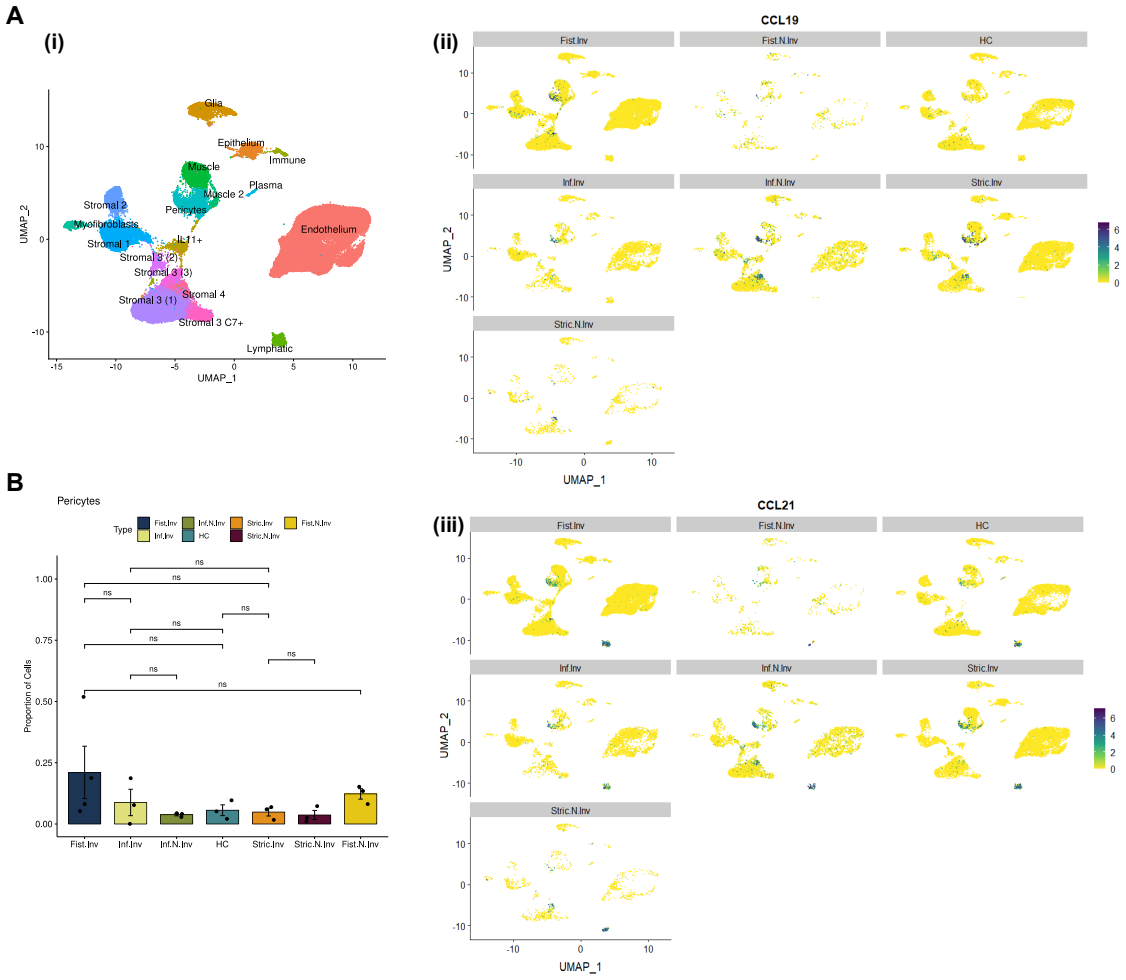


Figure A.3: Crohn's-Specific Subpopulation of Pericytes in Fistulating Disease. (A) UMAP embedding visualising scRNA-seq clusters of cell populations obtained from stromal cell isolation (i), with UMAP overlays showing the expression of *CCL19* (ii) and *CCL21* (iii) across these clusters. (B) Cellular composition differences in the stromal isolation reaction – pericytes.

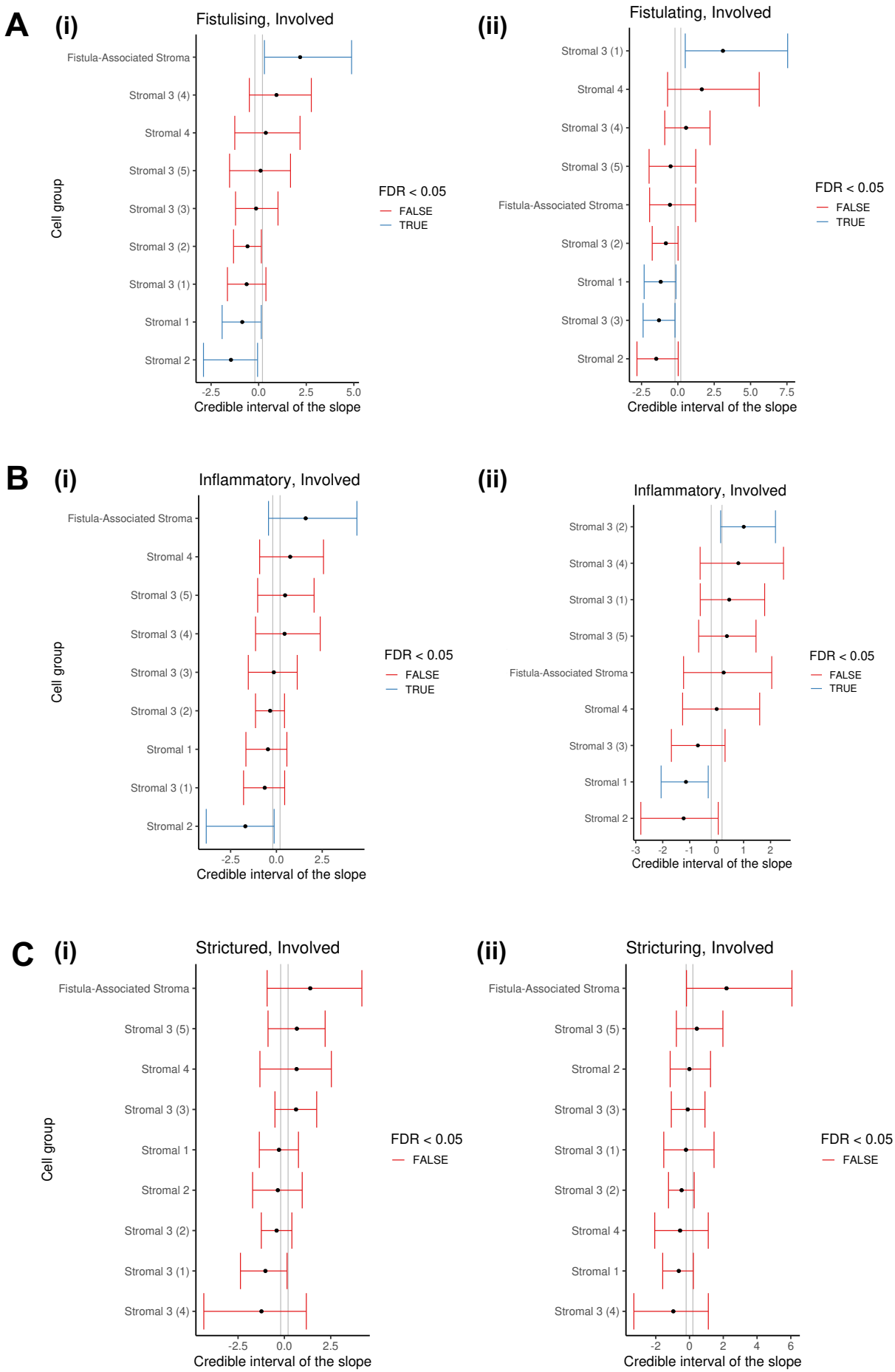

A
(i)
Fistulising, Involved
Fistula-Associated Stroma
Stromal 3 (4)
Stromal 4
Stromal 3 (5)
Stromal 3 (3)
Stromal 3 (2)
Stromal 3 (1)
Stromal 1
Stromal 2
Cell group
-2.5
0.0
2.5
5.0
Credible interval of the slope
FDR < 0.05
FALSE
TRUE
(ii)
Fistulating, Involved
Stromal 3 (1)
Stromal 4
Stromal 3 (4)
Stromal 3 (5)
Fistula-Associated Stroma
Stromal 3 (2)
Stromal 1
Stromal 3 (3)
Stromal 2
-2.5
0.0
2.5
5.0
7.5
Credible interval of the slope
FDR < 0.05
FALSE
TRUE
B
(i)
Inflammatory, Involved
Fistula-Associated Stroma
Stromal 4
Stromal 3 (5)
Stromal 3 (4)
Stromal 3 (3)
Stromal 3 (2)
Stromal 1
Stromal 3 (1)
Stromal 2
Cell group
-2.5
0.0
2.5
Credible interval of the slope
FDR < 0.05
FALSE
TRUE
(ii)
Inflammatory, Involved
Stromal 3 (2)
Stromal 3 (4)
Stromal 3 (1)
Stromal 3 (5)
Fistula-Associated Stroma
Stromal 4
Stromal 3 (3)
Stromal 1
Stromal 2
-3
-2
-1
0
1
2
Credible interval of the slope
FDR < 0.05
FALSE
TRUE
C
(i)
Strictured, Involved
Fistula-Associated Stroma
Stromal 3 (5)
Stromal 4
Stromal 3 (3)
Stromal 1
Stromal 2
Stromal 3 (2)
Stromal 3 (1)
Stromal 3 (4)
Cell group
-2.5
0.0
2.5
Credible interval of the slope
FDR < 0.05
FALSE
(ii)
Stricturing, Involved
Fistula-Associated Stroma
Stromal 3 (5)
Stromal 2
Stromal 3 (3)
Stromal 3 (1)
Stromal 3 (2)
Stromal 4
Stromal 1
Stromal 3 (4)
-2
0
2
4
6
Credible interval of the slope
FDR < 0.05
FALSE

Figure A.4: Compositional Analysis of Fibroblasts – Single-Cell RNA Data (*previous page*). Differential abundance analysis comparing fibroblasts in (A) fistulating Crohn's disease (CD) involved vs. healthy samples (i) and non-involved samples (ii), (B) inflammatory CD involved vs. healthy samples (i) and non-involved samples (ii), and (C) stricturing CD involved vs. healthy samples (i) and non-involved samples (ii), as visualised by forest plots. The false discovery rate (FDR) and significance threshold are set at $p < 0.05$.

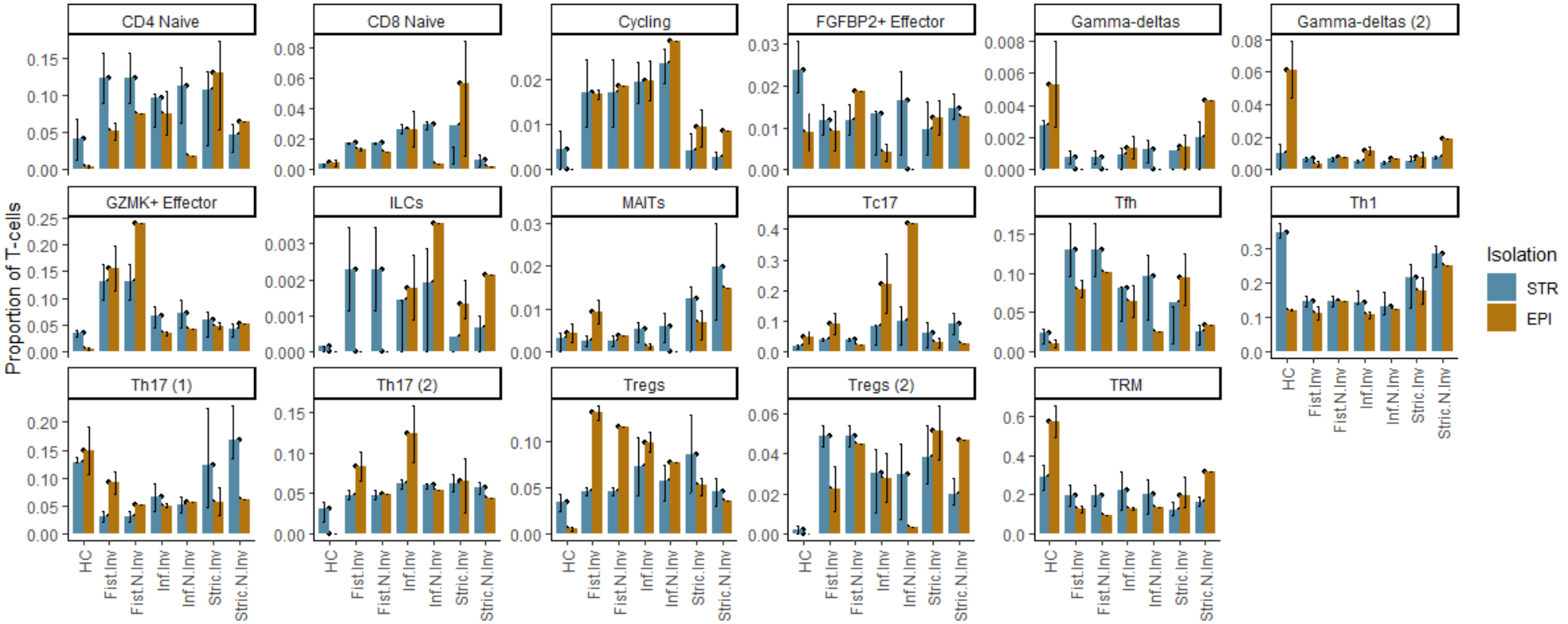


Figure A.5: T-cell abundance in epithelial and stromal compartments. Bar charts illustrating the proportion of T-cell subsets within the epithelial and stromal compartments across conditions. Error bars represent quantiles.

Figure A.6: Visium FFPE Cohort - H&E Overview. Patient sample cohort overview by H&E section (low resolution). Key: CD, Crohn's disease; DD, diverticular disease. Figures available here: https://figshare.com/s/aa4879ac817d5d66e976

Figure A.7: Overview of Visium and Xenium FFPE Cohort. Patient sample cohort overview, displaying H&E-stained sections and corresponding annotations for Visium and Xenium FFPE samples. Key: FOV, field of view; VIS, Visium; XEN, Xenium. Detailed figures are available at: https://figshare.com/s/ab93c2f917f951a6119b

Figure A.8: Visium Per Slide Analysis Outputs. Analysis for each slide, including overlays for cell type prediction signals, cluster distributions, top gene markers, and top transcription factors. Detailed figures are available at: https://figshare.com/s/66606d60547646e22f4c

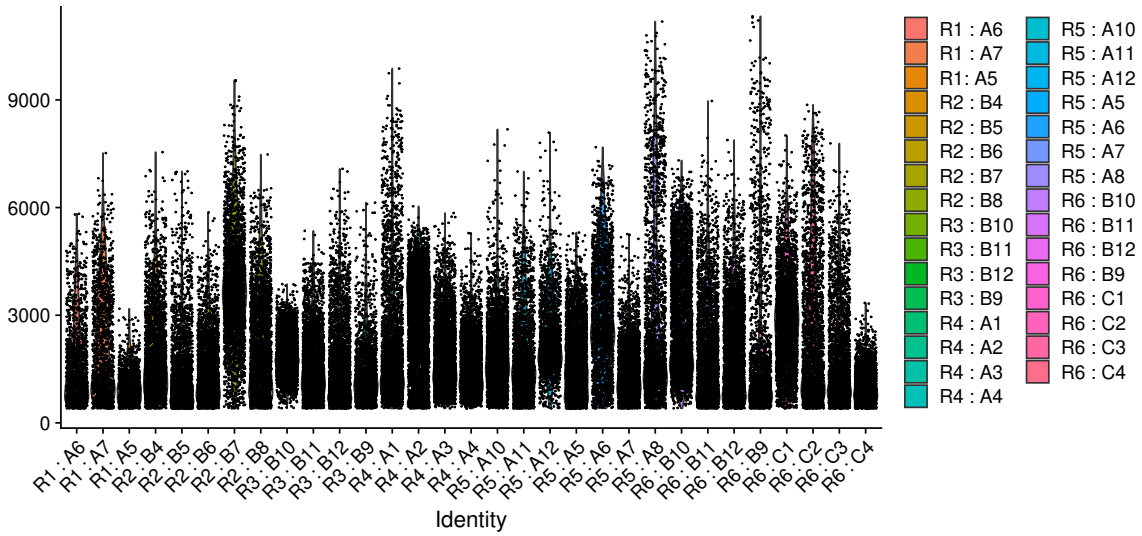
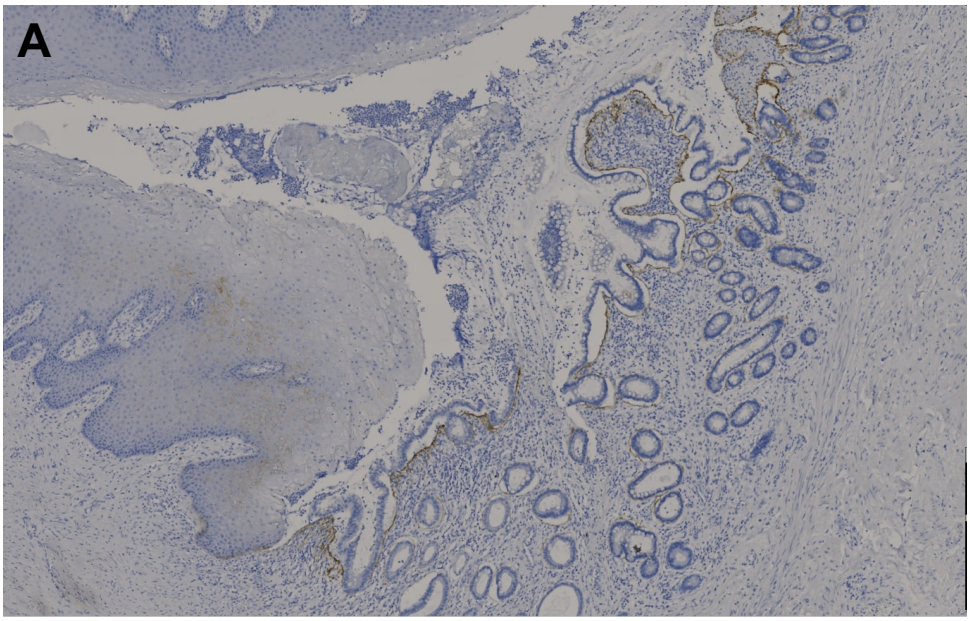
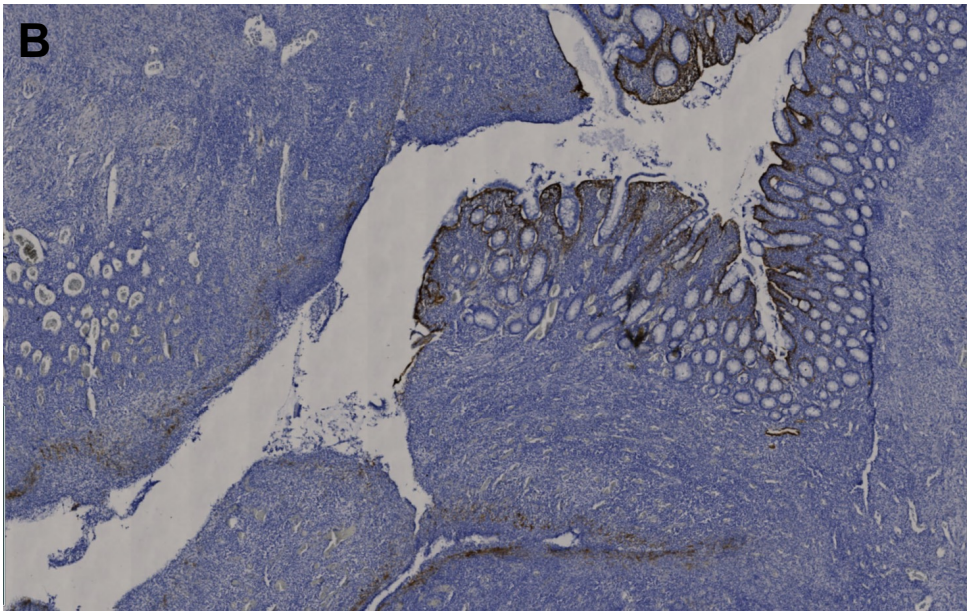


Figure A.9: Gene Recovery per Visium Slide. Violin plot illustrating the number of genes detected per spot for each Visium slide, with Visium Slide IDs displayed on the horizontal axis.

Figure A.10: Xenium Cluster Overlay Per Sample. Overlay of cluster distributions for each sample. Detailed figures are available at: https://figshare.com/s/7309d2659a56e32e3f83


A


B

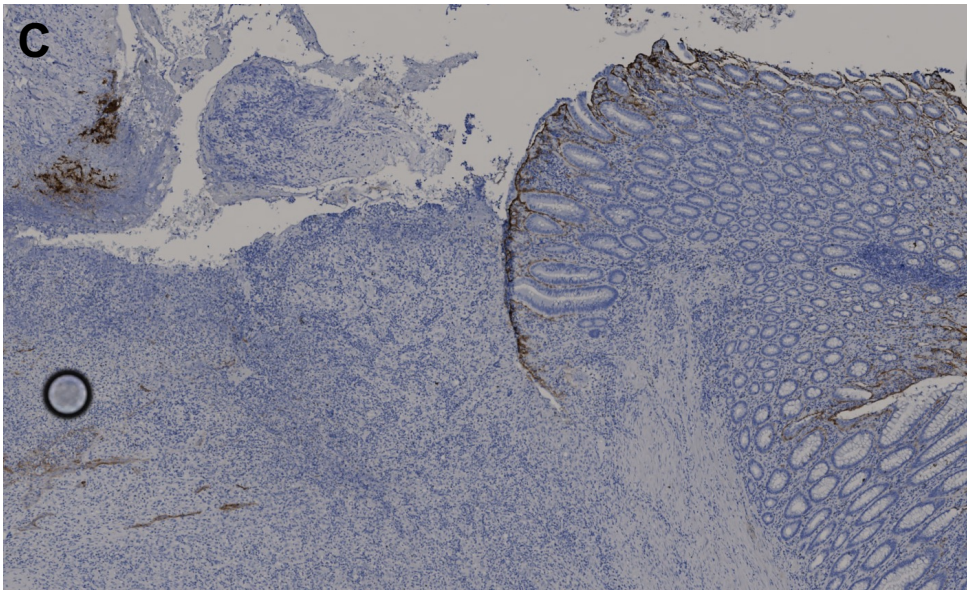

C

Figure A.11: Additional Immunohistochemistry Validation *(previous page)*. Validation by immunohistochemistry of selected sections showing loss of epithelium at the fistula tract: Patient JR 28598 (A), Patient BAY 86068 (B), and Patient BAY 113385 (C). Staining for the stromal marker F3 (pH 6, 1:250 dilution) is presented at ×1.76 original magnification.

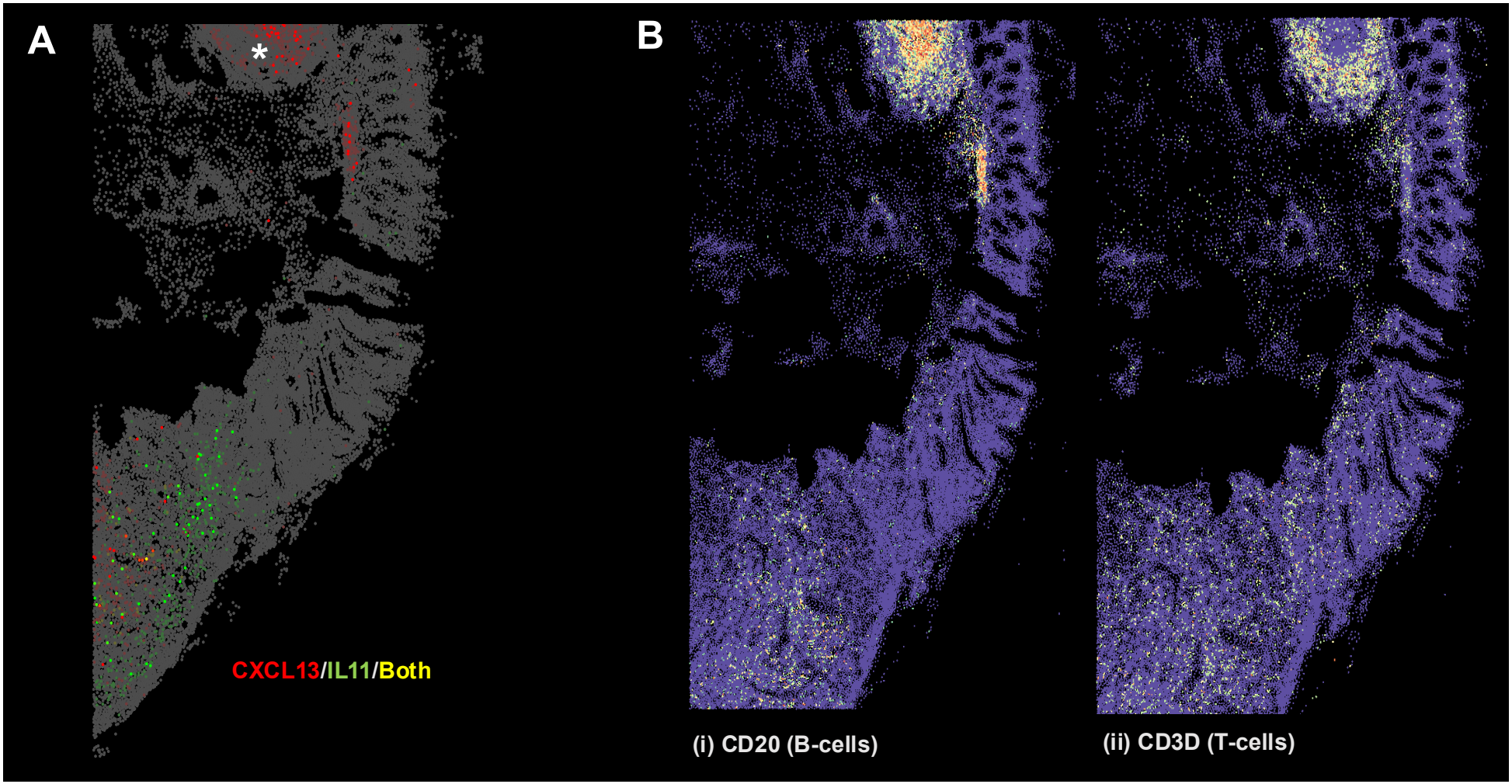


Figure A.12: Subcellular Spatial Characterisation of Immune Expression. Subcellular spatial transcriptomic distribution of *CXCL13* (red), *IL11* (green), and co-expression of both (yellow) in the section from Patient JR 47926 (A). Expression of the B-cell marker CD20 (B(i)) and the T-cell marker CD3D (B(ii)) is also shown. Asterisk denotes lymphoid follicle.

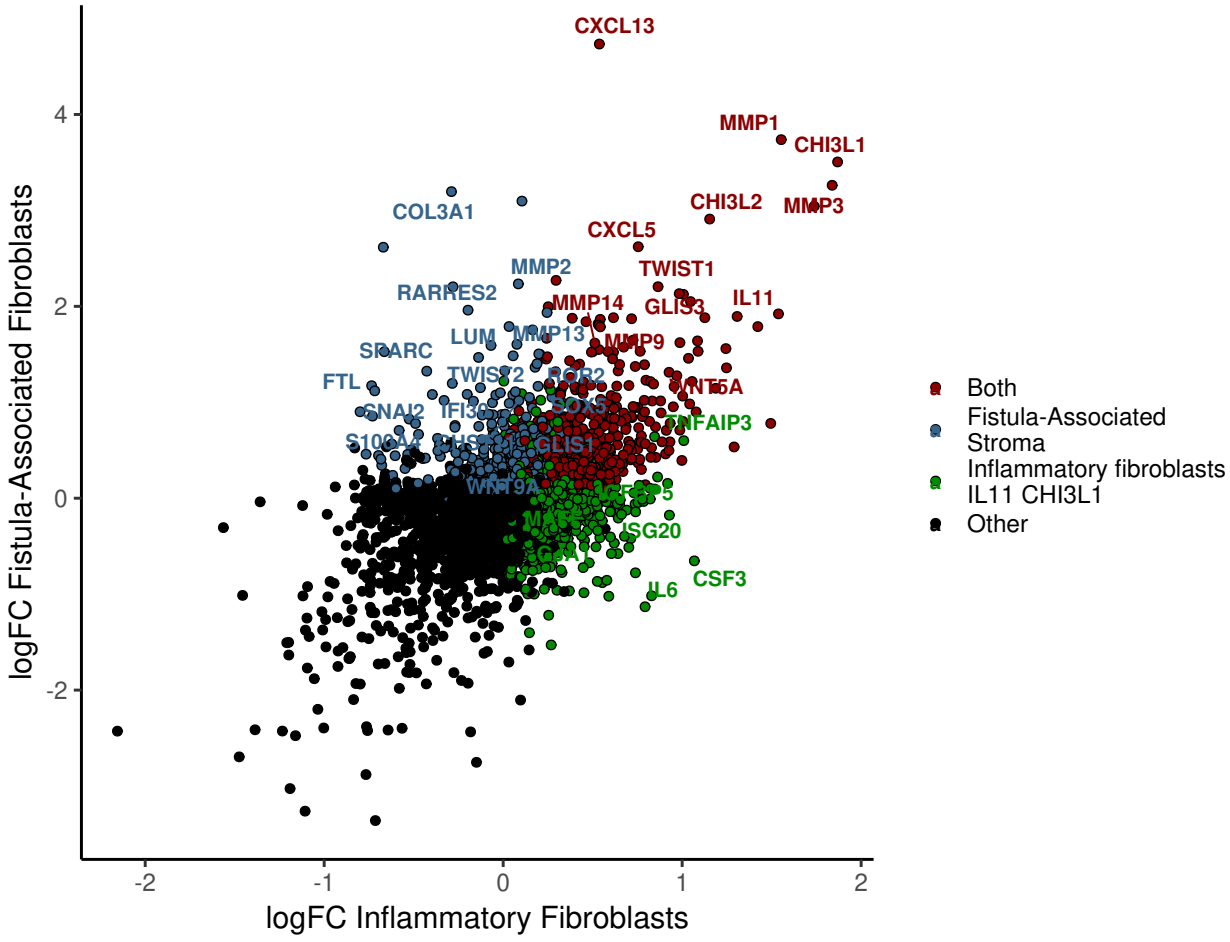


Figure A.13: Comparative Analysis of Fistula-Associated Fibroblasts and Inflammatory Fibroblasts. Correlation plot depicting the top genes expressed in fistula-associated fibroblasts and inflammatory fibroblasts IL11^{+} CHI3L1^{+} as reported in the dataset by Xavier et al (175). Genes coloured in red indicate shared expression, blue indicates genes exclusive to fistula-associated fibroblasts, and green indicates genes unique to inflammatory fibroblasts.

References

1. Jones GR, Lyons M, Plevris N, Jenkinson PW, Bisset C, Burgess C, et al. IBD prevalence in Lothian, Scotland, derived by capture-recapture methodology. Gut. 2019;68(11):1953-60.
2. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet. 2017;390(10114):2769-78.
3. Zhao J, Ng SC, Lei Y, Yi F, Li J, Yu L, et al. First prospective, population-based inflammatory bowel disease incidence study in mainland of China: the emergence of "western" disease. Inflamm Bowel Dis. 2013;19(9):1839-45.
4. Kim HJ, Hann HJ, Hong SN, Kim KH, Ahn IM, Song JY, et al. Incidence and natural course of inflammatory bowel disease in Korea, 2006-2012: a nationwide population-based study. Inflamm Bowel Dis. 2015;21(3):623-30.
5. Ng SC, Tang W, Ching JY, Wong M, Chow CM, Hui AJ, et al. Incidence and phenotype of inflammatory bowel disease based on results from the Asia-pacific Crohn's and colitis epidemiology study. Gastroenterology. 2013;145(1):158-65 e2.
6. Baumgart DC, Sandborn WJ. Crohn's disease. Lancet. 2012;380(9853):1590-605.
7. Nielsen OH, Rogler G, Hahnloser D, Thomsen OØ. Diagnosis and management of fistulizing Crohn's disease. Nature Clinical Practice Gastroenterology & Hepatology. 2009;6(2).
8. de Miguel Criado J, del Salto LG, Rivas PF, del Hoyo LF, Velasco LG, de las Vacas MI, et al. MR imaging evaluation of perianal fistulas: spectrum of imaging features. Radiographics. 2012;32(1):175-94.
9. Parks AG, Gordon PH, Hardcastle JD. A classification of fistula-in-ano. Br J Surg. 1976;63(1):1-12.
10. American Gastroenterological Association Clinical Practice C. American Gastroenterological Association medical position statement: perianal Crohn's disease. Gastroenterology. 2003;125(5):1503-7.
11. Hirten RP, Shah S, Sachar DB, Colombel JF. The Management of Intestinal Penetrating Crohn's Disease. Inflamm Bowel Dis. 2018;24(4):752-65.
12. Panés J, Rimola J. Perianal fistulizing Crohn's disease: pathogenesis, diagnosis and therapy. Nature Reviews Gastroenterology & Hepatology. 2017;14(11).
13. Schwartz DA, Loftus EV, Tremaine WJ, Panaccione R, Harmsen WS, Zinsmeister AR, et al. The natural history of fistulizing Crohn's disease in Olmsted County, Minnesota. Gastroenterology. 2002;122(4).
14. Peyrin-Biroulet L, Loftus EV, Colombel J-F, Sandborn WJ. The Natural History of Adult Crohn's Disease in Population-Based Cohorts. American Journal of Gastroenterology. 2010;105(2):289-97.
15. Solomon MJ. Fistulae and abscesses in symptomatic perianal Crohn's disease. Int J Colorectal Dis. 1996;11(5):222-6.
16. Fan Y, Zhang L, Omidakhsh N, Bohn RL, Thompson JS, Brodovicz KG, et al. Patients With Stricturing or Penetrating Crohn's Disease Phenotypes Report High Disease Burden and Treatment Needs. Inflamm Bowel Dis. 2023;29(6):914-22.
17. Chen G, Pedarla V, Null KD, Cazzetta SE, Khan QR, Schwartz DA. Health Care Costs and Resource Utilization Among Patients With Crohn's Disease With and Without Perianal Fistula. Inflamm Bowel Dis. 2022;28(6):870-7.
18. Cohen RD, Waters HC, Tang B, Rahman MI. Effects of fistula on healthcare costs and utilization for patients with Crohn's disease treated in a managed care environment. Inflamm Bowel Dis. 2008;14(12):1707-14.
19. Burisch J, Vardi H, Schwartz D, Friger M, Kiudelis G, Kupcinskas J, et al. Health-care costs of inflammatory bowel disease in a pan-European, community-based, inception

cohort during 5 years of follow-up: a population-based study. Lancet Gastroenterol Hepatol. 2020;5(5):454-64.
20. Misra R, Limdi J, Cooney R, Sakuma S, Brookes M, Fogden E, et al. Ethnic differences in inflammatory bowel disease: Results from the United Kingdom inception epidemiology study. World Journal of Gastroenterology. 2019;25(40):6145-57.
21. Kang B, Kim JE, Jung JH, Choe JY, Kim MJ, Choe YH, et al. Korean Children and Adolescents with Crohn's Disease Are More Likely to Present with Perianal Fistulizing Disease at Diagnosis Compared to Their European Counterparts. Pediatric Gastroenterology, Hepatology & Nutrition. 2020;23(1):49-.
22. Jangi S, Ruan A, Korzenik J, de Silva P. South Asian Patients With Inflammatory Bowel Disease in the United States Demonstrate More Fistulizing and Perianal Crohn Phenotype. Inflammatory Bowel Diseases. 2020;26(12).
23. Brant SR, Okou DT, Simpson CL, Cutler DJ, Haritunians T, Bradfield JP, et al. Genome-Wide Association Study Identifies African-Specific Susceptibility Loci in African Americans With Inflammatory Bowel Disease. Gastroenterology. 2017;152(1):206-17 e2.
24. Shi HY, Levy AN, Trivedi HD, Chan FKL, Ng SC, Ananthakrishnan AN. Ethnicity Influences Phenotype and Outcomes in Inflammatory Bowel Disease: A Systematic Review and Meta-analysis of Population-based Studies. Clin Gastroenterol Hepatol. 2018;16(2):190-7 e11.
25. Jackson JF, 3rd, Dhere T, Repaka A, Shaukat A, Sitaraman S. Crohn's disease in an African-American population. Am J Med Sci. 2008;336(5):389-92.
26. Alli-Akintade L, Pruthvi P, Hadi N, Sachar D. Race and fistulizing perianal Crohn's disease. J Clin Gastroenterol. 2015;49(3):e21-3.
27. Chun J, Im JP, Kim JW, Lee KL, Choi CH, Kim H, et al. Association of Perianal Fistulas with Clinical Features and Prognosis of Crohn's Disease in Korea: Results from the CONNECT Study. Gut Liver. 2018;12(5):544-54.
28. Mak WY, Mak OS, Lee CK, Tang W, Leung WK, Wong MTL, et al. Significant Medical and Surgical Morbidity in Perianal Crohn's Disease: Results from a Territory-Wide Study. J Crohns Colitis. 2018;12(12):1392-8.
29. Henckaerts L, Van Steen K, Verstreken I, Cleynen I, Franke A, Schreiber S, et al. Genetic risk profiling and prediction of disease course in Crohn's disease patients. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2009;7(9).
30. Zeitz J, Fournier N, Labenz C, Biedermann L, Frei P, Misselwitz B, et al. Risk Factors for the Development of Fistulae and Stenoses in Crohn Disease Patients in the Swiss Inflammatory Bowel Disease Cohort. Inflamm Intest Dis. 2017;1(4):172-81.
31. Moradkhani A, Beckman LJ, Tabibian JH. Health-related quality of life in inflammatory bowel disease: psychosocial, clinical, socioeconomic, and demographic predictors. J Crohns Colitis. 2013;7(6):467-73.
32. Park SH, Aniwan S, Scott Harmsen W, Tremaine WJ, Lightner AL, Faubion WA, et al. Update on the Natural Course of Fistulizing Perianal Crohn's Disease in a Population-Based Cohort. Inflamm Bowel Dis. 2019;25(6):1054-60.
33. Thia KT, Sandborn WJ, Harmsen WS, Zinsmeister AR, Loftus EV, Jr. Risk factors associated with progression to intestinal complications of Crohn's disease in a population-based cohort. Gastroenterology. 2010;139(4):1147-55.
34. Louis E, Collard A, Oger AF, Degroote E, Aboul Nasr El Yafi FA, Belaiche J. Behaviour of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease. Gut. 2001;49(6):777-82.
35. Silverberg MS, Satsangi J, Ahmad T, Arnott ID, Bernstein CN, Brant SR, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. Can J Gastroenterol. 2005;19 Suppl A:5A-36A.
36. Pariente B, Cosnes J, Danese S, Sandborn WJ, Lewin M, Fletcher JG, et al. Development of the Crohn's disease digestive damage score, the Lemann score. Inflamm Bowel Dis. 2011;17(6):1415-22.

37. Fiorino G, Morin M, Bonovas S, Bonifacio C, Spinelli A, Germain A, et al. Prevalence of Bowel Damage Assessed by Cross-Sectional Imaging in Early Crohn's Disease and its Impact on Disease Outcome. J Crohns Colitis. 2017;11(3):274-80.
38. Gottgens KW, Jeuring SF, Sturkenboom R, Romberg-Camps MJ, Oostenbrug LE, Jonkers DM, et al. Time trends in the epidemiology and outcome of perianal fistulizing Crohn's disease in a population-based cohort. Eur J Gastroenterol Hepatol. 2017;29(5):595-601.
39. Ahmad T, Armuzzi A, Bunce M, Mulcahy-Hawes K, Marshall SE, Orchard TR, et al. The molecular classification of the clinical manifestations of Crohn's disease. Gastroenterology. 2002;122(4):854-66.
40. Hellers G, Bergstrand O, Ewerth S, Holmstrom B. Occurrence and outcome after primary treatment of anal fistulae in Crohn's disease. Gut. 1980;21(6):525-7.
41. Williams DR, Coller JA, Corman ML, Nugent FW, Veidenheimer MC. Anal complications in Crohn's disease. Dis Colon Rectum. 1981;24(1):22-4.
42. Nguyen GC, Torres EA, Regueiro M, Bromfield G, Bitton A, Stempak J, et al. Inflammatory bowel disease characteristics among African Americans, Hispanics, and non-Hispanic Whites: characterization of a large North American cohort. Am J Gastroenterol. 2006;101(5):1012-23.
43. Schwartz DA, Ghazi LJ, Regueiro M, Fichera A, Zoccali M, Ong EM, et al. Guidelines for the multidisciplinary management of Crohn's perianal fistulas: summary statement. Inflamm Bowel Dis. 2015;21(4):723-30.
44. Cayci M, Bostanci EB, Turhan N, Karaman K, Dalgic T, Ozer I, et al. The analysis of clinico-pathologic characteristics in patients who underwent surgery due to stricturing and non-perineal fistulizing forms of Crohn's disease: a retrospective cohort study. Int J Surg. 2015;15:49-54.
45. Annese V, Beaugerie L, Egan L, Biancone L, Bolling C, Brandts C, et al. European Evidence-based Consensus: Inflammatory Bowel Disease and Malignancies. J Crohns Colitis. 2015;9(11):945-65.
46. Beaugerie L, Itzkowitz SH. Cancers Complicating Inflammatory Bowel Disease. N Engl J Med. 2015;373(2):195.
47. Canavan C, Abrams KR, Mayberry J. Meta-analysis: colorectal and small bowel cancer risk in patients with Crohn's disease. Aliment Pharmacol Ther. 2006;23(8):1097-104.
48. Lutgens MW, Vleggaar FP, Schipper ME, Stokkers PC, van der Woude CJ, Hommes DW, et al. High frequency of early colorectal cancer in inflammatory bowel disease. Gut. 2008;57(9):1246-51.
49. Adegbola SO, Dibley L, Sahnan K, Wade T, Verjee A, Sawyer R, et al. Burden of disease and adaptation to life in patients with Crohn's perianal fistula: a qualitative exploration. Health Qual Life Outcomes. 2020;18(1):370.
50. Mahadev S, Young JM, Selby W, Solomon MJ. Quality of life in perianal Crohn's disease: what do patients consider important? Dis Colon Rectum. 2011;54(5):579-85.
51. Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. Lancet. 2017;389(10080):1741-55.
52. Allocca M, Fiorino G, Danese S. Cross-sectional imaging modalities in Crohn's disease. Dig Dis. 2013;31(2):199-201.
53. Zallot C, Peyrin-Biroulet L. Deep remission in inflammatory bowel disease: looking beyond symptoms. Curr Gastroenterol Rep. 2013;15(3):315.
54. Noor NM, Lee JC, Bond S, Dowling F, Brezina B, Patel KV, et al. A biomarker-stratified comparison of top-down versus accelerated step-up treatment strategies for patients with newly diagnosed Crohn's disease (PROFILE): a multicentre, open-label randomised controlled trial. Lancet Gastroenterol Hepatol. 2024;9(5):415-27.
55. Sands BE, Anderson FH, Bernstein CN, Chey WY, Feagan BG, Fedorak RN, et al. Infliximab maintenance therapy for fistulizing Crohn's disease. N Engl J Med. 2004;350(9):876-85.

56. Sands BE, Blank MA, Patel K, van Deventer SJ, Study AI. Long-term treatment of rectovaginal fistulas in Crohn's disease: response to infliximab in the ACCENT II Study. Clin Gastroenterol Hepatol. 2004;2(10):912-20.
57. Colombel JF, Sandborn WJ, Rutgeerts P, Enns R, Hanauer SB, Panaccione R, et al. Adalimumab for maintenance of clinical response and remission in patients with Crohn's disease: the CHARM trial. Gastroenterology. 2007;132(1):52-65.
58. Le Baut G, Peyrin-Biroulet L, Bouguen G, Gornet JM, Stefanescu C, Amiot A, et al. Anti-TNF therapy for genital fistulas in female patients with Crohn's disease: a nationwide study from the Groupe d'Etude Therapeutique des Affections Inflammatoires du tube Digestif (GETAID). Aliment Pharmacol Ther. 2018;48(8):831-8.
59. Bouguen G, Huguet A, Amiot A, Viennot S, Cholet F, Nachury M, et al. Efficacy and Safety of Tumor Necrosis Factor Antagonists in Treatment of Internal Fistulizing Crohn's Disease. Clin Gastroenterol Hepatol. 2020;18(3):628-36.
60. Attauabi M, Burisch J, Seidelin JB. Efficacy of ustekinumab for active perianal fistulizing Crohn's disease: a systematic review and meta-analysis of the current literature. Scand J Gastroenterol. 2021;56(1):53-8.
61. Feagan BG, Schwartz D, Danese S, Rubin DT, Lissoos TW, Xu J, et al. Efficacy of Vedolizumab in Fistulising Crohn's Disease: Exploratory Analyses of Data from GEMINI 2. J Crohns Colitis. 2018;12(5):621-6.
62. Ayoub F, Odenwald M, Micic D, Dalal SR, Pekow J, Cohen RD, et al. Vedolizumab for perianal fistulizing Crohn's disease: systematic review and meta-analysis. Intest Res. 2022;20(2):240-50.
63. McCurdy JD, Stwalley D, Olsen MA, Deepak P. Comparative Effectiveness of Biologic Therapies in Preventing Penetrating Complications in Patients With Crohn's Disease. Clin Gastroenterol Hepatol. 2024;22(2):377-85 e5.
64. Adegbola SO, Sarafian M, Sahnan K, Pechlivanis A, Phillips RKS, Warusavitarne J, et al. Lack of anti-TNF drugs levels in fistula tissue - a reason for nonresponse in Crohn's perianal fistulating disease? Eur J Gastroenterol Hepatol. 2022;34(1):18-26.
65. Chudy-Onwugaje KO, Christian KE, Farraye FA, Cross RK. A State-of-the-Art Review of New and Emerging Therapies for the Treatment of IBD. Inflamm Bowel Dis. 2019;25(5):820-30.
66. Reinisch W, Colombel JF, D'Haens GR, Rimola J, DeHaas-Amatsaleh A, McKevitt M, et al. OP18 Efficacy and safety of filgotinib for the treatment of perianal fistulizing Crohn's Disease: Results from the phase 2 DIVERGENCE 2 study. Journal of Crohn's and Colitis. 2022/01/21;16(Supplement_1).
67. Bell SJ, Williams AB, Wiesel P, Wilkinson K, Cohen RC, Kamm MA. The clinical course of fistulating Crohn's disease. Aliment Pharmacol Ther. 2003;17(9):1145-51.
68. Garcia-Olmo D, Garcia-Arranz M, Herreros D, Pascual I, Peiro C, Rodriguez-Montes JA. A phase I clinical trial of the treatment of Crohn's fistula by adipose mesenchymal stem cell transplantation. Dis Colon Rectum. 2005;48(7):1416-23.
69. Garcia-Olmo D, Herreros D, Pascual I, Pascual JA, Del-Valle E, Zorrilla J, et al. Expanded adipose-derived stem cells for the treatment of complex perianal fistula: a phase II clinical trial. Dis Colon Rectum. 2009;52(1):79-86.
70. Panes J, Garcia-Olmo D, Van Assche G, Colombel JF, Reinisch W, Baumgart DC, et al. Expanded allogeneic adipose-derived mesenchymal stem cells (Cx601) for complex perianal fistulas in Crohn's disease: a phase 3 randomised, double-blind controlled trial. Lancet. 2016;388(10051):1281-90.
71. Panes J, Bouma G, Ferrante M, Kucharzik T, Nachury M, de la Portilla de Juan F, et al. INSPECT: A Retrospective Study to Evaluate Long-term Effectiveness and Safety of Darvadstrocel in Patients With Perianal Fistulizing Crohn's Disease Treated in the ADMIRE-CD Trial. Inflamm Bowel Dis. 2022;28(11):1737-45.
72. Park MY, Yoon YS, Lee JL, Park SH, Ye BD, Yang SK, et al. Comparative perianal fistula closure rates following autologous adipose tissue-derived stem cell transplantation or treatment with anti-tumor necrosis factor agents after seton placement in patients with Crohn's disease: a retrospective observational study. Stem Cell Res Ther. 2021;12(1):401.

73. Bataille F. Morphological characterisation of Crohn's disease fistulae. Gut. 2004;53(9).
74. Siegmund B, Feakins RM, Bamias G, Ludvig JC, Teixeira FV, Rogler G, et al. Results of the Fifth Scientific Workshop of the ECCO (II): Pathophysiology of Perianal Fistulizing Disease. Journal of Crohn's and Colitis. 2016;10(4).
75. Park JM, Kim J, Lee YJ, Bae SU, Lee HW. Inflammatory bowel disease-associated intestinal fibrosis. J Pathol Transl Med. 2023;57(1):60-6.
76. Oberhuber G, Stangl PC, Vogelsang H, Schober E, Herbst F, Gasche C. Significant association of strictures and internal fistula formation in Crohn's disease. Virchows Arch. 2000;437(3):293-7.
77. Tonelli F, Ficari F. Pathological features of Crohn's disease determining perforation. J Clin Gastroenterol. 1991;13(2):226-30.
78. Maggi L, Capone M, Giudici F, Santarlasci V, Querci V, Liotta F, et al. CD4+CD161+ T lymphocytes infiltrate Crohn's disease-associated perianal fistulas and are reduced by anti-TNF-alpha local therapy. Int Arch Allergy Immunol. 2013;161(1):81-6.
79. Bruckner RS, Spalinger MR, Barnhoorn MC, Feakins R, Fuerst A, Jehle EC, et al. Contribution of CD3+CD8- and CD3+CD8+ T Cells to TNF-alpha Overexpression in Crohn Disease-Associated Perianal Fistulas and Induction of Epithelial-Mesenchymal Transition in HT-29 Cells. Inflamm Bowel Dis. 2021;27(4):538-49.
80. Lee JM, Dedhar S, Kalluri R, Thompson EW. The epithelial-mesenchymal transition: new insights in signaling, development, and disease. The Journal of cell biology. 2006;172(7):973-81.
81. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. The Journal of clinical investigation. 2009;119(6):1420-8.
82. Shook D, Keller R. Mechanisms, mechanics and function of epithelial-mesenchymal transitions in early development. Mechanisms of development. 2003;120(11):1351-83.
83. Nieto MA, Huang RY-J, Jackson RA, Thiery JP. EMT: 2016. Cell. 2016;166(1):21-45.
84. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. Nature Reviews Molecular Cell Biology. 2019;20(2):69-84.
85. LeBleu VS, Taduri G, O'Connell J, Teng Y, Cooke VG, Woda C, et al. Origin and function of myofibroblasts in kidney fibrosis. Nature medicine. 2013;19(8):1047-53.
86. Flier SN, Tanjore H, Kokkotou EG, Sugimoto H, Zeisberg M, Kalluri R. Identification of Epithelial to Mesenchymal Transition as a Novel Source of Fibroblasts in Intestinal Fibrosis. Journal of Biological Chemistry. 2010;285(26):20202-12.
87. Jiang H, Shen J, Ran Z. Epithelial–mesenchymal transition in Crohn's disease. Mucosal Immunology. 2018;11(2):294-303.
88. Glocker E-O, Kotlarz D, Boztug K, Gertz EM, Schäffer AA, Noyan F, et al. Inflammatory Bowel Disease and Mutations Affecting the Interleukin-10 Receptor. New England Journal of Medicine. 2009;361(21):2033-45.
89. Bataille F, Rohrmeier C, Bates R, Weber A, Rieder F, Brenmoehl J, et al. Evidence for a role of epithelial mesenchymal transition during pathogenesis of fistulae in Crohn's disease. Inflammatory Bowel Diseases. 2008;14(11).
90. Scharl M, Frei S, Pesch T, Kellermeier S, Arikkat J, Frei P, et al. Interleukin-13 and transforming growth factor β synergise in the pathogenesis of human intestinal fistulae. Gut. 2013;62(1).
91. Leeb SN, Vogl D, Gunckel M, Kiessling S, Falk W, Göke M, et al. Reduced migration of fibroblasts in inflammatory bowel disease: role of inflammatory mediators and focal adhesion kinase1. Gastroenterology. 2003;125(5):1341-54.
92. Brenmoehl J, Miller SN, Hofmann C, Vogl D, Falk W, Schölmerich J, et al. Transforming growth factor-β1 induces intestinal myofibroblast differentiation and modulates their migration. World Journal of Gastroenterology. 2009;15(12):1431-.
93. Brenmoehl J, Lang M, Hausmann M, Leeb SN, Falk W, Scholmerich J, et al. Evidence for a differential expression of fibronectin splice forms ED-A and ED-B in Crohn's disease (CD) mucosa. Int J Colorectal Dis. 2007;22(6):611-23.

94. Meier JKH, Scharl M, Miller SN, Brenmoehl J, Hausmann M, Kellermeier S, et al. Specific differences in migratory function of myofibroblasts isolated from Crohn's disease fistulae and strictures. Inflammatory Bowel Diseases. 2011;17(1):202-12.
95. Ortiz-Masiá D, Gisbert-Ferrándiz L, Bauset C, Coll S, Mamie C, Scharl M, et al. Succinate Activates EMT in Intestinal Epithelial Cells through SUCNR1: A Novel Protagonist in Fistula Development. Cells. 2020;9(5):1104-.
96. Scharl M, Weber A, Fürst A, Farkas S, Jehle E, Pesch T, et al. Potential role for SNAIL family transcription factors in the etiology of Crohn's disease-associated fistulae. Inflammatory bowel diseases. 2011;17(9):1907-16.
97. Bates RC, Bellovin DI, Brown C, Maynard E, Wu B, Kawakatsu H, et al. Transcriptional activation of integrin beta6 during the epithelial-mesenchymal transition defines a novel prognostic indicator of aggressive colon carcinoma. The Journal of clinical investigation. 2005;115(2):339-47.
98. Frei SM, Pesch T, Lang S, Weber A, Jehle E, Vavricka SR, et al. A role for tumor necrosis factor and bacterial antigens in the pathogenesis of Crohn's disease-associated fistulae. Inflamm Bowel Dis. 2013;19(13):2878-87.
99. Ortiz-Masià D, Salvador P, Macias-Ceja DC, Gisbert-Ferrándiz L, Esplugues JV, Manyé J, et al. WNT2b Activates Epithelial-mesenchymal Transition Through FZD4: Relevance in Penetrating Crohn´s Disease. Journal of Crohn's and Colitis. 2020;14(2):230-9.
100. Frei SM, Hemsley C, Pesch T, Lang S, Weber A, Jehle E, et al. The Role for Dickkopf-Homolog-1 in the Pathogenesis of Crohn's Disease-Associated Fistulae. PLoS ONE. 2013;8(11):e78882-e.
101. Sullivan DE, Ferris M, Pociask D, Brody AR. Tumor necrosis factor-alpha induces transforming growth factor-beta1 expression in lung fibroblasts through the extracellular signal-regulated kinase pathway. American journal of respiratory cell and molecular biology. 2005;32(4):342-9.
102. Bates RC, Mercurio AM. Tumor Necrosis Factor-α Stimulates the Epithelial-to-Mesenchymal Transition of Human Colonic Organoids. Molecular Biology of the Cell. 2003;14(5):1790-800.
103. Ruffolo C, Scarpa M, Faggian D, Romanato G, De Pellegrin A, Filosa T, et al. Cytokine Network in Chronic Perianal Crohn's Disease and Indeterminate Colitis After Colectomy. Journal of Gastrointestinal Surgery. 2007;11(1):16-21.
104. Hodzic Z, Schill EM, Bolock AM, Good M. IL-33 and the intestine: The good, the bad, and the inflammatory. Cytokine. 2017;100:1-10.
105. Haddow JB, Musbahi O, MacDonald TT, Knowles CH. Comparison of cytokine and phosphoprotein profiles in idiopathic and Crohn's disease-related perianal fistula. World Journal of Gastrointestinal Pathophysiology. 2019;10(4):42-53.
106. Plevris N, Jenkinson PW, Arnott ID, Jones GR, Lees CW. Higher anti-tumor necrosis factor levels are associated with perianal fistula healing and fistula closure in Crohn's disease. European Journal of Gastroenterology & Hepatology. 2020;32(1):32-7.
107. Schuppan D. Fistulising Crohn's disease: MMPs gone awry. Gut. 2004;53(5).
108. von Lampe B. Differential expression of matrix metalloproteinases and their tissue inhibitors in colon mucosa of patients with inflammatory bowel disease. Gut. 2000;47(1):63-73.
109. Santana A. Attenuation of dextran sodium sulphate induced colitis in matrix metalloproteinase-9 deficient mice. World Journal of Gastroenterology. 2006;12(40):6464-.
110. Castaneda FE, Walia B, Vijay–Kumar M, Patel NR, Roser S, Kolachala VL, et al. Targeted Deletion of Metalloproteinase 9 Attenuates Experimental Colitis in Mice: Central Role of Epithelial-Derived MMP. Gastroenterology. 2005;129(6):1991-2008.
111. Goffin L, Fagagnini S, Vicari A, Mamie C, Melhem H, Weder B, et al. Anti-MMP-9 Antibody. Inflammatory Bowel Diseases. 2016;22(9):2041-57.
112. de Bruyn M, Ferrante M. Failure of MMP-9 Antagonists in IBD: Demonstrating the Importance of Molecular Biology and Well-Controlled Early Phase Studies. Journal of Crohn's and Colitis. 2018;12(9):1011-3.

113. Nelson AR, Fingleton B, Rothenberg ML, Matrisian LM. Matrix Metalloproteinases: Biologic Activity and Clinical Implications. Journal of Clinical Oncology. 2000;18(5):1135-.
114. Kirkegaard T, Hansen A, Bruun E, Brynskov J. Expression and localisation of matrix metalloproteinases and their natural inhibitors in fistulae of patients with Crohn's disease. Gut. 2004;53(5):701-9.
115. Lee IT, Lin CC, Wu YC, Yang CM. TNF-alpha induces matrix metalloproteinase-9 expression in A549 cells: role of TNFR1/TRAF2/PKCalpha-dependent signaling pathways. J Cell Physiol. 2010;224(2):454-64.
116. McGregor CGC, Tandon R, Simmons A. Pathogenesis of Fistulating Crohn's Disease: A Review. Cell Mol Gastroenterol Hepatol. 2023;15(1):1-11.
117. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature. 2011;474(7351):307-17.
118. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nature genetics. 2017;49(2):256-61.
119. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119-24.
120. Liu Z, Liu R, Gao H, Jung S, Gao X, Sun R, et al. Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. Nat Genet. 2023;55(5):796-806.
121. Latiano A, Palmieri O, Cucchiara S, Castro M, D'Incà R, Guariso G, et al. Polymorphism of the IRGM Gene Might Predispose to Fistulizing Behavior in Crohn's Disease. The American Journal of Gastroenterology. 2009;104(1):110-6.
122. Cleynen I, González JR, Figueroa C, Franke A, McGovern D, Bortlík M, et al. Genetic factors conferring an increased susceptibility to develop Crohn's disease also influence disease phenotype: results from the IBDchip European Project. Gut. 2013;62(11):1556-65.
123. Franke A, McGovern DPB, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nature genetics. 2010;42(12):1118-25.
124. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nature genetics. 2008;40(8):955-62.
125. Schnitzler F, Friedrich M, Wolf C, Stallhofer J, Angelberger M, Diegelmann J, et al. The NOD2 Single Nucleotide Polymorphism rs72796353 (IVS4+10 A>C) Is a Predictor for Perianal Fistulas in Patients with Crohn's Disease in the Absence of Other NOD2 Mutations. PloS one. 2015;10(7):e0116044-e.
126. Park HJ, Jung ES, Kong KA, Park E-M, Cheon JH, Choi JH. Identification of OCTN2 variants and their association with phenotypes of Crohn's disease in a Korean population. Scientific Reports. 2016;6(1):22887-.
127. Vermeire S, Pierik M, Hlavaty T, Claessens G, van Schuerbeeck N, Joossens S, et al. Association of organic cation transporter risk haplotype with perianal penetrating Crohn's disease but not with susceptibility to IBD. Gastroenterology. 2005;129(6).
128. Zhang M, Wang X, Jiang X, Yang X, Wen C, Zhi M, et al. Polymorphisms of the TNF Gene and Three Susceptibility Loci Are Associated with Crohn's Disease and Perianal Fistula Crohn's Disease: A Study among the Han Population from South China. Medical science monitor : international medical journal of experimental and clinical research. 2019;25:9637-50.
129. Li Y, Wang Z, Wu X, Wang G, Gu G, Ren H, et al. Intestinal mucosa-derived DNA methylation signatures in the penetrating intestinal mucosal lesions of Crohn's disease. Scientific reports. 2021;11(1).
130. Bickel M. The role of interleukin-8 in inflammation and mechanisms of regulation. J Periodontol. 1993;64(5 Suppl):456-60.

131. Mitsuhashi S, Feldbrugge L, Csizmadia E, Mitsuhashi M, Robson SC, Moss AC. Luminal Extracellular Vesicles (EVs) in Inflammatory Bowel Disease (IBD) Exhibit Proinflammatory Effects on Epithelial Cells and Macrophages. Inflamm Bowel Dis. 2016;22(7):1587-95.
132. Kessel C, Lavric M, Weinhage T, Brueckner M, de Roock S, Dabritz J, et al. Serum biomarkers confirming stable remission in inflammatory bowel disease. Sci Rep. 2021;11(1):6690.
133. Ananthakrishnan AN, Bernstein CN, Iliopoulos D, Macpherson A, Neurath MF, Ali RAR, et al. Environmental triggers in IBD: a review of progress and evidence. Nature Reviews Gastroenterology & Hepatology. 2018;15(1):39-49.
134. Zhang M, Sun K, Wu Y, Yang Y, Tso P, Wu Z. Interactions between Intestinal Microbiota and Host Immune Response in Inflammatory Bowel Disease. Front Immunol. 2017;8:942.
135. West RL, Van Der Woude CJ, Endtz HP, Hansen BE, Ouwedijk M, Boelens HAM, et al. Perianal Fistulas in Crohn's Disease Are Predominantly Colonized by Skin Flora: Implications for Antibiotic Treatment? Digestive Diseases and Sciences. 2005;50(7):1260-3.
136. Haac BE, Palmateer NC, Seaton ME, VanYperen R, Fraser CM, Bafford AC. A Distinct Gut Microbiota Exists Within Crohn's Disease–Related Perianal Fistulae. Journal of Surgical Research. 2019;242:118-28.
137. Tozer PJ, Rayment N, Hart AL, Daulatzai N, Murugananthan AU, Whelan K, et al. What role do bacteria play in persisting fistula formation in idiopathic and Crohn's anal fistula? Colorectal Disease. 2015;17(3):235-41.
138. Cane G, Ginouvès A, Marchetti S, Buscà R, Pouysségur J, Berra E, et al. HIF-1α mediates the induction of IL-8 and VEGF expression on infection with Afa/Dr diffusely adhering E. coli and promotes EMT-like behaviour. Cellular Microbiology. 2010;12(5):640-53.
139. Chandrakesan P, Roy B, Jakkula LUMR, Ahmed I, Ramamoorthy P, Tawfik O, et al. Utility of a bacterial infection model to study epithelial–mesenchymal transition, mesenchymal–epithelial transition or tumorigenesis. Oncogene. 2014;33(20):2639-54.
140. Jain U, Ver Heul AM, Xiong S, Gregory MH, Demers EG, Kern JT, et al. Debaryomyces is enriched in Crohn's disease intestinal tissue and impairs healing in mice. Science. 2021;371(6534):1154-9.
141. Li Y, Zhu W, Zuo L, Shen B. The Role of the Mesentery in Crohn's Disease. Inflammatory Bowel Diseases. 2016;22(6):1483-95.
142. Ha CWY, Martin A, Sepich-Poore GD, Shi B, Wang Y, Gouin K, et al. Translocation of Viable Gut Microbiota to Mesenteric Adipose Drives Formation of Creeping Fat in Humans. Cell. 2020;183(3).
143. Ratto C, Litta F, Lucchetti D, Parello A, Boninsegna A, Arena V, et al. Immunopathological characterization of cryptoglandular anal fistula: a pilot study investigating its pathogenesis. Colorectal Disease. 2016;18(12):O436-O44.
144. van Onkelen RS, Gosselink MP, van Meurs M, Melief MJ, Schouten WR, Laman JD. Pro-inflammatory cytokines in cryptoglandular anal fistulas. Techniques in coloproctology. 2016;20(9):619-25.
145. Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature. 2009;459(7244).
146. Pott J, Kabat AM, Maloy KJ. Intestinal Epithelial Cell Autophagy Is Required to Protect against TNF-Induced Apoptosis during Chronic Colitis in Mice. Cell host & microbe. 2018;23(2):191-202.e4.
147. Hahn S, Nam M-O, Noh JH, Lee DH, Han HW, Kim DH, et al. Organoid-based epithelial to mesenchymal transition (OEMT) model: from an intestinal fibrosis perspective. Scientific Reports. 2017;7(1).

148. Buttó LF, Pelletier A, More SK, Zhao N, Osme A, Hager CL, et al. Intestinal Stem Cell Niche Defects Result in Impaired 3D Organoid Formation in Mouse Models of Crohn's Disease-like Ileitis. Stem Cell Reports. 2020;15(2):389-407.
149. Rodansky ES, Johnson LA, Huang S, Spence JR, Higgins PDR. Intestinal organoids: a model of intestinal fibrosis for evaluating anti-fibrotic drugs. Experimental and molecular pathology. 2015;98(3).
150. Hibiya S, Tsuchiya K, Hayashi R, Fukushima K, Horita N, Watanabe S, et al. Long-term Inflammation Transforms Intestinal Epithelial Cells of Colonic Organoids. Journal of Crohn's and Colitis. 2016:jjw186-jjw.
151. Meir M, Salm J, Fey C, Schweinlin M, Kollmann C, Kannapin F, et al. Enteroids Generated from Patients with Severe Inflammation in Crohn's Disease Maintain Alterations of Junctional Proteins. Journal of Crohn's and Colitis. 2020;14(10):1473-87.
152. Nanki K, Fujii M, Shimokawa M, Matano M, Nishikori S, Date S, et al. Somatic inflammatory gene mutations in human ulcerative colitis epithelium. Nature. 2020;577(7789):254-9.
153. Rivera-Nieves J, Bamias G, Vidrich A, Marini M, Pizarro TT, McDuffie MJ, et al. Emergence of perianal fistulizing disease in the SAMP1/YitFc mouse, a spontaneous model of chronic ileitis. Gastroenterology. 2003;124(4):972-82.
154. Mc Laughlin D, Murphy P, Puri P. Altered Tbx1 gene expression is associated with abnormal oesophageal development in the adriamycin mouse model of oesophageal atresia/tracheo-oesophageal fistula. Pediatric Surgery International. 2014;30(2):143-9.
155. Chang C-J, Huang C-C, Chen P-R, Lai Y-J. Remodeling Matrix Synthesis in a Rat Model of Aortocaval Fistula and the Cyclic Stretch: Impaction in Pulmonary Arterial Hypertension-Congenital Heart Disease. International journal of molecular sciences. 2020;21(13).
156. Bruckner RS, Nissim-Eliraz E, Marsiano N, Nir E, Shemesh H, Leutenegger M, et al. Transplantation of Human Intestine Into the Mouse: A Novel Platform for Study of Inflammatory Enterocutaneous Fistulas. Journal of Crohn's and Colitis. 2019;13(6):798-806.
157. Mamie C, Bruckner RS, Lang S, Shpigel NY, Turina M, Rickenbacher A, et al. MMP9 expression in intestinal fistula from patients with fistulizing CD and from human xenograft mouse model. Tissue barriers. 2021:1994350-.
158. Flacs M, Collard M, Doblas S, Zappa M, Cazals-Hatem D, Maggiori L, et al. Preclinical Model of Perianal Fistulizing Crohn's Disease. Inflamm Bowel Dis. 2020;26(5):687-96.
159. Yu X, Abbas-Aghababazadeh F, Chen YA, Fridley BL. Statistical and Bioinformatics Analysis of Data from Bulk and Single-Cell RNA Sequencing Experiments. Methods Mol Biol. 2021;2194:143-75.
160. Stubbington MJT, Rozenblatt-Rosen O, Regev A, Teichmann SA. Single-cell transcriptomics to explore the immune system in health and disease. Science. 2017;358(6359):58-63.
161. Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. Nature. 2017;541(7637):331-8.
162. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009;6(5):377-82.
163. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med. 2018;50(8):1-14.
164. Corridoni D, Chapman T, Antanaviciute A, Satsangi J, Simmons A. Inflammatory Bowel Disease Through the Lens of Single-cell RNA-seq Technologies. Inflamm Bowel Dis. 2020;26(11):1658-68.
165. Jovic D, Liang X, Zeng H, Lin L, Xu F, Luo Y. Single-cell RNA sequencing technologies and applications: A brief overview. Clin Transl Med. 2022;12(3):e694.
166. Jia C, Hu Y, Kelly D, Kim J, Li M, Zhang NR. Accounting for technical noise in differential expression analysis of single-cell RNA sequencing data. Nucleic Acids Res. 2017;45(19):10978-88.

167. Parikh K, Antanaviciute A, Fawkner-Corbett D, Jagielowicz M, Aulicino A, Lagerholm C, et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. Nature. 2019;567(7746):49-55.
168. Kinchen J, Chen HH, Parikh K, Antanaviciute A, Jagielowicz M, Fawkner-Corbett D, et al. Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease. Cell. 2018;175(2):372-86 e17.
169. Smillie CS, Biton M, Ordovas-Montanes J, Sullivan KM, Burgin G, Graham DB, et al. Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell. 2019;178(3):714-30 e22.
170. Corridoni D, Antanaviciute A, Gupta T, Fawkner-Corbett D, Aulicino A, Jagielowicz M, et al. Single-cell atlas of colonic CD8(+) T cells in ulcerative colitis. Nat Med. 2020;26(9):1480-90.
171. Jasso GJ, Jaiswal A, Varma M, Laszewski T, Grauel A, Omar A, et al. Colon stroma mediates an inflammation-driven fibroblastic response controlling matrix remodeling and healing. PLoS Biol. 2022;20(1):e3001532.
172. Martin JC, Chang C, Boschetti G, Ungaro R, Giri M, Grout JA, et al. Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. Cell. 2019;178(6):1493-508.e20.
173. Jaeger N, Gamini R, Cella M, Schettini JL, Bugatti M, Zhao S, et al. Single-cell analyses of Crohn's disease tissues reveal intestinal intraepithelial T cells heterogeneity and altered subset distributions. Nature Communications. 2021;12(1).
174. Yokoi T, Murakami M, Kihara T, Seno S, Arase M, Wing JB, et al. Identification of a unique subset of tissue-resident memory CD4(+) T cells in Crohn's disease. Proc Natl Acad Sci U S A. 2023;120(1):e2204269120.
175. Kong L, Pokatayev V, Lefkovith A, Carter GT, Creasey EA, Krishna C, et al. The landscape of immune dysregulation in Crohn's disease revealed through single-cell transcriptomic profiling in the ileum and colon. Immunity. 2023;56(12):2855.
176. Stahl PL, Salmen F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science. 2016;353(6294):78-82.
177. Williams CG, Lee HJ, Asatsuma T, Vento-Tormo R, Haque A. An introduction to spatial transcriptomics for biomedical research. Genome Med. 2022;14(1):68.
178. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015;348(6233):aaa6090.
179. Fawkner-Corbett D, Antanaviciute A, Parikh K, Jagielowicz M, Geros AS, Gupta T, et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. Cell. 2021;184(3):810-26 e23.
180. Krausgruber T, Redl A, Barreca D, Doberer K, Romanovskaia D, Dobnikar L, et al. Single-cell and spatial transcriptomics reveal aberrant lymphoid developmental programs driving granuloma formation. Immunity. 2023;56(2):289-306 e7.
181. Lyubetskaya A, Rabe B, Fisher A, Lewin A, Neuhaus I, Brett C, et al. Assessment of spatial transcriptomics for oncology discovery. Cell Rep Methods. 2022;2(11):100340.
182. Jung N, Kim TK. Spatial transcriptomics in neuroscience. Exp Mol Med. 2023;55(10):2105-15.
183. Ferri-Borgogno S, Zhu Y, Sheng J, Burks JK, Gomez JA, Wong KK, et al. Spatial Transcriptomics Depict Ligand-Receptor Cross-talk Heterogeneity at the Tumor-Stroma Interface in Long-Term Ovarian Cancer Survivors. Cancer Res. 2023;83(9):1503-16.
184. Ozato Y, Kojima Y, Kobayashi Y, Hisamatsu Y, Toshima T, Yonemura Y, et al. Spatial and single-cell transcriptomics decipher the cellular environment containing HLA-G+ cancer cells and SPP1+ macrophages in colorectal cancer. Cell Rep. 2023;42(1):111929.
185. Lohoff T, Ghazanfar S, Missarova A, Koulena N, Pierson N, Griffiths JA, et al. Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. Nat Biotechnol. 2022;40(1):74-85.

186. Garrido-Trigo A, Corraliza AM, Veny M, Dotti I, Melon-Ardanaz E, Rill A, et al. Macrophage and neutrophil heterogeneity at single-cell spatial resolution in human inflammatory bowel disease. Nat Commun. 2023;14(1):4506.
187. Mennillo E, Kim YJ, Lee G, Rusu I, Patel RK, Dorman LC, et al. Single-cell and spatial multi-omics highlight effects of anti-integrin therapy across cellular compartments in ulcerative colitis. bioRxiv. 2024.
188. Engelhardt KR, Shah N, Faizura-Yeop I, Kocacik Uygun DF, Frede N, Muise AM, et al. Clinical outcome in IL-10– and IL-10 receptor–deficient patients with or without hematopoietic stem cell transplantation. Journal of Allergy and Clinical Immunology. 2013;131(3):825-30.e9.
189. Fawkner-Corbett D, Gerós AS, Antanaviciute A, Simmons A. Isolation of human fetal intestinal cells for single-cell RNA sequencing. STAR Protocols. 2021;2(4):100890-.
190. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. Brief Bioinform. 2013;14(2):144-61.
191. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36(5):411-20.
192. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289-96.
193. Mangiola S, Roth-Schulze AJ, Trussart M, Zozaya-Valdes E, Ma M, Gao Z, et al. sccomp: Robust differential composition and variability analysis for single-cell data. Proc Natl Acad Sci U S A. 2023;120(33):e2203828120.
194. Dann E, Henderson NC, Teichmann SA, Morgan MD, Marioni JC. Differential abundance testing on single-cell data using k-nearest neighbor graphs. Nat Biotechnol. 2022;40(2):245-53.
195. Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM, 3rd, et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol. 2018;19(1):224.
196. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, et al. The Reactome pathway Knowledgebase. Nucleic Acids Res. 2016;44(D1):D481-7.
197. Petukhov V, Xu RJ, Soldatov RA, Cadinu P, Khodosevich K, Moffitt JR, et al. Cell segmentation in imaging-based spatial transcriptomics. Nat Biotechnol. 2022;40(3):345-54.
198. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-7.
199. Lun ATL, Riesenfeld S, Andrews T, Dao TP, Gomes T, participants in the 1st Human Cell Atlas J, et al. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. Genome Biol. 2019;20(1):63.
200. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 2015;16:278.
201. Cabello-Aguilar S, Alame M, Kon-Sun-Tack F, Fau C, Lacroix M, Colinge J. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. Nucleic Acids Res. 2020;48(10):e55.
202. Risso D, Perraudeau F, Gribkova S, Dudoit S, Vert JP. A general and flexible method for signal extraction from single-cell RNA-seq data. Nat Commun. 2018;9(1):284.
203. Aibar S, Gonzalez-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods. 2017;14(11):1083-6.
204. Jorgensen PB, Fenton TM, Morbe UM, Riis LB, Jakobsen HL, Nielsen OH, et al. Identification, isolation and analysis of human gut-associated lymphoid tissues. Nat Protoc. 2021;16(4):2051-67.
205. Schreurs R, Drewniak A, Bakx R, Corpeleijn WE, Geijtenbeek THB, van Goudoever JB, et al. Quantitative comparison of human intestinal mononuclear leukocyte isolation techniques for flow cytometric analyses. J Immunol Methods. 2017;445:45-52.

206. Uronen-Hansson H, Persson E, Nilsson P, Agace W. Isolation of Cells from Human Intestinal Tissue.
207. Scharl M, Rogler G. Pathophysiology of fistula formation in Crohn's disease. World journal of gastrointestinal pathophysiology. 2014;5(3).
208. Booth C, Potten CS. Effects of IL-11 on the growth of intestinal epithelial cells in vitro. Cell Prolif. 1995;28(11):581-94.
209. Fung KY, Louis C, Metcalfe RD, Kosasih CC, Wicks IP, Griffin MDW, et al. Emerging roles for IL-11 in inflammatory diseases. Cytokine. 2022;149:155750.
210. Nayar S, Morrison JK, Giri M, Gettler K, Chuang LS, Walker LA, et al. A myeloid-stromal niche and gp130 rescue in NOD2-driven Crohn's disease. Nature. 2021;593(7858):275-81.
211. Crawley SW, Shifrin DA, Jr., Grega-Larson NE, McConnell RE, Benesh AE, Mao S, et al. Intestinal brush border assembly driven by protocadherin-based intermicrovillar adhesion. Cell. 2014;157(2):433-46.
212. Gregorieff A, Stange DE, Kujala P, Begthel H, van den Born M, Korving J, et al. The ets-domain transcription factor Spdef promotes maturation of goblet and paneth cells in the intestinal epithelium. Gastroenterology. 2009;137(4):1333-45 e1-3.
213. O'Callaghan B, Bosch AM, Houlden H. An update on the genetics, clinical presentation, and pathomechanisms of human riboflavin transporter deficiency. J Inherit Metab Dis. 2019;42(4):598-607.
214. Sakthianandeswaren A, Christie M, D'Andreti C, Tsui C, Jorissen RN, Li S, et al. PHLDA1 expression marks the putative epithelial stem cells and contributes to intestinal tumorigenesis. Cancer Res. 2011;71(10):3709-19.
215. Brenner CM, Choudhary M, McCormick MG, Cheung D, Landesberg GP, Wang JF, et al. BAG3: Nature's Quintessential Multi-Functional Protein Functions as a Ubiquitous Intra-Cellular Glue. Cells. 2023;12(6).
216. Kelleher M, Singh R, O'Driscoll CM, Melgar S. Carcinoembryonic antigen (CEACAM) family members and Inflammatory Bowel Disease. Cytokine Growth Factor Rev. 2019;47:21-31.
217. Nonnecke EB, Castillo PA, Dugan AE, Almalki F, Underwood MA, De La Motte CA, et al. Human intelectin-1 (ITLN1) genetic variation and intestinal expression. Sci Rep. 2021;11(1):12889.
218. Liu Y, Wan Y, Jiang Y, Zhang L, Cheng W. GPX4: The hub of lipid oxidation, ferroptosis, disease and treatment. Biochim Biophys Acta Rev Cancer. 2023;1878(3):188890.
219. Teubner W, Meinl W, Florian S, Kretzschmar M, Glatt H. Identification and localization of soluble sulfotransferases in the human gastrointestinal tract. Biochem J. 2007;404(2):207-15.
220. Verset L, Tommelein J, Moles Lopez X, Decaestecker C, Mareel M, Bracke M, et al. Epithelial expression of FHL2 is negatively associated with metastasis-free and overall survival in colorectal cancer. Br J Cancer. 2013;109(1):114-20.
221. Liu J, Fu N, Yang Z, Li A, Wu H, Jin Y, et al. The genetic and epigenetic regulation of CD55 and its pathway analysis in colon cancer. Front Immunol. 2022;13:947136.
222. Qin Q, Xu Y, He T, Qin C, Xu J. Normal and disease-related biological functions of Twist1 and underlying molecular mechanisms. Cell Res. 2012;22(1):90-106.
223. Torsello B, De Marco S, Bombelli S, Chisci E, Cassina V, Corti R, et al. The 1ALCTL and 1BLCTL isoforms of Arg/Abl2 induce fibroblast activation and extra cellular matrix remodelling differently. Biol Open. 2019;8(3).
224. Chiang TS, Lin MC, Tsai MC, Chen CH, Jang LT, Lee FS. ADP-ribosylation factor-like 4A interacts with Robo1 to promote cell migration by regulating Cdc42 activation. Mol Biol Cell. 2019;30(1):69-81.
225. Huang B, Chen Z, Geng L, Wang J, Liang H, Cao Y, et al. Mucosal Profiling of Pediatric-Onset Colitis and IBD Reveals Common Pathogenics and Therapeutic Pathways. Cell. 2019;179(5):1160-76.e24.

226. Schafer S, Viswanathan S, Widjaja AA, Lim WW, Moreno-Moral A, DeLaughter DM, et al. IL-11 is a crucial determinant of cardiovascular fibrosis. Nature. 2017;552(7683):110-5.
227. Zhao L, Westerhoff M, Hornick JL, Krausz T, Antic T, Xiao SY, et al. Loss of microfibril-associated protein 5 (MFAP5) expression in colon cancer stroma. Virchows Arch. 2020;476(3):383-90.
228. Xu H, Acott TS, Wirtz MK. Identification and expression of a novel type I procollagen C-proteinase enhancer protein gene from the glaucoma candidate region on 3q21-q24. Genomics. 2000;66(3):264-73.
229. Voon DC, Thiery JP. The Emerging Roles of RUNX Transcription Factors in Epithelial-Mesenchymal Transition. Adv Exp Med Biol. 2017;962:471-89.
230. Vanuytsel T, Senger S, Fasano A, Shea-Donohue T. Major signaling pathways in intestinal stem cells. Biochim Biophys Acta. 2013;1830(2):2410-26.
231. Reya T, Clevers H. Wnt signalling in stem cells and cancer. Nature. 2005;434(7035):843-50.
232. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. Nat Rev Mol Cell Biol. 2014;15(12):786-801.
233. Lee KW, Yeo SY, Gong JR, Koo OJ, Sohn I, Lee WY, et al. PRRX1 is a master transcription factor of stromal fibroblasts for myofibroblastic lineage progression. Nat Commun. 2022;13(1):2793.
234. Lan Y, Ovitt CE, Cho ES, Maltby KM, Wang Q, Jiang R. Odd-skipped related 2 (Osr2) encodes a key intrinsic regulator of secondary palate growth and morphogenesis. Development. 2004;131(13):3207-16.
235. Martin JF, Bradley A, Olson EN. The paired-like homeo box gene MHox is required for early events of skeletogenesis in multiple lineages. Genes Dev. 1995;9(10):1237-49.
236. Bodnar RJ, Satish L, Yates CC, Wells A. Pericytes: A newly recognized player in wound healing. Wound Repair Regen. 2016;24(2):204-14.
237. Armulik A, Genove G, Betsholtz C. Pericytes: developmental, physiological, and pathological perspectives, problems, and promises. Dev Cell. 2011;21(2):193-215.
238. Mahley RW. Central Nervous System Lipoproteins: ApoE and Regulation of Cholesterol Metabolism. Arterioscler Thromb Vasc Biol. 2016;36(7):1305-15.
239. Stechschulte LA, Hinds TD, Jr., Khuder SS, Shou W, Najjar SM, Sanchez ER. FKBP51 controls cellular adipogenesis through p38 kinase-mediated phosphorylation of GRalpha and PPARgamma. Mol Endocrinol. 2014;28(8):1265-75.
240. Murphy-Ullrich JE. Thrombospondin 1 and Its Diverse Roles as a Regulator of Extracellular Matrix in Fibrotic Disease. J Histochem Cytochem. 2019;67(9):683-99.
241. David G, Bai XM, Van der Schueren B, Marynen P, Cassiman JJ, Van den Berghe H. Spatial and temporal changes in the expression of fibroglycan (syndecan-2) during mouse embryonic development. Development. 1993;119(3):841-54.
242. Clevers H, Nusse R. Wnt/beta-catenin signaling and disease. Cell. 2012;149(6):1192-205.
243. Lewis A, Sanchez S, Berti G, Pan-Castillo B, Nijhuis A, Mehta S, et al. Small-molecule Wnt inhibitors are a potential novel therapy for intestinal fibrosis in Crohns disease. Clin Sci (Lond). 2022;136(19):1405-23.
244. Claessen MM, Schipper ME, Oldenburg B, Siersema PD, Offerhaus GJ, Vleggaar FP. WNT-pathway activation in IBD-associated colorectal carcinogenesis: potential biomarkers for colonic surveillance. Cell Oncol. 2010;32(4):303-10.
245. Gregorieff A, Pinto D, Begthel H, Destree O, Kielman M, Clevers H. Expression pattern of Wnt signaling components in the adult intestine. Gastroenterology. 2005;129(2):626-38.
246. Wang Y, Liu B, Zhao G, Lee Y, Buzdin A, Mu X, et al. Spatial transcriptomics: Technologies, applications and experimental considerations. Genomics. 2023;115(5):110671.
247. Magro F, Langner C, Driessen A, Ensari A, Geboes K, Mantzaris GJ, et al. European consensus on the histopathology of inflammatory bowel disease. Journal of Crohn's and Colitis. 2013;7(10).

248. Sekikawa A, Fukui H, Fujii S, Takeda J, Nanakin A, Hisatsune H, et al. REG Ialpha protein may function as a trophic and/or anti-apoptotic factor in the development of gastric cancer. Gastroenterology. 2005;128(3):642-53.
249. Mao H, Jia J, Sheng J, Zhang S, Huang K, Li H, et al. Protective and anti-inflammatory role of REG1A in inflammatory bowel disease induced by JAK/STAT3 signaling axis. Int Immunopharmacol. 2021;92:107304.
250. Mollenhauer J, Herbertz S, Holmskov U, Tolnay M, Krebs I, Merlo A, et al. DMBT1 encodes a protein involved in the immune defense and in epithelial differentiation and is highly unstable in cancer. Cancer Res. 2000;60(6):1704-10.
251. Diegelmann J, Czamara D, Le Bras E, Zimmermann E, Olszak T, Bedynek A, et al. Intestinal DMBT1 expression is modulated by Crohn's disease-associated IL23R variants and by a DMBT1 variant which influences binding of the transcription factors CREB1 and ATF-2. PLoS One. 2013;8(11):e77773.
252. Renner M, Bergmann G, Krebs I, End C, Lyer S, Hilberg F, et al. DMBT1 confers mucosal protection in vivo and a deletion variant is associated with Crohn's disease. Gastroenterology. 2007;133(5):1499-509.
253. Hayes P, Dhillon S, O'Neill K, Thoeni C, Hui KY, Elkadri A, et al. Defects in NADPH Oxidase Genes NOX1 and DUOX2 in Very Early Onset Inflammatory Bowel Disease. Cell Mol Gastroenterol Hepatol. 2015;1(5):489-502.
254. Gustafsson JK, Johansson MEV. The role of goblet cells and mucus in intestinal homeostasis. Nat Rev Gastroenterol Hepatol. 2022;19(12):785-803.
255. Dzialo E, Czepiel M, Tkacz K, Siedlar M, Kania G, Blyszczuk P. WNT/beta-Catenin Signaling Promotes TGF-beta-Mediated Activation of Human Cardiac Fibroblasts by Enhancing IL-11 Production. Int J Mol Sci. 2021;22(18).
256. Sarabia-Sanchez MA, Robles-Flores M. WNT Signaling in Stem Cells: A Look into the Non-Canonical Pathway. Stem Cell Rev Rep. 2024;20(1):52-66.
257. Ma L, Zhang X, Zhang C, Hou B, Zhao H. FOSL1 knockdown ameliorates DSS-induced inflammation and barrier damage in ulcerative colitis via MMP13 downregulation. Exp Ther Med. 2022;24(3):551.
258. Hoeft K, Schaefer GJL, Kim H, Schumacher D, Bleckwehl T, Long Q, et al. Platelet-instructed SPP1(+) macrophages drive myofibroblast activation in fibrosis in a CXCL4-dependent manner. Cell Rep. 2023;42(2):112131.
259. Harvey NL, Gordon EJ. Deciphering the roles of macrophages in developmental and inflammation stimulated lymphangiogenesis. Vasc Cell. 2012;4(1):15.
260. Levantovsky RM, Tastad C, Zhang J, Gettler K, Sabic K, Werner R, et al. Multimodal single-cell analyses reveal mechanisms of perianal fistula in diverse patients with Crohn's disease. Med. 2024.
261. Poindexter NJ, Williams RR, Powis G, Jen E, Caudle AS, Chada S, et al. IL-24 is expressed during wound repair and inhibits TGFalpha-induced migration and proliferation of keratinocytes. Exp Dermatol. 2010;19(8):714-22.
262. Mao R, Kurada S, Gordon IO, Baker ME, Gandhi N, McDonald C, et al. The Mesenteric Fat and Intestinal Muscle Interface: Creeping Fat Influencing Stricture Formation in Crohn's Disease. Inflamm Bowel Dis. 2019;25(3):421-6.
263. Rieder F. The gut microbiome in intestinal fibrosis: environmental protector or provocateur? Sci Transl Med. 2013;5(190):190ps10.
264. Rieder F, Fiocchi C. Intestinal fibrosis in IBD--a dynamic, multifactorial process. Nat Rev Gastroenterol Hepatol. 2009;6(4):228-35.
265. Mukherjee PK, Nguyen QT, Li J, Zhao S, Christensen SM, West GA, et al. Stricturing Crohn's Disease Single-Cell RNA Sequencing Reveals Fibroblast Heterogeneity and Intercellular Interactions. Gastroenterology. 2023;165(5):1180-96.
266. Jagt JZ, van Rheenen PF, Thoma SMA, Gower J, Reimering-Hartgerink PB, van der Wielen H, et al. The top 10 research priorities for inflammatory bowel disease in children and young adults: results of a James Lind Alliance Priority Setting Partnership. Lancet Gastroenterol Hepatol. 2023;8(8):690-1.

267. Geldof J, Iqbal N, LeBlanc JF, Anandabaskaran S, Sawyer R, Buskens C, et al. Classifying perianal fistulising Crohn's disease: an expert consensus to guide decision-making in daily practice and clinical trials. Lancet Gastroenterol Hepatol. 2022;7(6):576-84.